Draft for UCLA Law, Economics & Organization Workshop
Please Do Not Cite Without Permission

Explorations in the Theory of Optimal Consumption Taxes

James R. Hines Jr.[1]
University of Michigan and NBER

Edward J. McCaffery[2]
University of Southern California
California Institute of Technology

Abstract

Mirrlees (1971) developed the theory of optimal income taxation using a one period model with an income (labor earnings) tax and no savings. In a multi-period world with savings, the choice of tax system affects the impact of marginal tax rates on labor supply. Under a progressive cash-flow, postpaid consumption or (equivalently) expenditure tax, where the tax falls on present period expenditures, the taxpayer can avoid present period taxation by saving and facing a deferred tax on ultimate consumption. Developing an optimal expenditure tax rate function depends on the taxpayer's perceived tax rate on future consumption, and hence on her motives for saving. There is apt to be significant heterogeneity of savings motives in a broad population. This article presents several possible specifications. We conclude that under plausible assumptions, the optimum annual expenditure tax rate structure can be significantly more positively increasing in income (that is, more "progressive") than under typical optimum income tax models, and that what is perhaps the canonical finding of the Mirrlees literature—that the marginal tax rate declines to zero at the highest income level—does not characterize an expenditure tax implemented on a yearly basis.

[1] jrhines@umich.edu.
[2] emccaffe@hss.caltech.edu. An earlier version of this paper was presented at the Harvard Law School Seminar on Current Reseach in Taxation, and we thank Harvard and the conference organizers for their support, and all the attendees for their comments, especially Jeff Strnad for his detailed commentary,

1. Introduction

James Mirrlees 1971 famously developed a theory of optimal income taxation, building on the much earlier optimal commodity tax work of Frank Ramsey 1927. See also Mirrlees and Diamond 1971, Bankman and Griffith 1987, Auerbach and Hines 2002. Whereas Ramsey demonstrated that an optimally efficient government should differentiate commodity tax rates based on demand conditions, generally leading to the "inverse elasticity rule" whereby inelastically demanded goods should be most highly taxed, Mirrlees analyzed the properties of income taxes designed to redistribute income in the most efficient possible manner, given the information limits facing governments. Mirrlees modeled a tax on income or labor earnings in a one-period setting that lacked savings, by specification. The principal implications of the Mirrlees model and its progeny are that, given a redistributive social welfare function and with certain restrictions on technology and skills, the marginal tax rate on labor earnings should (1) never be negative, (2) attain its peak in the middle of the income range and be declining at the upper ranges of income for workers with high wage-earning ability, and (3) reach zero on the last dollar earned by the highest ability wage-earner. Progressivity in average tax rates is achieved by means of cash (or in-kind) transfers, sometimes referred to as "demogrants," combined with the pattern of often increasing, but intermittently declining, marginal rates.

The intuitions behind the complex mathematics of the Mirrlees model are simple enough to state. The pattern of optimal tax rates reflects tradeoffs between the distorting effects of taxes on labor supply decisions, on the one hand, and the benefits of producing and then redistributing income via the tax system, on the other. While higher marginal tax rates at low income levels have the undesirable feature of discouraging work effort by low ability taxpayers, they also offer the prospect of raising significant tax revenue from all taxpayers, including the high income or ability ones. The government can then redistribute the resulting tax revenues to all citizens, including the

poor, via demogrants. High marginal tax rates at upper incomes and abilities, in contrast, while they also distort labor supply, here of high income/ability taxpayers, raise only modest amounts of revenue, because they only apply to the relatively more elastic high ability workers. Hence the declining marginal tax rate over the upper ranges of the income distribution. In the limiting case, it is better—for all, under the standard Paretian condition—for a worker to work than not to work. Hence the zero tax rate on the otherwise fully-deterred highest potential wage earner. Another way to express the same concept is that there is nothing for society to gain by imposing a positive marginal tax rate on the most able worker, and thereby distorting his or her labor supply, since such taxes yield no greater revenue than lesser distorting alternatives.

Many have debated how general the Mirrlees results are, and what the precise details should be—where, exactly, the marginal rate should begin to decline and where (if at all) it should reach zero. For example, Diamond 1998 reports that the top marginal rate does not reach zero in his model with an unbounded distribution of individual types. Dahan and Strawczynski 2000 argue that Diamond's result is attributable not to the unbounded distribution of individual types but instead to the very particular form of the utility function that Diamond (and some subsequent authors) adopted for modeling convenience. Stiglitz (1982) shows in contrast that the top income tax rate can be *negative* in a setting in which labor of different types is not perfectly substitutable. There are yet other generalizations of the basic Mirrlees result, in a flourishing optimum income tax literature. See Auerbach and Hines 2002 for a partial survey.

Despite its highly abstract quality and nearly four decades of qualifications, questions, and refinements, Mirrlees' influence can be seen strongly in contemporary tax systems, such as that presently obtaining in the United States. The American tax system features tax rate curves that peak in the middle income ranges and then gradually decline, once the effect of payroll taxes and

more particular features such as the phase-outs of items of deduction and exclusion are aggregated onto the income tax. See Figure 1, which is bounded at $100,000 of labor income, and which ignores the low taxation of the yield to capital, an important source of revenues for the upper-income. Meanwhile, the political argument against high marginal tax rates on high earners has been a powerful one, at least since the presidency of Ronald Reagan in the 1980s. See generally Steurle 2005, Slemrod and Bakija 2006, Feldstein 2006.

At nearly the same time of the dominance of a Mirrlees inspired, optimum tax approach to the questions of tax's rate structure, there has been a growing consensus in favor of consumption-based taxes to replace some or all of the current income tax. See generally Auerbach forthcoming, Shaviro 2007, McCaffery 2005a and b, Bradford 1996, Weisbach 2006, Graetz 2007, Hines 2007. A good deal of the popularity of consumption-based taxes, at least among economists and like-minded policy-makers, stems from the efficiency advantages of not taxing savings and the simplification benefits of not needing to tax, and therefore measure, "income." Analysts consider direct taxation of the return to saving to be extremely distortionary because inter-temporal consumption elasticities—the responsiveness of allocating consumption among time periods to the "price" of savings—are large, implying that relatively small tax differentials or "wedges" between present and deferred consumption can produce large compensated changes in behavior,[3] and therefore generate large efficiency costs. Further, the tax wedges between consumption in differing periods grow exponentially as the time between periods lengthen. This feature of inter-temporal taxation led to certain celebrated findings of Judd 1985 and Chamley 1986, namely that efficient taxation entails zero taxes on returns to saving in the steady state: an observation long found in the less technical political economic and legal literature on taxation (see e.g., Mill 1848,

[3] Consistent with standard practice in neoclassical public finance, we consider compensated changes in behavior, taking out the income effect to isolate the substitution effect of the tax at issue.

Andrews 1974, Klein 1977). Subsequent work has considered settings in which private markets are incomplete in one dimension or another, or government tax instruments limited, finding that, in such circumstances, the optimal inter-temporal tax rate need not be zero. See generally Correia 1996, and the numerical examples provided by Golosov, Kocherlakota and Tsyvinski 2003; Golosov and Tsyvinski 2005; and Golosov, Tsyvinski and Werning 2006. Indeed, many of these models of incomplete markets and limited government instruments share the striking feature that the behavior of taxpayers with significant assets on hand (i.e., financial wealth) is more responsive to marginal tax rates than is the behavior of otherwise similar individuals with less savings, such that optimal tax configurations generally call for discouraging savings as a means to extract revenue from taxpayers with heavy unavoidable tax burdens on labor income (that is, driving up their inelasticity of their labor supply to the tax rate). While it is possible to construct examples of stylized economies in which markets are sufficiently incomplete to make such tax policy a sensible course of action, the high efficiency cost of distorting inter-temporal choice through taxation has created a normative presumption in favor of zero, or close to zero, marginal tax rates on returns to saving, which has also been the more politically viable option.

There has however been no general union of the Mirrlees optimal tax literature with contemporary calls for consumption-based tax reform, applying optimal tax principles to tax reform options being actively considered by governments world-wide. Our central question in this preliminary exploration is what an optimum expenditure tax rate structure might look like. In posing this question, we have one foot solidly in the world of practical political constraints, considering a broadly practical tax reform option, a form of consumption or expenditure tax developed first at any length by Kaldor (1955) and later advanced by Andrews 1974, Bradford et al. 1977, in the "USA" tax plan advanced in the U.S. Congress in the mid 1990s (as discussed in Seidman 1995), and furthered discussed in McCaffery 2002. (The optimal tax literature, in

contrast, tends to be unconstrained in its political dimensions, and yet typically makes highly restrictive assumptions in its representative agent models in order to produce tractable analyses). The problem, even thus delimited, turns out to be surprisingly difficult, in significant part because of the challenges associated with specifying compelling models of individual savings behavior. See generally Bernheim 2002, Koltlikoff 1989. Our results are halting and preliminary, yet cast significant doubt on what might otherwise be taken to be the implications of the Mirrlees model, namely that marginal tax rates should be declining in consumption, and should reach zero for the individual with the highest consumption level.

2. Three forms of Taxation

It is important to get certain basic terms down, for there are alternative forms of consumption taxation that are not equivalent under varying marginal tax rates, and confusion can result for want of clear definitions. Much of the "income versus consumption" debate wages over the past several decades among public finance economists, lawyers, academics and policymakers has stumbled over semantics.

An income tax falls on all present period earnings, from "whatever source derived," in the language of the XVIth Amendment to the U.S. Constitution and Section 61 of the U.S. Internal Revenue Code ("IRC"), that is, from labor or capital (and, arguably, from beneficent transfers as well) (see Simons 1938, McCaffery 2005a). The inclusion of the yield to capital in the income tax base makes it a "double tax" on earnings that are not immediately consumed. Mill 1848. Largely to avoid this double-tax sting and its attendant distortions, critics, even before the efficiency-based analysis of Judd 1955 and Chamley 1986, have long advocated some form of consumption tax that would exempt savings, or the yield to capital, from the tax base. Kaldor 1955, Andrews 1974. Early theorists posited the equivalence of two forms of consumption-based taxes under

plausible assumptions, namely the "prepaid" or "yield-exempt" model and the "postpaid" or cash-flow one (Andrews 1974, Bradford et al 1977, Bradford 1986). If tax rates do not change and there is no effect on the rate of return on savings, then levying a tax upfront, and never again, as in a typical wage or payroll tax, is indeed equivalent to deferring the tax and levying it at the single time of ultimate private preclusive use, as in a typical expenditure, sales or value-added tax. Andrews 1974, McCaffery 2005a and b. Consider the simple formula for the future value (FV) of a present amount of principal (P), invested over n periods at a rate of return r:

$$FV = P\,(1 + r)^{n}$$

It does not matter when the government reduces this by a tax rate, t, leaving the taxpayer to consume (1 – t), under the commutative principle of multiplication:

$$(1 - t)\,P\,(1 + r)^{n} = P\,(1 + r)^{n}\,(1 - t)$$

Under variable marginal tax rates, however, the equivalence of "prepaid" and "postpaid" consumption taxes is destroyed. The tax rate at the time of initial earnings may not be the same as that at the time of consumption. This analytic fact has led many to argue for flat or constant-rate expenditure taxes, because it is thought to be a norm not to tax capital or its yield. But this argument, generated by policymakers invoking norms of "fairness"—and hence parallel to but distinct from the efficiency-based arguments of Judd 1985 and Chamley 1986, discussed above—presumes that the normative reason for choosing a consumption tax turns on the principled non-taxation of the yield to savings. McCaffery 2005a. It need not be so. Policymakers might desire a progressive tax on expenditures for reasons unrelated to the non-taxation of capital or its yield, or at least not pre-committed to such non-taxation. For example, a utility-maximizing

outcome might be achieved under the social welfare function with a progressive expenditure tax, because individuals in the face of the tax may continue to work and save, avoiding present period taxation but improving social welfare both by the positive externalities of their contributions to the capital stock and by the implicit deferred tax on the savings. We do not in this Article compare the social optimum *across* tax bases, income versus consumption (and possibly multiple specifications of consumption tax), but rather sketch out the contours of a likely optimum tax rate schedule *within* a particular consumption tax base, a logically prior task.[4]

Once the assumption of flat or constant tax rates under a consumption tax base is set aside, the analytics show that a pure or ideal income tax "double taxes" all savings, whatever their use; a prepaid consumption or wage tax ignores the yield to all savings, again whatever their use; while a progressive postpaid consumption, expenditure or (all equivalently) cash-flow tax allows capital transactions (borrowing, saving, investing) to sometimes increase, and sometimes decrease, the total burden of taxation, in present value terms, depending on the use of the associated savings. McCaffery 2005a. In this article, in line with other contemporary writers (e.g. Auerbach forthcoming), we analyze only a progressive postpaid or cash-flow expenditure consumption tax—an expenditure tax, as we shall henceforth refer to it—and not a wage tax. We also, and again largely for purely practical and political reasons, limit our analysis to an expenditure tax

[4] That is, we take it as a logically prior task, given the practical political constraints we accepted above. Thus the broad delineation of our project is to consider prepaid consumption or wage taxes—where the Mirrlees analysis holds pretty much as is, since he developed his insights in a one-period model; possibly an ideal income tax, to the extent this is in fact considered a viable policy option (we have our doubts, given the history of exceptions to the taxation of savings under the U.S. income tax, which has made this tax, at best, a "hybrid" income-consumption one, and, at worst, a "wage tax in drag" (see McCaffery 2002 and 2005a); and a postpaid consumption or expenditure tax each with its own "optimal" rate structure, and then to compare the combined *base and rate* structure. We understand that there are other logics as, for example, developing the optimal tax base and rate structure, combined, in a universe free from all practical and political constraints.

levied on an annual returns basis.[5] Under such a tax, capital transactions (borrowing, saving, investing) that are used to "smooth" consumption profiles, so as to bring annual (or other period) consumption closer to the level of average annual lifetime labor earnings—as for example borrowing in youth, to anticipate earnings, and saving for retirement in mid-life, to defer earnings—lower the average burden of taxes under a marginal rate system with monotonically increasing marginal tax rates (to illustrate). In contrast, capital transactions that allow for greater differences between consumption and average annual labor earnings—as for example supra-marginal returns from investments or net accessions from beneficent transfers—raise the average tax burden on the individual. Thus it becomes essential under a cash-flow or expenditure style consumption tax to specify the reasons for taxpayer savings, a task we take up below in Section 4, and to set out an optimal rate structure.

3. Modeling an Optimum Cash-Flow Consumption Tax

In order to progress beyond Mirrlees to consider optimal tax rates under an expenditure tax, we need to add at least a second period to the model, to allow for the possibility of savings. Following Mirrlees, we take all individuals to share the same utility function, with consumption and leisure as arguments; the utility function is now modified to include two types of consumption:

[5] As we note below, the assumption of annualized returns for an individualized tax strikes as us realistic. It does mean that we rule out "consumption averaging," along the lines of Vickrey income averaging. For a variety of reasons, however, including the ability to consumption-smooth by self help vehicles readily available in capital markets (e.g., leasing or installment sale purchases), we think the problem of "consumption bunching" tends to be overstated in the literature. See McCaffery 2005a for fuller discussion. Annual returns without an averaging or other "look-back" mechanism do, however, mean that the government fails to use some potentially valuable information, namely prior period expenditures, which information could be used by a Mirrlees inspired decision-maker. Hence our assumption of annual returns is another concession to real-world constraints.

$$U(C_1, C_2, L),$$

in which C_1 and C_2 represent first and second period consumption, respectively. We assume that consumption depends on first-period labor earnings, given by the product of a wage rate, W, and the labor decision 1 – L, where L is normalized to represent the portion of time spent on leisure, 0 > L > 1. We also assume that there are no labor earnings in period 2, representing a period of retirement from the labor market. Finally, consumers are assumed to have access to a capital market that permits borrowing (although there is no borrowing in our model, givn the specifications of wage income in period 1 only) and lending at rate r. (This rate is quite possibly a function of the tax system, but for much of the analysis we take it to be exogenous to the model, perhaps reflecting the impact of a world capital market.) The budget constraint that an individual cannot spend more in present value than his first period labor earnings can be expressed as:

$$C_1 + \frac{C_2}{(1+r)} \leq W(1-L).$$

This simple model holds fixed the timing of the retirement decision, quite possibly an important restriction given the potential impact of consumption taxation on financial resources available in retirement. Note that the model does not necessarily assume that C_2 consists entirely of own-consumption in period 2, a point that becomes relevant in the subsequent discussion of bequests and inter-generational savings.

Under Mirrlees, the sole deterrent effect of the tax rate on the individual was felt on the labor supply decision. The declining marginal tax rate among high earners, falling to zero for the highest dollar earned, followed. At least given the standard no-envy Paretian constraint, the

government will always prefer revenue, which can be used for utility-enhancing transfers, to no revenue, and some work effort in the limiting case to none. With taxes added to Mirrlees' one-period model, the representative agent faces a budget constraint reflecting that consumption cannot exceed after-tax income. With a two period model with savings, the agent's constraint is

$$C_1 + T(C_1) + \frac{C_2 + T(C_2)}{(1+r)} \leq W(1-L),$$

In which T(C) is the (consumption basis) tax obligation associated with consumption level C.

We hold the T(C) function to be unchanging over time, although it should be understood as being based on a vector of marginal tax rates applied to different levels of consumption, producing a unique average tax rate for each level of consumption over an initial zero bracket, if there is such a bracket. Holding the T(C) function to be unchanging assumes away the macroeconomic difficulties of different tax rate schedules over time, and the associated problem of government credibility in committing to a future tax rate schedule. This is a simplifying assumption, but note here, that the critical variable is the taxpayer's *perceived* or anticipated tax rate at a later time; the problem has a similar structure to modeling savings and investment decisions in the face of varying inflation rates (indeed, the inflation rate can be seen to operate like a tax), and so here we assume that taxpayers rationally expect no change.

The critical distinction from Mirrlees is that agents in this simple model now face two, not one, control variables: how much to work, 1 – L, and how much to save, $C_2/(1+r)$. It is thus essential to the modeling task to develop realistic models of why people save, for this will affect taxpayer's analysis of the taxes due in the later period. We turn to this issue next.

4. Motives for Savings

Most contemporary taxpayers do not save, of course, and for them, and over large ranges of the earnings scale, income more or less equals consumption, and so the Mirrlees specification is an excellent approximation to the economic environment they face. Kawachi et al. 2006, Moomau forthcoming.

But the rich, as they say, are different. High-earning and wealthy taxpayers do save. It is also over the upper-income ranges where the Mirrlees results are most dramatic, and have been most influential. The central question in developing a model of optimal expenditure tax rates is thus to ascertain why the upper-income save. Unfortunately, this is not easy; specifying a compelling model of individual savings behavior has been a holy grail of sorts for economists of all stripes. See generally Bernheim 2002, Kotlikoff 1989 for helpful background. In much of the discussion that follows, we posit that at least high income/ability individuals reach a point of full or partial "consumption satiation," in which they no longer seek to consume all available resources in the present period but instead reserve some resources for a later period. Again, the question remains why they do this.

In this Section, we canvass some possible motives behind savings behavior, and note their relevance for a theory of optimal expenditure tax rates. We also note at the start that we see no reason for assuming homogeneity among the population of savers, even though heterogeneity makes the modeler's task difficult.

A. *Life-Cycle or "Permanent Income" Savings*

The early speculation of Keynes that individuals used savings to make for a "permanent income" has been supplanted with the "life cycle" savings model typically attributed to Modigliani and

Brumberg 1954; see also Friedman 1957 and Modigliani 1988. The idea is simple and intuitive: people save to convert uneven labor market earnings into smooth consumption profiles. For an individual with no sources of outside financing, as from altruistically linked family members, this would mean borrowing in youth, saving in the peak earnings years of mid-life, and drawing down on the saved reserves in late-life. Although there is good reason to question whether or not a sizeable number of individuals follow a "rational life cycle hypothesis" in their savings behavior (Shefrin and Thaler 1988, but see Scholz et al. 2006), much of American retirement savings policy seems built with a life-cycle model in mind: for example, the minimum distribution rules under tax-favored pension plans within the current U.S. income tax, or the very structure of the social security system. These mechanisms encourage, and in some sense compel, taxpayers to follow a life cycle model.

One implication of life-cycle savings on the optimal expenditure tax rate schedule under our simple model is that the social goal of permitting (or encouraging) individuals to "smooth" their consumption is inconsistent with certain expenditure tax schedules. For example, one might envision a tax schedule with marginal tax rates that fall rapidly with consumption over certain ranges, much as the Mirrlees model can imply rapidly falling income tax rates, as for example in some of the simulations considered by Tuomala 1990. With such rapidly declining tax rates in the schedule, however, it is entirely possible that individuals would choose not to smooth their consumption, but instead select consumption profiles that vary wildly between accounting periods, in order to benefit from the reduced tax rates available at higher levels of expenditure in the present or any prior period. The problem—and it is intrinsic to any periodic consumption tax schemes, such as one based on annual returns—is that the present discounted value of lifetime taxes depends not merely on the present discounted value of lifetime consumption, but also on the details of its distribution across years, or other limited accounting period.

This point can be made considerably more general. Consider a generalization of a Mirrlees-type setting, in which the economy is populated by individuals with identical utility functions but differing abilities, all of whom live for two periods. The first-order conditions corresponding to the individual's utility maximizing choice of first and second period consumption, and labor supply, are:

$$\frac{\partial u}{\partial C_1} = \lambda\left[1 + T'\left(C_1\right)\right]$$

$$\frac{\partial u}{\partial C_2} = \frac{\lambda}{\left(1 + r\right)}\left[1 + T'\left(C_2\right)\right]$$

$$\frac{\partial u}{\partial L} = \lambda W ,$$

in which λ is the Lagrange multiplier corresponding to the individual's budget constraint, and $T'\left(C_i\right)$ is the derivative of the (assumed differentiable) tax function, or marginal consumption tax rate, evaluated at C_i.

As is well understood from the Mirrlees problem, which is set in the domain of information economics, self-selection constraints do not bind on the individuals at the top of the skill distribution. See, for example, explanations provided by Stiglitz 1987, Tuomala 1990 and Auerbach and Hines 2002. This means that the tax rate applicable to the most skilled individual is chosen not to redistribute income, but to support efficient behavior, since there is effectively no constraint on the choice of this tax rate. In the setting considered by Mirrlees and others, there is

only one decision margin, the labor/leisure choice, and efficiency therefore implies that the marginal income tax rate is zero for the most able individual.

In a setting with inter-temporal choice, however, there are multiple decision margins, and the choice of the marginal tax rate on the highest level of consumption carries implications for all of these, including not only the labor/leisure choice but also the choice between consumption in one period and another. It is clear that efficiency on the labor/leisure margin requires a zero marginal tax rate on consumption, as above. The difficulty with imposing a zero tax rate on the highest feasible level of consumption, however, is that this simultaneously distorts the inter-temporal allocation of consumption by the most able individual, if that individual's consumption in other periods is subject to positive consumption taxation. Consequently, the optimal tax configuration will imply positive, and not zero, taxation of the most able individual.

Consider the case in which utility is additively separable in consumption and leisure, and the utility function (counterfactually) exhibits no preference for consumption in one period over another.[6] Given this lack of preference, when confronted with a positive rate of interest, consumers will not smooth their consumption perfectly between the two periods, but instead consume less in the first period than in the second, in order to benefit from the ability to earn interest on consumption foregone in the first period. Denote by C_1^* consumption of the most able individual in the first period, and by C_2^* consumption by the same individual in the second period. It is clear from above that $C_2^* > C_1^*$, but it is also clear that, with a continuum of types, there will be many individuals other than the highest ability type whose second period consumption level also exceeds C_1^*. The desire to impose positive marginal taxes on

[6] This is akin to having a personal discount rate of 0, which is not generally observed.

consumption by these individuals entails subjecting the highest type to positive taxation on first period expenditures. This, in turn, implies that marginal tax rates on second period expenditures should be positive in order to avoid creating too great an inter-temporal distortion.

This simple example highlights two aspects of the general result. The first is that this argument for non-zero taxation of the highest consumption level in the life-cycle model requires that individuals choose not to smooth their consumption perfectly between periods. The second is that the optimal tax configuration entails positive marginal tax rates on at least some periods of consumption by the most able individual. These are in fact very mild requirements, since they are likely to be violated only in special circumstances, though it is very difficult to say much in general about the second requirement in particular. Certainly the distortion introduced by the taxation of consumption by the same individual in different periods will affect tax rates throughout the tax schedule, very likely dampening differences, but the quantitative features of optimal tax rates in such settings tend to be extremely sensitive to fine modeling details.

The general point can be made analytically, though in lieu of a complete characterization of the optimal tax problem, we proceed heuristically. Consider the choice of an efficient top marginal consumption tax rate, denoted, as in the example, $T'\left(C_2^*\right)$. (Recall that C_2^* is the highest consumption level any individual ever obtains.) If this tax rate is chosen optimally, then a small perturbation in the tax rate should have no effect on the welfare of the most able individual. For analytic simplicity, consider a perturbation that reduces leisure by 1/W, so lifetime consumption rises by one unit, and utility from leisure falls by λ. This change in the marginal tax rate on second period consumption, and accompanying endogenous change in labor supply, will affect consumption in the first and second period. Denote the change in second period consumption by

Δ, and assume (as is standard in this type of exercise) that any change in tax revenue collected from the highest type is recycled to that same individual, so the budget constraint implies that the change in first period consumption is given by $1-\frac{\Delta}{(1+r)}$.

The change in utility for the most able individual is then given by:

$$-\lambda+\left(1-\frac{\Delta}{(1+r)}\right)\lambda\left[1+T'\left(C_1^*\right)\right]+\Delta\frac{\lambda}{(1+r)}\left[1+T'\left(C_2^*\right)\right]$$

Setting this expression equal to zero, along with some simple algebraic manipulation, produces:

$$T'\left(C_2^*\right)=T'\left(C_1^*\right)\left[1-\frac{(1+r)}{\Delta}\right].$$

This equation expresses the marginal tax rate on second period consumption by the highest ability type as a function of the marginal tax rate on his or her first period consumption, and the change in second period consumption that would accompany a second period tax perturbation of sufficient magnitude to induce a change in labor supply sufficient to produce one unit of income. In attempting to interpret this expression, it is important to bear in mind that the tax system is taken to be optimal, so the utility level of the highest ability type does not change during the hypothetical perturbation. As a result, all changes in labor supply and consumption can be taken to be compensated changes. Consequently, first period consumption will fall, since the price of first period consumption relative to leisure does not change, whereas first period consumption becomes more expensive relative to second period consumption. (In order to induce greater labor

supply in a compensated setting, the perturbation must take the form that the marginal tax rate on second period consumption falls.) If first period consumption did not change at all, the unit rise in first period labor income would imply that $\Delta = (1+r)$. Since the substitution effect causes first period consumption to fall, it follows that $\Delta > (1+r)$, which in turn implies that $T'(C_2^*) > 0$. The optimal expenditure tax profile entails positive marginal tax rates on all levels of consumption.

As this exercise illustrates, the optimal marginal tax rate applying to the highest level of consumption lies between zero and the marginal tax rate on consumption by the most able type in other periods when consumption levels are lower. This result stems ultimately from the inability of a consumption tax levied on an annual basis to distinguish between high ability individuals in low consumption years and lower ability individuals in high consumption years. Obviously a consumption tax levied strictly on a lifetime basis, along the lines perhaps of the system outlined by Vickrey 1947, would not have this feature. In practice, however, such systems have never been used, nor do they seem likely to be adopted any time soon, and so we continue to assume, as we have above, an annual expenditure tax.

Note to readers: the remainder of this draft is especially preliminary and tentative; our apologies.

B. ***Precautionary Savings***

Another motive for savings is to guard against prospective emergencies, as in periods of low earnings or heightened medical needs. See for example Kotlikoff, Shoven and Spivak 1987. If the "precaution" is simply a hedge against periods of lower earnings, as for example insurance

against job loss, the analysis collapses into that above, because we have assumed no earnings in the second period. If the precautionary savings are meant as a reserve against other heightened expenditure needs, such as medical expenses, a critical issue here is the tax system's treatments of such expenditures. If we assume that medical expenses are deductible from the tax base, as under present law (Internal Revenue Code Section 213) the rational taxpayer can assume a low or zero rate of tax on precautionary savings. Thus the ex ante perceived tax rate on non-present period consumption will be lower, and will indeed approach zero, as under Mirrlees, but now without an explicitly declining marginal tax rate structure in labor market earnings.

C. *Bequest Savings*

Studies consistently show large amounts of savings are passed inter-generationally, in the form of bequests. Bernheim 2002, Kotlikoff and Summers 1981, Hurd 1987. The *fact* of bequest savings, however, does not fully explain the *motive* or *subjective intention* behind such savings. But because our project involves looking at the effect on present period labor supply effort from a deferred tax on savings, it is important to consider the precise motives for savings in greater detail. This subsection considers three distinct possibilities.

1. *Bequest Savings as Deferred Consumption by Heirs*

The possibility most immediately available from the label of "bequest savings" is that present period taxpayers are actively motivated to save by a desire to pass consumption opportunities along to their heirs. In such a case, C_2 will take place by a later generation. The question is what tax rate (as well as what tradeoff rate in own-utility) might a present period saver put on period 2 consumption by her heirs?

Note that both of the above possibilities—consumption smoothing, as in the life-cycle model, or precautionary savings—can apply to bequest savings. A rational saver might presume that her heirs will be at a lower level of consumption than she is in and thus, under a consistent progressive expenditure tax, will face a lower rate bracket. (Note that we do not consider the potentially important issue of labor market deterrent effects at the second and lower generation levels). There is at least anecdotal evidence to suggest that savers might prefer a tax or legal structure in which access to capital by their heirs is limited, by kind or amount. Many wealthy individuals in fact create "spendthrift," educational trusts, or even "dynasty trusts" for their heirs, limiting their access to capital to certain levels or specified reasons. This suggests that savers may not be motivated to enable high expenditures by heirs, so might not be deterred by high implicit tax rates on high expenditures.

Likewise, aside from the possibility, considered below, that bequests represent unused precautionary savings at the first generation—the rainy day never having come—the saver may be anticipating the precautionary needs of her heirs. Once again, the implicit tax rate considered by the first generation saver will be lower than her own consumption bracket, approaching zero.

2. *Bequest Savings as Deferred Precautionary Savings*

As noted, the rational saver might be anticipating that her heirs could face emergencies, other than low-earnings, and so may presume lower tax rates on period 2 consumption. Indeed, bequest savings might come from unused precautionary savings at the first generation, so to that extent problem is same as above.

3. *Bequest Savings as Residuum (Consumption Value of Capital)*

Finally, we consider the possibility that savings result from the first generation saver's own utility from capital: that is, that taxpayers are motivated to build up large stores of capital for their own sake, and prefer having them to own-consumption. In such a case, the taxpayer today faces a perceived subjective tax rate of zero on continued savings, again paralleling Mirrlees without an explicitly declining marginal rate structure.

A similar result obtains—also similar to the precautionary savings discussion noted above—from an ultimate philanthropic intent, or a mixed motivation in which charities to whom payments are tax deductible are the residuum. The rational worker in a prior period, saving in anticipation of one day potentially making a large philanthropic gift, will face an implicit tax rate of 0 on her additional work effort today.

5. Behavioral Economics Models

In the analysis above, although we have considered different possible motives for savings, we assumed that taxpayers are "rational," in the standard economics sense of individuals acting consistently on the basis of a well-behaved utility function. But many models of savings presume that individuals do not so behave—that they follow a "behavioral" life cycle model in which mental accounts, myopia, and other forms of time-inconsistent preferences matter. Thaler and Sheffrin 1981, Sheffrin and Thaler 1988, Thaler 1994. In this Section, we offer some comments on optimal consumption tax rates under a behavioral model.

[Tentative musings: One possibility is that is that people are myopic, in which case they do not save at all, and we wonder if we can even say that a tax system that taxes them and then distributes value back in kind, in the future, in a way that they cannot arbitrage today, can be

optimal. Of course this is the kind of paternalism that makes behavioral models suspect. Perhaps we can also make the point here, that if people are myopic, you do not want an income tax, because of the wedges, etc., but you also don't want a "hybrid" that is basically an income tax with cash-flow treatment reserved just for savings, because then the truly myopic borrows to save today, getting more present period consumption via the tax break, and even less savings. See http://ssrn.com/abstract=899302 . So a consistent progressive expenditure tax is an attractive option. And then, one thing we can say tying in above analysis, is that you don't ever want the rapidly declining rates, or you pour fuel on the fire, and make the irrational myopic a rational myopic.]

6. Empirical Observations

Designing an optimal rate schedule for an expenditure tax requires having a good sense of why people save. As we have noted above, however, pinning down the precise motives for savings is a difficult task, made harder by the fact that there is no ground rule constraining all people to be the same. The motives for savings are almost certainly mixed, within and across individuals. We suspect that much savings by the "upper middle" class is indeed life-cycle, with perhaps a mixed of precautionary savings thrown in, and the fact of any bequests simply a residuum. There are also no doubt many myopic, behavioral life-cycle actors. Among the very wealthy, the fact that most savers continue to save suggests either a bequest motive or a utility from owning capital, or both.

Our principal objective in the preceding analysis has been to show that under all or most plausible reasons for savings, the Mirrlees results of a declining marginal tax rate on the upper-income, reaching zero, do not hold under a consistent expenditure tax.

7. Conclusions

[and suggestions for further research, etc.]

8. References

Andrews, William D. 1974. A Consumption-Type or Cash Flow Personal Income Tax. *Harvard Law Review* 87: 1113-88.

Auerbach, Alan C. forthcoming. Fundamental Tax Reform in the 21st Century. In John W. Diamond and George R. Zodrow, editors, *Fundamental Tax Reform: Issues, Choices and Implications* (Cambridge, Mass; MIT Press, forthcoming 2008).

Auerbach, Alan C. and James R. Hines. 2002. Efficiency and Taxation. In *Handbook of Public Economics,* Volume 3, eds. Alan Auerbach and Martin Feldstein, (New York: Elsevier Science Pub. Co., 2002), 1347-1421.

Bankman, Joseph, and Thomas Griffith. 1987. Social Welfare and the Rate Structure: A New Look at Progressive Taxation. *California Law Review* 75:1905-67.

Bernheim, B. Douglas. 1997. Rethinking Saving Incentives. In *Fiscal Policy: Lessons from Economic Research,* ed. Alan Auerbach (Cambridge, MA: MIT Press, 1997), 259-311.

———. 2002. Taxation and Saving. In *Handbook of Public Economics,* Volume 3, eds. Alan Auerbach and Martin Feldstein, (New York: Elsevier Science Pub. Co., 2002), 1174-1249.

Bradford, David F. 1986. *Untangling the Income Tax*. Cambridge, MA: Harvard Univ. Press.

———. 1996. *Fundamental Issues in Consumption Taxation*. Washington, DC: AEI Press.

Bradford, David F. et al. 1977. *Blueprints for Basic Tax Reform*. Washington, DC: US Department of Treasury.

Correia, Isabel H. 1996. Should Capital Income Be Taxed in the Steady State? *Journal of Public Economics 60*, 147-151.

Congressional Budget Office. 2005. Effective Marginl Rates on Labor Income. Available online at www.cbo.gov/ftpdoc.cfm?index=6854&type=1

Diamond, Peter A., and James A. Mirrlees. 1971. Optimal Taxation and Public Production I: Production Efficiency. *American Economic Review* 61(1):8-27.

Feldstein, Martin. 2006. Capital Income Taxation. Economist's Voice.

Friedman, Milton. 1957. *A Theory of the Consumption Function.* Princeton, NJ: Princeton University Press.

Golosov, Mikhail et al. 2003. Optimal Indirect and Capital Taxation. *Review of Economic Studies 70*, 569-587.

Golosov, Mikhail and Aleh Tsyvinski. 2005. Optimal Taxation with Endogenous Insurance Markets. NBER Working Paper No. 11185.

Graetz, Michael J. 2007. Tax Reform Unraveling. *Journal of Economic Perspectives 21*, 69-90.

Hines, James R., Jr. 2007. Taxing Consumption and Other Sins. *Journal of Economic Perspectives 21*, 49-68.

Hurd, Michael . 1987. Savings of the Elderly and Desired Bequests. *American Economic Review. 77,* 298-312.

Kaldor, Nicholas. 1955. *An Expenditure Tax.* (London: Allen & Unwin).

Kawachi, Jannette, Karen E. Smith, and Eric Toder. 2006. Making Maximum Use of Tax-Deferred Retirement Accounts. Washington, DC: The Urban Institute.

Kotlikioff, Lawrence J. 1989. *What Determines Savings?* (Cambridge, Mass, MIT Press).

Kotlikoff, Lawrence J., John Shoven and Avia Spivak. 1987. Annuity Insurance, Savings, and Inquality. In Kotlikoff 1989.

Kotlikoff, Lawrence J., and Avia Spivak. 1981. The Family as an Incomplete Annuities Market. *Journal of Political Economy* 89:372-91.

Kotlikoff, Lawrence J., and Lawrence Summers. 1981. The Role of Intergenerational Transfers in Aggregate Capital accumulation. *Journal of Political Economy* 89:706-32.

McCaffery, Edward J. 2002. Fair Not Flat: How to Make the Tax system Better and Simpler. (Chicago: Universiity of Chicago Press).

_____. 2005a. A New Understanding of Tax. *Michigan Law Review* 103:807-938.

———. 2005b. Three Views of Tax. *Canadian Journal of Law & Jurisprudence* 18: 153-64.

_____. forthcoming. "Behavioral Economics and Fundamental Tax Reform," in John W. Diamond and George R. Zodrow, eds, Fundamental Tax Reform: Issues, Choices and Implications (MIT Press, 2008), draft version available at http://ssrn.com/abstract=899302.

Mill, John Stuart. 1848. *Principles of Political Economy*. edited by Jonathan Riley. New York: Oxford University Press (1994).

Mirrlees, James. 1971. An Exploration in the Theory of Optimum Taxes. *Review of Economic Studies* 38: 175-208.

Modigliani, Francisco. 1988. The Role of Intergenerational Transfers and Life Cycle Saving in the Accumulation of Wealth. *Journal of Economic Perspectives* 2:15-40.

Modigliani, Francisco, and R. Brumberg. 1954. Utility Analysis and the Consumption Function Function: An Interpretation of Cross-Section Data. In *Post-Keynesian Economics*, ed. K.K. Kurihara (New Brunswick, NJ: Rutgers University Press, 1954), 388-436.

Moomau, Pamela H. forthcoming. Comments on Edward J. McCaffery, Behavioral Economics and Fundamental Tax Reform, in John W. Diamond and George R. Zodrow, editors, *Fundamental Tax Reform: Issues, Choices, and Implications*, (Cambridge, Mass; MIT Press, forthcoming 2008).

Ramsey, F.P. 1927. A Contribution to the Theory of Taxation. *Economic Journal 37*, 47-61.

Scholz, John Karl, Ananth Seshadri and Surachai Khitatrakun. 2006. Are Americans Saving 'Optimally' for Retirement? *Journal of Political Economy* (forthcoming).

Seidman, Laurence. 1995. The USA Tax.

Shaviro, Daniel. 2007. Beyond the Pro-Consumption Tax Consensus. Working paper, NYU, available on-line at http://ssrn.com/abstract=955351

Shefrin, Hersh M., and Richard H. Thaler. 1988. The Behavioral Life-Cycle Hypothesis. *Economic Inquiry* 26(4):609-43.

Simons, Henry C. 1938. *Personal Income Taxation: The Definition of Income as a Problem of Fiscal Policy*. Chicago: The University of Chicago Press.

Steuerle, Gene. 2005. *Contemporary US Tax Policy*. Washington, DC: The Urban Institute Press.

Stiglitz, Joseph E. 1982. Self-Selection and Pareto Efficient Taxation. *Journal of Public Economics 17*, 213-240.

Stiglitz, Joseph E. 1987. Pareto Efficient and Optimal Taxation and the New New Welfare Economics. In *Handbook of Public Economics,* Volume 2, eds. Alan Auerbach and Martin Feldstein, (New York: Elsevier Science Pub. Co.), 991-1042.

Thaler, Richard H. 1994. Psychology and Savings Policies. *American Economic Review* 84:186-92.

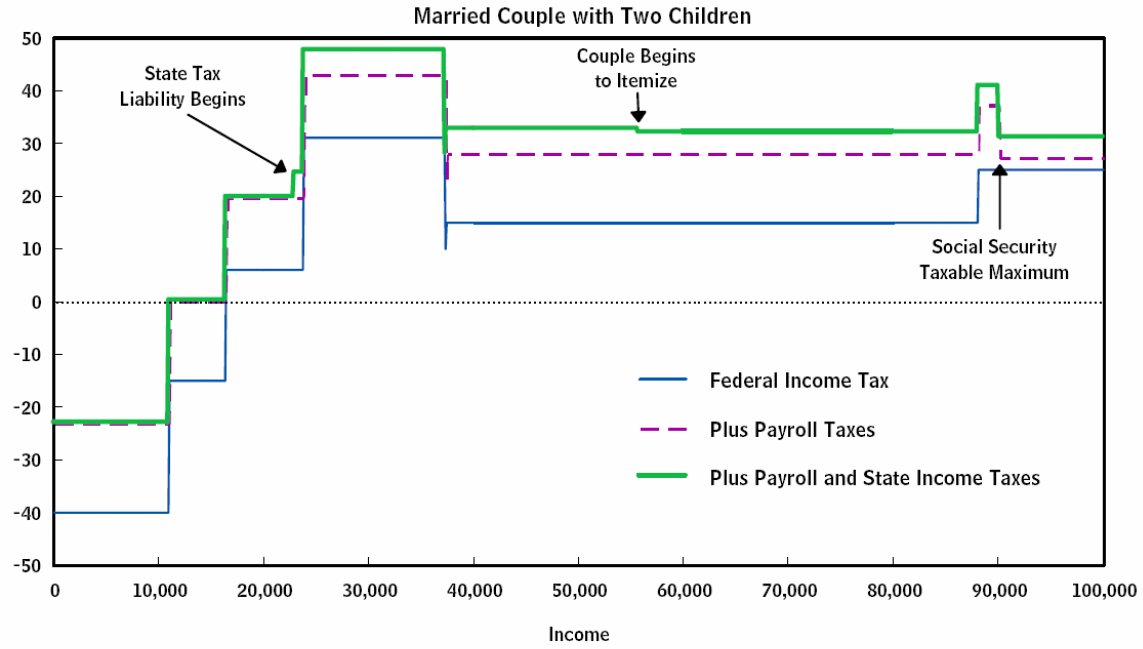
Thaler, Richard H., and Hersh M. Shefrin. 1981. An Economic Theory of Self-Control. *Journal of Political Economy* 89(2):392-406.

Tuomala, Matti. 1990. *Optimal Income Taxation and Redistribution.* Oxford, UK: Oxford University Press.

Vickrey, William. 1947. *Agenda for Progressive Taxation.* New York: Ronald Press.

Weisbach, David A. 2006. The Case for a Consumption Tax. *Tax Notes* 110:1357-59.

(Tax rate in percent)



Source: Congressional Budget Office.

Notes: These examples assume that the taxpayers are a single filer with no dependents, a single filer who has one dependent and files as a head of household, or a married couple filing jointly with two dependents. All income is from wages (in the case of the married couple, those wages are earned by one spouse). The taxpayers have itemized deductions worth 18 percent of their income and claim the greater of those deductions or the standard deduction. (Forty percent of the itemized deductions are assumed to be state and local taxes, and the rest are charitable contributions and mortgage interest.)

State taxes are assumed to be 5 percent of federal taxable income.

Marginal rates are computed as a percentage of compensation before the employer's share of payroll taxes has been paid.

Figure 1: Effective Marginal Tax Rates on Labor Income (CBO 205)