5

New Dynamic Public Finance: A User's Guide

Mikhail Golosov, *MIT and NBER*
Aleh Tsyvinski, *Harvard University and NBER*
Iván Werning, *MIT and NBER*

1 Introduction

New Dynamic Public Finance is a recent literature that extends the static Mirrlees [1971] framework to dynamic settings.[1] The approach addresses a broader set of issues in optimal policy than its static counterpart, while not relying on exogenously specified tax instruments as in the representative-agent Ramsey approach often used in macroeconomics.

In this paper we show that this alternative approach can be used to revisit three issues that have been extensively explored within representative-agent Ramsey setups. We show that this alternative approach delivers insights and results that contrast with those from the Ramsey approach. First, it is optimal to introduce a positive distortion in savings that implicitly discourages savings (Diamond and Mirrlees 1978, Rogerson 1985, Golosov, Kocherlakota, and Tsyvinski 2003). This contrasts with the Chamley-Judd (Judd 1985, Chamley 1986) result, obtained in Ramsey settings, that capital should go untaxed in the long run.[2] Second, when workers' skills evolve stochastically due to shocks that are not publicly observable, their labor income tax rates are affected by aggregate shocks: Perfect tax smoothing, as in Ramsey models (Barro 1979, Lucas and Stokey 1983, Judd 1989, Kingston 1991, Zhu 1992, Chari, Christiano, and Kehoe 1994), may not be optimal with uncertain and evolving skills.[3] In contrast, it is optimal to smooth labor distortions when skills are heterogenous but constant or affected by shocks that are publicly observable (Werning 2007). Finally, the nature of the time-consistency problem is very different from that arising within Ramsey setups. The problem is, essentially, about learning and using acquired information, rather than taxing sunk capital: A benevolent government is tempted to exploit information collected in the past. Indeed, capital is not directly at the root of the problem, in that even if the government controlled all capital accumulation in the economy—or in an economy without capital—a time-consistency problem arises.

1.1 User's Guide

We call this paper "a user's guide" because our main goal is to provide the reader with an overview of three implications of the dynamic Mirrlees literature that differ from those of Ramsey's. Our workhorse model is a two-period economy that allows for aggregate uncertainty regarding government purchases or rates of returns on savings, as well as idiosyncratic uncertainty regarding workers' productivity. The model is flexible enough to illustrate some key results in the literature. Moreover, its tractability allows us to explore some new issues. We aim to comprehensively explore the structure of distortions and its dependence on parameters within our dynamic Mirrleesian economy. Papers by Albanesi and Sleet (2006), Golosov and Tsyvinski (2006a) and Kocherlakota (2005) include some similar exercises, but our simple model allows us to undertake a more comprehensive exploration.[4] Although some of our analysis is based on numerical simulations, our focus is qualitative: We do not seek definitive quantitative answers from our numerical exercises, rather our goal is to illustrate qualitative features and provide a feel for their quantitative importance.

The presence of private information regarding skills and the stochastic evolution of skills introduces distortions in the marginal decisions of agents. We focus attention on two such wedges. The first wedge is a consumption-labor wedge (or, simply, a labor wedge) that measures the difference between the marginal rate of substitution and transformation between consumption and labor. The second wedge is the intertemporal (or capital) wedge, defined as the difference between the expected marginal rate of substitution of consumption between periods and the return on savings. In this paper, our focus is distinctively on these wedges—which are sometimes termed "implicit marginal tax rates"—rather than on explicit tax systems that implement them. However, we do devote a section to discussing the latter.

1.2 Ramsey and Mirrlees Approaches

The representative-agent Ramsey model has been extensively used by macroeconomists to study optimal policy problems in dynamic settings.[5] Examples of particular interest to macroeconomists include: the smoothing of taxes and debt management over the business cycle, the taxation of capital in the long run, monetary policy, and a variety of time inconsistency problems.

This approach studies the problem of choosing taxes within a given set of available tax instruments. Usually, to avoid the first-best, it is assumed that taxation must be proportional. Lump-sum taxation, in particular, is prohibited. A benevolent government then sets taxes to finance its expenditures and maximize the representative agent's utility. If, instead, lump-sum taxes were allowed, then the unconstrained first-best optimum would be achieved. One criticism of the Ramsey approach is that the main goal of the government is to mimic lump-sum taxes with an imperfect set of instruments. However, very little is usually said about why tax instruments are restricted or why they take a particular form. Thus, as has been previously recognized, the representative-agent Ramsey model does not provide a theoretical foundation for distortionary taxation. Distortions are simply assumed and their overall level is largely determined exogenously by the need to finance some given level of government spending.

The Mirrlees approach to optimal taxation is built on a different foundation. Rather than starting with an exogenously restricted set of tax instruments, Mirrlees's (1971) starting point is an informational friction that endogenizes the feasible tax instruments. The crucial ingredient is to model workers as heterogenous with respect to their skills or productivity. Importantly, workers' skills and work effort are not directly observed by the government. This private information creates a tradeoff between insurance (or redistribution) and incentives. Even when tax instruments are not constrained, distortions arise from the solution to the planning problem.

Since tax instruments are not restricted, without heterogeneity the first-best would be attainable. That is, if everyone shared the same skill level then a simple lump-sum tax—that is, an income tax with no slope—could be optimally imposed. The planning problem is then equivalent to the first-best problem of maximizing utility subject only to the economy's resource constraints. This extreme case emphasizes the more general point that a key determinant of distortions is the desire to redistribute or insure workers with respect to their skills. As a result, the level of taxation is affected by the distribution of skills and risk aversion, among other things.

1.3 Numerical Results

We now summarize the main findings from our numerical simulations. We begin with the case without aggregate uncertainty.

We found that the main determinants for the size of the labor wedge are agents' skills, the probability with which skill shocks occurs, risk aversion, and the elasticity of labor supply. Specifically, we found that the labor wedges in the first period, or for those in the second period not suffering the adverse shock, are largely unaffected by the size or probability of the adverse shock; these parameters affect these agents only indirectly through the ex-ante incentive compatibility constraints. Higher risk aversion leads to higher labor wedges because it creates a higher desire to redistribute or insure agents. As for the elasticity of labor supply, we find two opposing effects on the labor wedge: A lower elasticity leads to smaller welfare losses from redistribution but also leads to less pre-tax income inequality, for a given distribution of skills, making redistribution less desirable.

Turning to the capital wedge, we find that two key determinants for its size are the size of the adverse future shock and its probability. A higher elasticity of labor may decrease the savings wedge if it decreases the desire to redistribute. More significantly, we derive some novel predictions for capital wedges when preferences over consumption and labor are nonseparable. The theoretical results in dynamic Mirrleesian models have been derived by assuming additively-separable utility between consumption and labor. In particular, the derivation of the Inverse Euler optimality condition, which ensures a positive capital wedge, relies on this separability assumption. Little is known about the solution of the optimal problem when preferences are not separable. Here we partially fill this gap with our numerical explorations. The main finding of the model with a nonseparable utility function is that the capital wedge may be negative. We show that the sign of the wedge depends on whether consumption and labor are complements or substitutes in the utility function, as well as on whether skills are expected to trend up or down.

We now describe the cases with aggregate uncertainty. Most of our numerical findings are novel here, since aggregate shocks have remained almost unexplored within the Mirrleesian approach.[6]

When it comes to aggregate shocks, an important insight from representative-agent Ramsey models is that tax rates on labor income should be smoothed across time (Barro 1979) and aggregate states of nature

(Lucas and Stokey 1983).[7] As shown by Werning (2007), this notion does not depend on the representative-agent assumption, as it extends to economies with heterogenous agents subject to linear or nonlinear taxation. Thus, in our setup perfect tax smoothing obtains as long as all idiosyncratic uncertainty regarding skills is resolved in the first period.

In our numerical exercises we also consider the case where idiosyncratic uncertainty persists into the second period. We find that labor wedges then vary across aggregate shocks. Thus, perfect tax smoothing—where the wedges for each skill type are perfectly invariant to aggregate states—does not hold. Tax rates vary because individual skill shocks and aggregate shocks are linked through the incentive constraints. Interestingly, aggregate shocks do not increase or decrease tax rates uniformly. In particular, we find that a positive aggregate shock (from a higher return on savings or a lower government expenditure) lowers the spread between labor wedges across skill types in the second period.

2 An Overview of the Literature

The dynamic Mirrleesian literature builds on the seminal work by Mirrlees (1971), Diamond and Mirrlees (1978), Atkinson and Stiglitz (1976) and Stiglitz (1987).[8,9] These authors laid down the foundation for analyzing optimal non-linear taxation with heterogeneous agents and private information. Many of the more recent results build on the insights first developed in those papers. The New Dynamic Public Finance literature extends previous models by focusing on the stochastic evolution of skills and aggregate shocks. Thus, relative to the representative agent Ramsey approach, commonly pursued by macroeconomists, it places greater emphasis on individual heterogeneity and uncertainty; whereas, relative to traditional work in public finance it places uncertainty, at the aggregate and individual level, at the forefront of the analysis.

Werning (2002) and Golosov, Kocherlakota, and Tsyvinski (2003) incorporated Mirrleesian framework into the standard neoclassical growth model. Werning (2002) derived the conditions for the optimality of smoothing labor income taxes over time and across states. Building on the work of Diamond and Mirrlees (1978) and Rogerson (1985), Golosov et al. (2003) showed that it is optimal to distort savings in a general class of economies where skills of agents evolve stochastically over time. Kocherlakota (2005) extended this result to an economy with aggregate shocks. We discuss these results in section 4. Werning (2002), Shimer and Werning (2005), and Abraham and Pavoni (2003) study optimal taxation when capital is not observable and its rate of return is not taxed. da Costa and Werning (2002), Golosov and Tsyvinski (2006b), and da Costa (2005) consider economies where individual borrowing and lending are not observable so that non-linear distortions of savings are not feasible, but the government may still uniformly influence the rate of return by taxing the observable capital stock.

Unlike the taxation of savings, less work has been done in studying optimal labor wedges in the presence of stochastic skills shocks. Battaglini and Coate (2005) show that if the utility of consumption is linear, labor taxes of all agents asymptotically converge to zero. Risk neutrality, however, is crucial to this result. Section 5 of this paper explores dynamic behavior of labor wedges for risk averse agents in our two-period economy.

Due to space constraints we limit our analysis in the main body of the paper only to capital and labor taxation. At this point we briefly mention recent work on other aspects of tax policy. Farhi and Werning (2007) analyze estate taxation in a dynastic model with dynamic private information. They show that estate taxes should be progressive: Richer parents should face a higher marginal tax rate on bequests. This result is a consequence of the optimality of mean reversion in consumption across generations, which tempers the intergenerational transmission of welfare. Rich parents must face lower net rates of return on their transfers so that they revert downward towards the mean, while poor parents require the opposite to revert upwards. Albanesi (2006) considers optimal taxation of entrepreneurs. In her setup an entrepreneur exerts unobservable effort that affects the rate of return of the project. She shows that the optimal intertemporal wedge for the entrepreneurs can be either positive or negative. da Costa and Werning (2005) study a monetary model with heterogeneous agents with privately observed skills, where they prove the optimality of the Friedman rule, that the optimal inflationary tax is zero.

The analysis of optimal taxation in response to aggregate shocks has traditionally been studied in the macro-oriented Ramsey literature. Werning (2002, 2007) reevaluated the results on tax smoothing in a model with private information regarding heterogeneous skills. In his setup, all idiosyncratic uncertainty after the initial period is due to unobservable shock. In Section 6, for the two period economy introduced in this paper, we explore the extent of tax smoothing in response to aggregate shocks when unobservable idiosyncratic shocks are also present in the second period.

Some papers, for example Albanesi and Sleet (2006), Kocherlakota (2005), and Golosov and Tsyvinski (2006a), consider implementing optimal allocations by the government using tax policy. Those analyses assume that no private markets exist to insure idiosyncratic risks and agents are able to smooth consumption over time by saving at a market interest rate. Prescott and Townsend (1984) show that the first welfare theorem holds in economies with unrestricted private markets and the efficient wedges can be implemented privately without any government intervention. When markets are very efficient, distortionary taxes are redundant. However, if some of the financial transactions are not observable, the competitive equilibrium is no longer constrained efficient. Applying this insight, Golosov and Tsyvinski (2006b) and Albanesi (2006) explore the implications of unobservability in financial markets on optimal tax interventions. We discuss some of these issues in section 4.

In step with theoretical advances, several authors have carried out quantitative analyses of the size of the distortion and welfare gains from improving tax policy. For example, Albanesi and Sleet (2006) study the size of the capital and labor wedges in a dynamic economy. However they are able to conduct their analyses only for the illustrative case of *i.i.d.* shocks to skills. Moving to the other side of the spectrum, with permanent disability shocks, Golosov and Tsyvinski (2006a) show that the welfare gains from improving disability insurance system might be large. Recent work by Farhi and Werning (2006a) develops a general method for computing the welfare gains from partial reforms, starting from any initial incentive compatible allocations with flexible skill processes, that introduce optimal savings distortions.

All the papers discussed above assume that the government has full commitment power. The more information is revealed by agents about their types, the stronger is the incentive of the government to deviate from the originally promised tax sequences. This motivated several authors to study optimal taxation in environments where the government cannot commit. Optimal taxation without commitment is technically a much more challenging problem since the simplest versions of the Revelation Principle do not hold in such an environment. One of the early contributors was Roberts (1984) who studies an economy where individuals have constant skills which are private information. Bisin and Rampini (2006) study a two period version of this problem. Sleet and Yeltekin (2005) and Acemoglu, Golosov, and Tsyvinski (2006) show conditions under which even the simplest versions of the Revelation Principle can be applied along the equilibrium path. We discuss these issues in section 4.

3 A Two-Period Mirrleesian Economy

In this section we introduce a two-period Mirrleesian economy with uncertainty.

3.1 Preferences

There is a continuum of workers that are alive in both periods and maximize their expected utility

$$\mathbb{E}[u(c_1) + v(n_1) + \beta(u(c_2) + v(n_2))],$$

where c_t represents consumption and n_t is a measure of work effort.

With two periods, the most relevant interpretation of our model is that the first period represents relatively young workers, say those aged 20–45, while the second period represents relatively older workers and retired individuals, say, those older than 45.[10]

3.2 Skills

Following Mirrlees (1971), workers are, at any time, heterogenous with respect to their skills, and these skills are privately observed by workers. The output y produced by a worker with skill θ and work effort n is given by the product, effective labor: $y = \theta n$. The distribution of skills is independent across workers.

For computational reasons, we work with a finite number of skill types in both periods. Let the skill realizations for the first period be $\theta_1(i)$ for $i = 1, 2, \ldots, N_1$ and denote by $\pi_1(i)$ their ex ante probability distribution, equivalent to the ex post distribution in the population. In the second period the skill becomes $\theta_2(i, j)$ for $j = 1, 2, \ldots, N_2(i)$ where $\pi_2(j \mid i)$ is the conditional probability distribution for skill type j in the second period, given skill type i in the first period.

3.3 Technology

We assume production is linear in efficiency units of labor supplied by workers. In addition, there is a linear savings technology.

We consider two types of shocks in the second period: (1) a shock to the rate of return; and (2) a shock to government expenditures in the second period. To capture both shocks we introduce a state of the world $s \in S$, where S is some finite set, which is realized at the beginning of period $t = 2$. The rate of return and government expenditure in the second period are functions of s. The probability of state s is denoted by $\mu(s)$.

The resource constraints are

$$\sum_i (c_1(i) - y_1(i))\pi_1(i) + K_2 \leq R_1 K_1 - G_1, \tag{1}$$

$$\sum_{i,j} (c_2(i,j) - y_2(i,j))\pi_2(j \mid i)\pi_1(i) \leq R_2(s)K_2 - G_2(s), \quad \text{for all } s \in S, \tag{2}$$

where K_2 is capital saved between periods $t = 1$ and $t = 2$, and K_1 is the endowed level of capital.

An important special case is one without aggregate shocks. In that case we can collapse both resource constraints into a single present value condition by solving out for K_2:

$$\sum_i \left(c_i(i) - y_1(i) + \frac{1}{R_2} \sum_j [c_2(i,j) - y_2(i,j)]\pi_2(i,j) \right) \pi_1(i) \leq R_1 K_1 - G_1 - \frac{1}{R_2} G_2. \tag{3}$$

3.4 Planning Problem

Our goal is to characterize the optimal tax policy without imposing any *ad hoc* restrictions on the tax instruments available to a government. The only constraints on taxes arise endogenously because of the informational frictions. It is convenient to carry out our analysis in two steps. First, we describe how to find the allocations that maximize social welfare function subject to the informational constraints. Then, we discuss how to find taxes that in competitive equilibrium lead to socially efficient allocations. Since we do not impose any restrictions on taxes a priori, the tax instruments available to the government may be quite rich. The next section describes features that such a system must have.

To find the allocations that maximize social welfare it is useful to think about a fictitious social planner who collects reports from the workers about their skills and allocates consumption and labor according to those reports. Workers make skill reports i_r and j_r to the planner in the first and second period, respectively. Given each skill type i, a reporting strategy is a choice of a first-period report i_r and a plan for the second period report $j_r(j, s)$ as a function of the true skill realization j and the aggregate shock. Since skills are private information, the allocations must be such that no worker has an incentive to misreport his type. Thus, the allocations must satisfy the following incentive constraint

$$u(c_1(i))+v\left(\frac{y_1(i)}{\theta_1(i)}\right)+\beta\sum_{s,j}\left[u(c_2(i,j,s))+v\left(\frac{y_2(i,j,s)}{\theta_2(i,j)}\right)\right]\pi_2(i\mid j)\mu(s) \tag{4}$$

$$\geq u(c_1(i_r))+v\left(\frac{y_1(i_r)}{\theta_1(i)}\right)$$

$$+\beta\sum_{s,j}\left[u(c_2(i_r,j_r(j,s),s))+v\left(\frac{y_2(i_r,j_r(j,s),s)}{\theta_2(i,j)}\right)\right]\pi_2(j\mid i)\mu(s),$$

for all alternative feasible reporting strategies i_r and $j_r(j, s)$.[11]

In our applications we will concentrate on maximizing a utilitarian social welfare function.[12] The *constrained efficient planning problem* maximizes expected discounted utility

$$\sum_i\left[u(c_1(i))+v\left(\frac{y_1(i)}{\theta_1(i)}\right)+\beta\sum_{s,j}\left[u(c_2(i,j,s))+v\left(\frac{y_2(i,j,s)}{\theta_2(i,j)}\right)\right]\pi_2(j\mid i)\mu(s)\right]\pi_1(i),$$

subject to the resource constraints in (1) and (2) and the incentive constraints in (4). Let (c^*, y^*, k^*) denote the solution to this problem. To understand the implications of these allocations for the optimal tax policy, it is important to focus on three key relationships or wedges between marginal rates of substitution and technological rates of transformation:

The consumption-labor wedge (distortion) in $t = 1$ for type i is

$$\tau_{y_1}(i)\equiv 1+\frac{v'(y_1^*(i)/\theta_1(i))}{u'(c_1^*(i))\theta_1(i)}, \tag{5}$$

The consumption-labor wedge (distortion) at $t = 2$ for type (i, j) in state s is

$$\tau_{y_2}(i,j,s)\equiv 1+\frac{v'(y_2^*(i,j,s)/\theta_2(i,j))}{u'(c_1^*(i,j,s))\theta_2(i,j)}, \tag{6}$$

The intertemporal wedge for type i is

$$\tau_k(i)\equiv 1-\frac{u'(c_1^*(i))}{\beta\sum_{s,j}R_2(s)u'(c_2^*(i,j,s))\pi_2(j\mid i)\mu(s)}. \tag{7}$$

Note that in the absence of government interventions all the wedges are equal to zero.

4 Theoretical Results and Discussion

In this section we review some aspects of the solution to the planning problem that can be derived theoretically. In the next sections we illustrate these features in our numerical explorations.

4.1 Capital Wedges

We now characterize the intertemporal distortion, or implicit tax on capital. We first work with an important benchmark in which there are no skill shocks in the second period. That is, all idiosyncratic uncertainty is resolved in the first period. For this case we recover Atkinson and Stiglitz 's (1976) classical uniform taxation result, implying no intertemporal consumption distortion: Capital should not be taxed. Then, with shocks in the second period we obtain an Inverse Euler Equation, which implies a positive intertemporal wedge (Diamond and Mirrlees 1978, Golosov, Kocherlakota, and Tsyvinski 2003).

4.1.1 Benchmark: Constant Types and a Zero Capital Wedge In this section, we consider a benchmark case in which the skills of agents are fixed over time and there is no aggregate uncertainty. Specifically, assume that $N_2(i) = 1$, $\forall i$, and that $\theta_1(i) = \theta_2(i, j) = \theta(i)$. In this case the constrained efficient problem simplifies to:

$$\max \sum_i \left[u(c_1(i)) + v\left(\frac{y_1(i)}{\theta(i)}\right) + u(c_2(i)) + v\left(\frac{y_2(i)}{\theta(i)}\right)\right]\pi_1(i)$$

subject to the incentive compatibility constraint that $\forall i \in \{1, ..., N_1\}$, and $i_r \in \{1, ..., N_1\}$:

$$u(c_1(i)) + v\left(\frac{y_1(i)}{\theta(i)}\right) + \beta\left[u(c_2(i)) + v\left(\frac{y_2(i)}{\theta(i)}\right)\right] \geq u(c_1(i_r))$$
$$+ v\left(\frac{y_1(i_r)}{\theta(i)}\right) + \beta\left[u(c_2(i_r)) + v\left(\frac{y_2(i_r)}{\theta(i)}\right)\right],$$

and subject to the feasibility constraint,

$$\sum_i \left[c_1(i) - y_1(i) + \frac{1}{R_2}(c_2(i) - y_2(i)) \right] \pi_1(i) \le R_1 k_1 - G_1 - \frac{1}{R_2} G_2.$$

We can now prove a variant of a classic Atkinson and Stiglitz (1976) uniform commodity taxation theorem which states that the marginal rate of substitution should be equated across goods and equated to the marginal rate of transformation.

To see this note that only the value of total utility from consumption $u(c_1) + \beta u(c_2)$ enters the objective and incentive constraints. It follows that for any total utility coming from consumption $u(c_1(i)) + \beta u(c_2(i))$ it must be that resources $c_1(i) + (1/R_2)c_2(i)$ are minimized, since the resource constraint cannot be slack. The next proposition then follows immediately.

Proposition 1 *Assume that the types of agents are constant. A constrained efficient allocation satisfies*

$$u'(c_1(i)) = \beta R_2 u'(c_2(i)) \qquad \forall i$$

Note that if $\beta = R_2$ then $c_1(i) = c_2(i)$. Indeed, in this case the optimal allocation is simply a repetition of the optimal one in a static version of the model.

4.1.2 Inverse Euler Equation and Positive Capital Taxation We now return to the general case with stochastic types and derive a necessary condition for optimality: The Inverse Euler Equation. This optimality condition implies a positive marginal intertemporal wedge.

We consider variations around any incentive compatible allocation. The argument is similar to the one we used to derive Atkinson and Stiglitz's (1976) result. In particular, it shares the property that for any realization of i in the first period we shall minimize the resource cost of delivering the remaining utility from consumption.

Fix any first period realization i. We then increase second period utility $u(c_2(i, j))$ in a parallel way across second period realizations j. That is define $u(\tilde{c}_2(i, j; \Delta)) \equiv u(c_2(i, j)) + \Delta$ for some small Δ. To compensate, we decrease utility in the first period by $\beta\Delta$. That is, define $u(\tilde{c}_1(i; \Delta)) \equiv u(c_1(i)) - \beta\Delta$ for small Δ.

The crucial point is that such variations do not affect the objective function and incentive constraints in the planning problem. Only the resource constraint is affected. Hence, for the original allocation to be optimal it must be that $\Delta = 0$ minimizes the resources expended

$$\tilde{c}_1(i;\Delta)+\frac{1}{R_2}\sum_j \tilde{c}_2(i,j;\Delta)\pi(j\,|\,i)$$
$$=u^{-1}(u(c_1(i))-\beta\Delta)+\frac{1}{R_2}\sum_j u^{-1}(u(c_2(i,j))+\Delta)\pi(j\,|\,i)$$

for all i. The first order condition for this problem evaluated at $\Delta = 0$ then yields the Inverse Euler equation summarized in the next proposition, due originally to Diamond and Mirrlees (1978) and extended to an arbitrary process for skill shocks by Golosov, Kocherlakota, and Tsyvinski (2003).

Proposition 2 *A constrained efficient allocation satisfies an Inverse Euler Equation:*

$$\frac{1}{u'(c_1(i))}=\frac{1}{\beta R_2}\sum_j \frac{1}{u'(c_2(i,j))}\pi_2(j\,|\,i). \tag{8}$$

If there is no uncertainty in second period consumption, given the first period shock, the condition becomes

$$\frac{1}{u'(c_1)}=\frac{1}{\beta R_2}\frac{1}{u'(\bar{c}_2)} \quad \Rightarrow \quad u'(c_1)=\beta R_2 u'(\bar{c}_2), \tag{9}$$

which is the standard Euler equation that must hold for a consumer who optimizes savings without distortions.

Whenever consumption remains stochastic, the standard Euler equation must be distorted. This result follows directly by applying Jensen's inequality to the reciprocal function "$1/x$" in equation (8).[13]

Proposition 3 *Suppose that for some i, there exists j such that $0 < \pi(j\,|\,i) < 1$ and that $c_2(i, j)$ is not independent of j. Then the constrained efficient allocation satisfies:*

$$u'(c_1(i))<\beta R_2\sum_j u'(c_2(i,j))\pi_2(j\,|\,i) \quad \Rightarrow \quad \tau_k(i)>0.$$

The intuition for this intertemporal wedge is that implicit savings affect the incentives to work. Specifically, consider an agent who is contemplating a deviation. Such an agent prefers to implicitly save more than the agent who is planning to tell the truth. An intertemporal wedge worsens the return to such deviation.

The Inverse Euler Equation can be extended to the case of aggregate uncertainty (Kocherlakota 2005). At the optimum

$$\frac{1}{u'(c_1(i))}=\frac{1}{\beta\sum_s\left[R_2(s)\left[\sum_j \pi(j\mid i)[u'(c_2(i,j,s)]^{-1}\right]^{-1}\right]\mu(s)}$$

If there is no uncertainty regarding skills in the second period, this expression reduces to

$$u'(c_1(i))=\beta\sum_s[R_2(s)u'(c_2(i,s))]\mu(s)$$

so that the intertemporal marginal rate of substitution is undistorted. However, if the agent faces idiosyncratic uncertainty about his skills and consumption in the second period, Jensen's inequality implies that there is a positive wedge on savings:

$$u'(c_1(i))<\beta\sum_{j,s}\mu(s)\pi(j\mid i)R_2(s)u'(c_2(i,j,s)).$$

4.2 *Tax Smoothing*

One of the main results from the representative-agent Ramsey framework is that tax rates on labor income should be smoothed across time (Barro 1979) and states (Lucas and Stokey 1983).

This result extends to cases with heterogenous agents subject to linear or nonlinear taxation (Werning 2007), that is, where all the unobservable idiosyncratic uncertainty about skills is resolved in the first period. To see this, take $\theta_2(j,i)=\theta_1(i)=\theta(i)$. We can then write the allocation entirely in terms of the first period skill shock and the second period aggregate shock. The incentive constraints then only require truthful revelation of the first period's skill type i,

$$u(c_1(i))+v\left(\frac{y_1(i)}{\theta(i)}\right)+\beta\sum_s\left[u(c_2(i,s))+v\left(\frac{y_2(i,s)}{\theta(i)}\right)\right]\mu(s)\geq \tag{10}$$

$$u(c_1(i_r))+v\left(\frac{y_1(i_r)}{\theta(i)}\right)+\beta\sum_s\left[u(c_2(i_r,s))+v\left(\frac{y_2(i_r,s)}{\theta(i)}\right)\right]\mu(s)$$

for all i, i_r. Let $\psi(i,i_r)$ represent the Lagrangian multiplier associated with each of these inequalities.

The Lagrangian for the planning problem that incorporates these constraints can be written as

$$\sum_{i,i_r,s}\left\{(1+\psi(i,i_r))\left[u(c_1(i))+v\left(\frac{y_1(i)}{\theta(i)}\right)+\beta\left(u(c_2(i,s))+v\left(\frac{y_2(i,s)}{\theta(i)}\right)\right)\right]\right.$$
$$\left.-\psi(i,i_r)\left[u(c_1(i_r))+v\left(\frac{y_1(i_r)}{\theta(i)}\right)+\beta\left(u(c_2(i_r,s))+v\left(\frac{y_2(i_r,s)}{\theta(i)}\right)\right)\right]\right\}\mu(s)\pi_1(i)$$

To derive the next result we adopt an iso-elastic utility of work effort function $v(n) = -\kappa n^\gamma/\gamma$ with $\kappa > 0$ and $\gamma \geq 1$. The first-order conditions are then

$$u'(c_1(i))\eta^c(i)=\lambda_1 \qquad u'(c_2(i,s))\eta^c(i)=\lambda_2(s)$$
$$-\frac{1}{\theta(i)}v'\left(\frac{y_1(i)}{\theta(i)}\right)\eta^y(i)=\lambda_1 \qquad -\frac{1}{\theta(i)}v'\left(\frac{y_2(i,s)}{\theta(i)}\right)\eta^y(i)=\lambda_2(s)$$

where λ_1 and $\lambda_2(s)$ are first and second period multipliers on the resource constraints and where we define

$$\eta^c(i)\equiv 1+\sum_{i'}\left(\psi(i,i')-\psi(i',i)\frac{\pi(i')}{\pi(i)}\right)$$
$$\eta^y(i)\equiv 1+\sum_{i'}\left(\psi(i,i')-\psi(i',i)\left(\frac{\theta(i)}{\theta(i')}\right)\frac{\pi(i')}{\pi(i)}\right)$$

for notational convenience. Combining and cancelling terms then leads to

$$\tau_1\equiv 1+\frac{1}{\theta(i)}\frac{v'\left(\frac{y_1(i)}{\theta(i)}\right)}{u'(c_1(i))}=1-\frac{\eta^c(i)}{\eta^y(i)} \qquad \tau_2(s)\equiv 1+\frac{1}{\theta(i)}\frac{v'\left(\frac{y_2(i,s)}{\theta(i)}\right)}{u'(c_2(i,s))}=1-\frac{\eta^c(i)}{\eta^y(i)}$$

which proves that perfect tax smoothing is optimal in this case. We summarize this result in the next proposition, derived by Werning (2007) for a more general dynamic framework.

Proposition 4 *Suppose the disutility of work effort is isoelastic: $v(n) = -\kappa n^\gamma/\gamma$. Then when idiosyncratic uncertainty for skills is concentrated in the first period, so that $\theta_2(j, i) = \theta_1(i)$ then it is optimal to perfectly smooth marginal taxes on labor $\tau_1 = \tau_2(s) = \overline{\tau}$.*

Intuitively, tax smoothing results from the fact that the tradeoff between insurance and incentives remains constant between periods and across states. As shown by Werning (2007), if the distribution of skills varies across periods or aggregate states, then optimal marginal taxes should also vary with these shifts in the distribution. Intuitively, the tradeoff between insurance and incentives then shifts and taxes should adjust accordingly. In the numerical work in section 6 we examine another source for departures from the perfect tax smoothing benchmark.

4.3 Tax Implementations

In this section we describe the general idea behind *decentralization* or *implementation* of optimal allocations with tax instruments. The general goal is to move away from the direct mechanism, justified by the revelation principle to study constrained efficient allocations, and find tax systems so that the resulting competitive equilibrium yields these allocations. In general, the required taxes are complex nonlinear functions of all past observable actions, such as capital and labor supply, as well as aggregate shocks.

It is tempting to interpret the wedges defined in (5)–(7) as actual taxes on capital and labor in the first and second periods. Unfortunately, the relationship between wedges and taxes is typically less straightforward. Intuitively, each wedge controls only one aspect of worker's behavior (labor in the first or second period, or saving) taking all other choices fixed *at the optimal level.* For example, assuming that an agent supplies the socially optimal amount of labor, a savings tax defined by (7) would ensure that that agent also makes a socially optimal amount of savings. However, agents choose labor and savings jointly.[14]

In the context of our economy, taxes in the first period $T_1(y_1)$ can depend only on the observable labor supply of agents in that periods, and taxes in the second period $T_2(y_1, y_2, k, s)$ can depend on labor supply in both first and second period, as well as agents' wealth. In competitive equilibrium, agent i solves

$$\max_{\{c,y,k\}}\left\{u(c_1(i), y_1(i)/\theta_1(i)) + \beta\sum_{s,j}\left[u(c_2(i,j,s)) + v\left(\frac{y_2(i,j,s)}{\theta_2(i,j)}\right)\right]\pi_2(j\,|\,i)\mu(s)\right\}$$

subject to

$$c_1(i) + k(i) \le y_1(i) - T_1(y_1(i))$$

$$c_2(i, j, s) \le y_2(i, j, s) + R_2(s)k(i) - T_2(y_1(i), y_2(i, j, s), k(i), s).$$

We say that a tax system implements the socially optimal allocation $\{(c_1^*(i), y_1^*(i), c_2^*(i, j, s), y_2^*(i, j, s)\}$ if this allocation solves the agent's problem, given $T_1(y_1(i))$ and $T_2(y_1(i), y_2(i, j, s), k(i), s)$.

Generally, an optimal allocation may be implementable by various tax systems so $T_1(y_1(i))$ and $T_2(y_1(i), y_2(i, j, s), k(i), s)$ may not be uniquely determined. In contrast, all tax systems introduce the same wedges in agents' savings or consumption-leisure decisions. For this reason, in the numerical part of the paper we focus on the distortions defined in section 3, and omit the details of any particular implementation. In this section, however, we briefly review some of the literature on the details of implementation.

Formally, the simplest way to implement allocations is a *direct mechanism,* which assigns arbitrarily high punishments if individual's consumption and labor decisions in any period differ from those in the set of the allocations $\{(c_1^*(i), y_1^*(i), c_2^*(i, j, s), y_2^*(i, j, s)\}$ that solve the planning program. Although straightforward, such an implementation is highly unrealistic and severely limits agents' choices. A significant body of work attempts to find less heavy handed alternatives. One would like implementations to come close to using tax instruments currently employed in the United States and other advanced countries. Here we review some examples.

Albanesi and Sleet (2006) consider an infinitely repeated model where agents face *i.i.d.* skill shocks over time and there are no aggregate shocks. They show that the optimal allocation can be implemented by taxes that depend in each period only on agent's labor supply and capital stock (or wealth) in that period. The tax function $T_t(y_t, k_t)$ is typically non-linear in both of its arguments. Although simple, their implementation relies critically on the assumption that idiosyncratic shocks are *i.i.d.* and cannot be easily extended to other shocks processes.

Kocherlakota (2005) considers a different implementation that works for a wide range of shock processes for skills. His implementation separates capital from labor taxation. Taxes on labor in each period t depend on the whole history of labor supplies by agents up until period t and in general can be complicated non-linear functions. Taxes on capital are linear and also history dependent. Specifically, the tax rate on capital that is required is given by (written, for simplicity, for the case with no aggregate uncertainty)

$$\tilde{\tau}_k(i, j) = 1 - \frac{u'(c^*(i))}{\beta R_2 u'(c^*(i, j))} \tag{11}$$

Incidentally, an implication of this implementation is that, at the optimum, taxes on capital average out to zero and raise no revenue. That is, the conditional average over j for $\tilde{\tau}_k(i, j)$ given by equation (11) is zero when the Inverse Euler equation (8) holds. At first glance, a zero average tax rate may appear to be at odds with the positive intertemporal wedge $\tau_k(i)$ defined by equation (7) found in Proposition 3, but it is not: Savings are discouraged by this implementation. The key point is that the tax is not deterministic, but random. As a result, although the average net return on savings is unaffected by the tax, the net return $R_2(s)(1 - \tilde{\tau}_k(i, j, s))$ is made risky. Indeed, since net returns are negatively related to consumption, see equation (11), there is a risk-premium component (in the language of financial economics) to the expected return. This tax implementation makes saving strictly less attractive, just as the positive intertemporal wedge τ_k suggests.

In some applications the number of shocks that agents face is small and, with a certain structure, that allows for simple decentralizations. Golosov and Tsyvinski (2006a) study a model of disability insurance, where the only uncertainty agents face is whether, and when, they receive a permanent shock that makes them unable to work. In this scenario, the optimal allocation can be implemented by paying disability benefits to agents who have assets below a specified threshold, i.e., asset testing the benefits.

4.4 Time Inconsistency

In this section we argue that the dynamic Mirrlees literature and Ramsey literature are both prone to time-consistency problems. However, the nature of time inconsistency is very different in those two approaches.

An example that clarifies the notion of time inconsistency in Ramsey models is taxation of capital. The Chamley-Judd (Judd 1985, Chamley 1986) result states that capital should be taxed at zero in the long run. One of the main assumptions underlying this result is that a government can commit to a sequence of capital taxes. However, a benevolent government would choose to deviate from the prescribed sequence of taxes. The reason is that, once capital is accumulated, it is sunk, and taxing capital is no longer distortionary. A benevolent government would choose high capital taxes once capital is accumulated. The reasoning above motivates the analysis of time consistent policy as a game between a policy maker (government) and a continuum of economic agents (consumers).[15]

To highlight problems that arise when we depart from the benchmark of a benevolent planner with full commitment, it is useful to start with Roberts' (1984) example economy, where, similar to Mirrlees (1971), risk-averse individuals are subject to unobserved shocks affecting the marginal disutility of labor supply. But unlike the benchmark Mirrlees model, the economy is repeated T times, with individuals having perfectly persistent types. Under full commitment, a benevolent planner would choose the same allocation at every date, which coincides with the optimal solution of the static model. However, a benevolent government without full commitment cannot refrain from exploiting the information that it has collected at previous dates to achieve better risk sharing ex post. This turns the optimal taxation problem into a dynamic game between the government and the citizens. Roberts showed that as discounting disappears and $T \to \infty$, the unique sequential equilibrium of this game involves the highly inefficient outcome in which all types declare to be the worst type at all dates, supply the lowest level of labor and receive the lowest level of consumption. This example shows the potential inefficiencies that can arise once we depart from the case of full commitment, even with benevolent governments. The nature of time inconsistency in dynamic Mirrlees problems is, therefore, very different from that in a Ramsey model. In the dynamic Mirrlees model the inability of a social planner not to exploit information it learns about agents' types is a central issues in designing optimal policy without commitment. A recent paper by Bisin and Rampini (2006) considers the problem of mechanism design without commitment in a two-period setting. They show how the presence of anonymous markets acts as an additional constraint on the government, ameliorating the commitment problem.

Acemoglu, Golosov, and Tsyvinski (2006) depart from Roberts' (1984) framework and consider, instead of a finite-horizon economy, an infinite-horizon economy. This enables them to use punishment strategies against the government to construct a *sustainable mechanism,* defined as an equilibrium tax-transfer program that is both incentive compatible for the citizens and for the government (i.e., it satisfies a sustainability constraint for the government). The (best) sustainable mechanism implies that if the government deviates from the implicit agreement, citizens switch to supplying zero labor, implicitly punishing the government. The infinite-horizon setup enables them to prove that a version of the revelation principle, *truthful revelation along the equilibrium path,* applies and is a useful tool of analysis for this class of dynamic incentive problems with self-interested mechanism designers and without commitment.[16] The fact that the truthful revelation principle applies *only* along the equilibrium path is important, since it is actions off the equilibrium path that place restrictions on what type of mechanisms are allowed (these are encapsulated in the sustainability constraints). This enables them to construct sustainable mechanisms with the revelation principle along the equilibrium path, to analyze more general environments, and to characterize the limiting behavior of distortions and taxes.

4.5 The Government's Role As Insurance Provider

In the previous discussion we assumed that a government is the sole provider of insurance. However, in many circumstances, markets can provide insurance against shocks that agents experience. The presence of competitive insurance markets may significantly change optimal policy prescriptions regarding the desirability and extent of taxation and social insurance policies.

We assumed that individual asset trades and, therefore, agents' consumption, are publicly observable. In that case, following Prescott and Townsend (1984), Golosov and Tsyvinski (2006b) show that allocations provided by competitive markets are constrained efficient and the first welfare theorem holds. The competitive nature of insurance markets, even in the presence of private information, can provide optimal insurance as long as consumption and output are publicly observable. Note that individual insurance contracts, between agents and firms, would feature the same wedges as the social planning problem we studied, providing another motivation for focusing on wedges, rather than taxes that implement them.

In this paper we do not model explicitly reasons why private insurance markets may provide the inefficient level of insurance. Arnott and Stiglitz (1986, 1990), Greenwald and Stiglitz (1986), and Golosov and Tsyvinski (2006b) explore why markets may fail in the presence of asymmetric information.

5 Numerical Exercises

We now turn to numerical exercises with baseline parameters and perform several comparative-static experiments. The exercises we conduct strike a balance between flexibility and tractability. The two period setting is flexible enough to illustrate the key theoretical results and explore a few new ones. At the same time, it is simple enough that a complete solution of the optimal allocation is possible. In contrast, most work on Mirrleesian models has focused on either partial theoretical characterizations of the optimum, e.g., showing that the intertemporal wedge is positive (Golosov, Kocherlakota, and Tsyvinski, 2003) or on numerical characterizations for a particular skills processes, e.g., *i.i.d.* skills in Albanesi and Sleet (2006) or absorbing disability shocks in Golosov and Tsyvinski (2006a). In a recent paper, Farhi and Werning (2006a) take a different approach, by studying partial tax reforms—that fully capture the savings distortions implied by the Inverse Euler equation. The problem remains tractable even with empirically relevant skill processes.

5.1 Parameterization

When selecting parameters it is important to keep the following neutrality result in mind. With logarithmic utility, if productivity and government expenditures are scaled up within a period then: (1) the allocation for consumption is scaled by the same factor; (2) the allocation of labor is unaffected; and (3) marginal taxes rates are unaffected. This result is relevant for thinking about balanced growth in an extension of the model to an infinite horizon. It is also convenient in that it allows us to normalize, without any loss of generality, the second period shock for our numerical explorations.

We now discuss how we choose parameters for the benchmark example. We use the following baseline parameters. We first consider the case with no aggregate uncertainty. Assume that there is no discounting and that the rate of return on savings is equal to the discount factor: $R = \beta = 1$.

We choose the skill distribution as follows. In the first period, skills are distributed uniformly. Individual skills in the first period, $\theta_1(i)$, are equally spaced in the interval $[\underline{\theta}_1, \overline{\theta}_1]$. The probability of the realization of each skill is equal to $\pi_1(i) = 1/N_1$ for all i. We choose baseline parameters to be $\underline{\theta}_1 = 0.1$, $\overline{\theta}_1 = 1$, and $N_1 = 50$. Here, a relatively large number of skills allows us to closely approximate a continuous distribution of skills. In the second period, an agent can receive a skill shock. For computational tractability, we assume that there are only two possible shocks to an agent's skill in the second period, $N_2(i) = 2$ for all i. Skill shocks take the form of a proportional increase $\theta_2(i, 1) = \alpha_1\theta_1(i)$ or proportional decrease $\theta_2(i, 2) = \alpha_2\theta_1(i)$. For the baseline case, we set $\alpha_1 = 1$, and $\alpha_2 = 1/2$. This means that an agent in the second period can only receive an adverse shock α_2. We also assume that there is uncertainty about realization of skills and set $\pi_2(1 \mid i) = \pi_2(2 \mid i) = 1/2$. The agent learns his skill in the second period only at time $t = 2$. We chose the above parameterization of skills to allow a stark characterization of the main forces determining the optimum.[17]

We choose the utility function to be power utility. The utility of consumption is $u(c) = c^{1-\sigma}/(1 - \sigma)$. As our baseline we take $\sigma = 1$, so that $u(c) = \log(c)$. The utility of labor is given by $v(l) = -l^{\gamma}$; as our benchmark we set $\gamma = 2$.

We use the following conventions in the figures below:

1. The horizontal axis displays the first period skill type $i = 1, 2, \ldots, 50$;

2. The wedges (distortions) in the optimal solutions are labeled as follows:

(a) "Distortion $t = 1$" is the consumption-labor wedge in period 1: τ_{y_1};

(b) "Distortion high $t = 2$" is the consumption-labor wedge in period 2 for an agent with a high skill shock: $\tau_{y_2}(i, 1)$;

(c) "Distortion low $t = 2$" is the consumption-labor wedge in period 2 for an agent with a low skill shock: $\tau_{y_2}(i, 2)$;

(d) "Distortion capital" is the intertemporal (capital) wedge: $\tau_k(i)$.

5.2 *Characterizing the Benchmark Case*

In this section, we describe the numerical characterization of the optimal allocation. Suppose first that there were no informational friction and agents' skills were observable. Then the solution to the optimal program would feature full insurance. The agent's consumption would be equalized across realizations of shocks. Labor of agents would be increasing with their type. It is obvious that when skills are unobservable the unconstrained optimal allocation is not incentive compatible, as an agent with a higher skill would always prefer to claim to be of a lower type to receive the same consumption but work less. The optimal allocation with unobservable types balances two objectives of the social planner: Providing insurance and respecting incentive compatibility constraints.

The optimal allocation for the benchmark case with unobservable types is shown in figure 5.1 and figure 5.2. There is no bunching in either period: Agents of different skills are allocated different consumption and labor bundles.

First note that there is a significant deviation from the case of perfect insurance: agents' consumption increases with type, and consumption in the second period for an agent who claims to have a high shock is higher than that of an agent with the low shock. The intuition for this pattern of consumption is as follows. It is optimal for an agent with a higher skill to provide a higher amount of effective labor. One way to make provision of higher effective labor incentive compatible for an agent is to allocate a larger amount of consumption to him. Another way to reward an agent for higher effort is to increase his continuation value, i.e., allocate a higher amount of expected future consumption for such an agent.

We now turn our attention to the wedges in the constrained efficient allocation. In the unconstrained optimum with observable types, all wedges are equal to zero. We plot optimal wedges for the benchmark case in figure 5.3.

We see that the wedges are positive, indicating a significant departure from the case of perfect insurance. We notice that the consumption-labor wedge is equal to zero for the highest skill type in the first period and for the high realization of the skill shock in the second period: $\tau_{y_1}(\overline{\theta}_1) = \tau_{y_2}(\overline{\theta}_1, 1) = 0$. This result confirms a familiar "no distortion at the top" result due to Mirrlees (1971) which states that in a static context the consumption-labor decision of an agent with the highest skill is undistorted in the optimal allocation. The result that we obtain here is somewhat novel as we consider an economy with stochastically evolving skills, for which the "no distortion at the top" result have not yet been proven analytically.

We also see that the labor wedges at the bottom $\{\tau_{y_1}(\underline{\theta}_1), \tau_{y_2}(\underline{\theta}_1, 1), \tau_{y_2}(\underline{\theta}_1, 2)\}$ are strictly positive. A common result in the literature is that with a continuum of types, the tax rate at the bottom is zero if bunching types is not optimal. In our case, there is no bunching, but this result does not literally apply because we work with a discrete distribution of types.

We see that the intertemporal wedge is low for agents with low skills θ_1 in the first period yet is quite high for agents with high skills. The reason is that it turns out that lower skilled workers are quite well insured: Their consumption is not very volatile in the second period. It follows

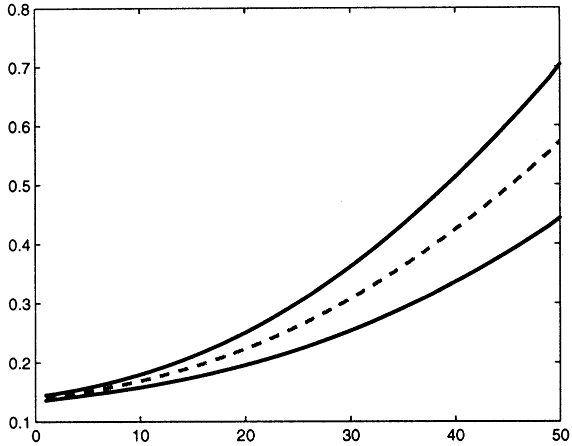

Figure 5.1
Consumption Allocation. Middle Dotted Line Shows First Period Consumption; Outer Solid Lines Are Second Period Consumption

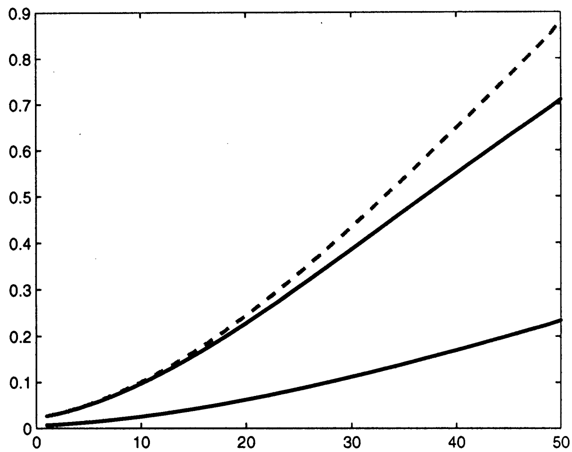

Figure 5.2
Effective Labor Allocation. Dashed Line Is for First Period. Solid Lines Are for Second Period, Top Is High Shock, Bottom Low Shock

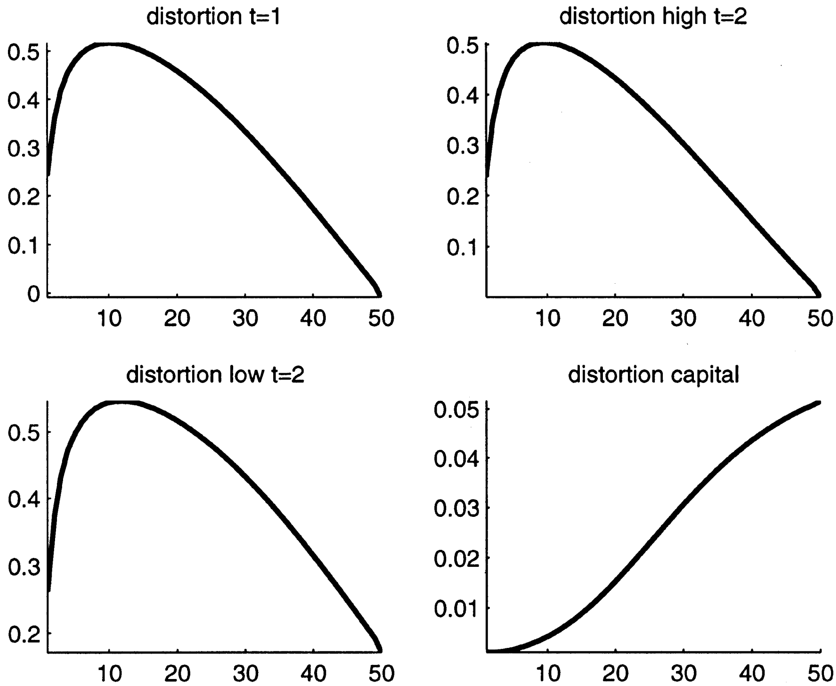

Figure 5.3
Benchmark Implicit Marginal Tax Rates

that the intertemporal distortion required is smaller. Note that figure 5.1 shows that consumption uncertainty in the second period increases with the first period shock.

5.3 *Effects of the Size of Second Period Shocks*

We now consider the effects of an increase in the size of the adverse second period shock affecting agents. This is an important exercise as it allows us to identify forces that distinguish the dynamic Mirrlees taxation in which skills stochastically change over time from a dynamic case in which types of agents do not change over time. We consider a range of shocks: From a very large shock ($\alpha_2 = 0.05$), that makes an agent almost disabled in the second period, to a small drop ($\alpha_2 = 0.95$) that barely changes the agent's skill. In figure 5.4 the bold line corresponds to the benchmark case of $\alpha_2 = 0.5$; the dashed lines correspond to $\alpha_2 =$ 0.6, 0.8, 0.9, and 0.95, while the dotted lines correspond to $\alpha_2 = 0.3$, 0.1, and 0.05 respectively.

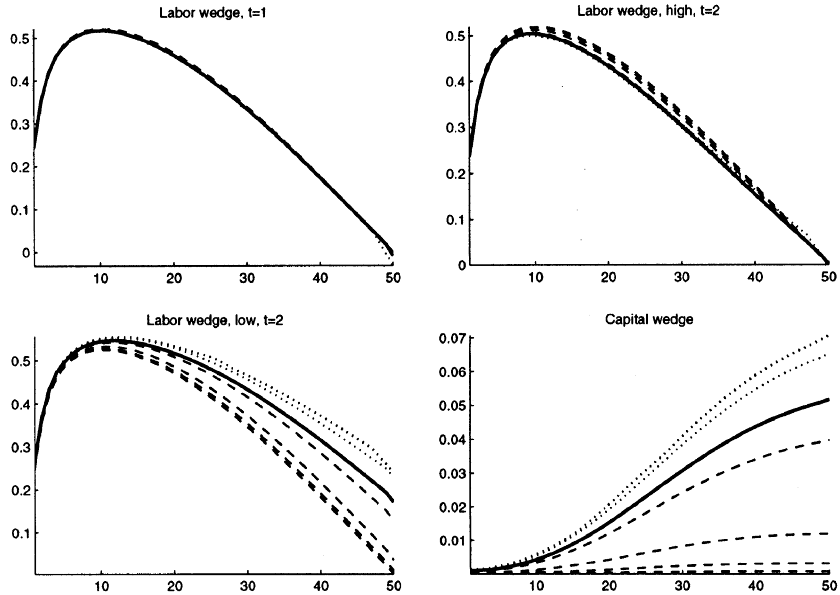

Figure 5.4
Varying α_2

We now describe the effects of an increase in the size of the skill shocks on the labor wedges. First notice that the size of the second period shocks practically does not affect the first period wedge schedule $\tau_{y_1}(\theta_1)$, and the shape and the level are preserved: Even when agents experience a high shock to their skills (e.g., $\alpha_2 = 0.05$), the schedule of labor wedges in the first period is, essentially, identical to the case when an agent experiences a very small shock ($\alpha_2 = 0.95$). Similarly, we don't see large changes in the marginal labor wedge schedule, $\tau_{y_2}(\cdot, 1)$, in the second period for the high realization of the shocks (i.e., if skills remain the same as in the previous period). Interestingly, the marginal tax on labor in the second period after a downward drop, $\tau_{y_2}(\cdot, 2)$ changes significantly. As α_2 increases, the shock to skill becomes smaller and the level of wedges at the top falls. To see this effect, compare the upper dotted line for $\alpha_2 = 0.05$ with the bottom dashed line for $\alpha_2 = 0.95$.

To summarize the discussion above, we conclude that the size of the second period shock only has significant effects on labor wedges for the agents who experience that shock and only in that period. Intuitively, the skill distribution for agents not affected by the shocks matters only indirectly, and, therefore, the labor wedge for those agents is affected only to a small degree.

We now proceed to characterize the effects of the size of shocks on the capital wedge. The intertemporal wedge becomes smaller and flatter when α_2 increases—compare, for example, the lower curve associated with $\alpha_2 = 0.95$ to the highest curve associated with $\alpha_2 = 0.05$. The reason is that consumption becomes less volatile in the second period when the skill drop is smaller. The inverse Euler equation then implies a smaller distortion. The intuition for this result is simple. If there were no skill shocks in the second period ($\alpha_2 = 1$) then, as we discussed above, the capital wedge is equal to zero. The higher is the wedge in the second period, the further away from the case of constant skills we are, therefore, the distortion increases. Also note that low α_2 (large shocks in the second period) significantly steepens the capital wedge profile.

We conclude that the shape and size of the capital wedge responds significantly to the size of the shocks that an agent may experience in the future.

5.4 Effects of the Probability of Second Period Shocks and Uncertainty

We now consider the effects of changing the probability of the adverse second period shock. This exercise is of interest because it allows us to investigate the effects of *uncertainty* about future skill realizations on the size and shape of wedges.

In figure 5.5, we show in bold the benchmark case where $\pi_2(2 \mid \cdot) = 0.5$; dashed line correspond to $\pi_2(2 \mid \cdot) = 0.7$ and 0.9 while the dotted lines correspond to $\pi_2(2 \mid \cdot) = 0.3$ and 0.1, respectively.

We first notice that the effects of the change in the probability of the adverse shock on labor wedge are similar to the case of increase in size of the adverse shock. That is, as the probability $\pi_2(2 \mid \cdot)$ of a drop in skills rises, the informational friction increases and so does the labor wedge.

For the intertemporal wedge there is an additional effect of changing the probability of the adverse skill shock. The wedge is the highest when uncertainty about skills is the highest: At the symmetric baseline case with $\pi_2(2 \mid \cdot) = 0.5$. Intuitively, the reason is that the uncertainty about next period's skill is maximized at $\pi_2(2 \mid \cdot) = 0.5$. It is uncertainty about future skills, rather than the level of next period's skill shock, that matters for the size of the capital wedge.

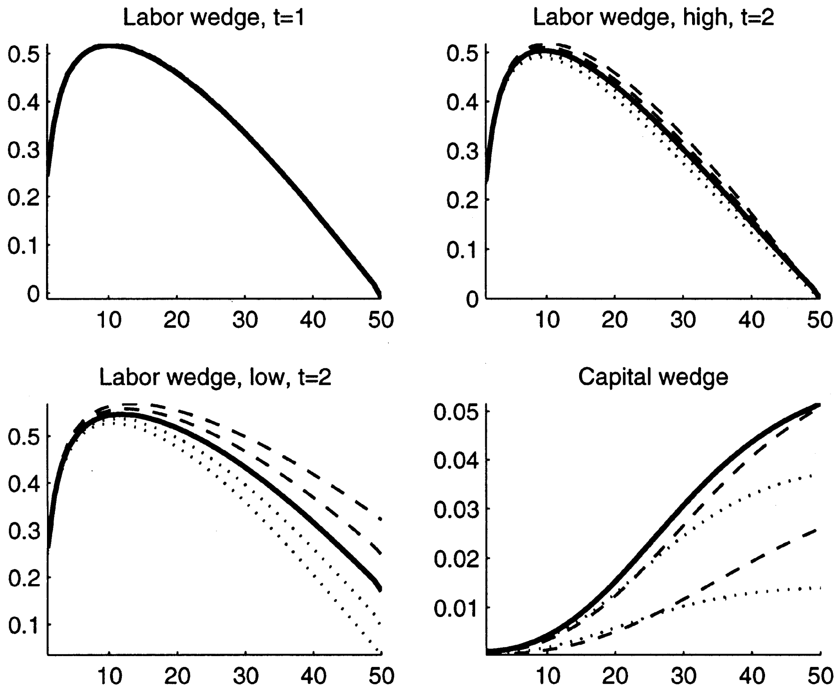

Figure 5.5
Varying the Probability of Skill Drop $\pi_2(2 \mid \cdot)$

5.5 *Effects of Changing Risk Aversion*

We proceed to explore effects of risk aversion on optimal wedges and allocations. This exercise is important as risk aversion determines the need for redistribution or insurance for an agent. Specifically, we change the risk aversion parameter σ in the utility function. The results are shown in figure 5.6. Our benchmark case of logarithmic utility $\sigma = 1$ is shown in bold. With dotted lines we plot lower risk aversions: $\sigma = 0.8, 0.5, 0.3$, and 0.1; and with dashed lines we plot higher risk aversions: $\sigma = 1.5$ and 3.

The immediate observation is that a higher degree of risk aversion leads to uniformly higher distortions. The intuition is again rather simple. We know that if $\sigma = 0$, so that utility is linear in consumption and an agent is risk neutral, private information about the skill would not affect the optimal allocation and the unconstrained allocation in which all wedges are equal to zero can be obtained. The higher is risk aversion, the higher is the desire of the social planner to redistribute and insure agents. Therefore, all distortions rise.

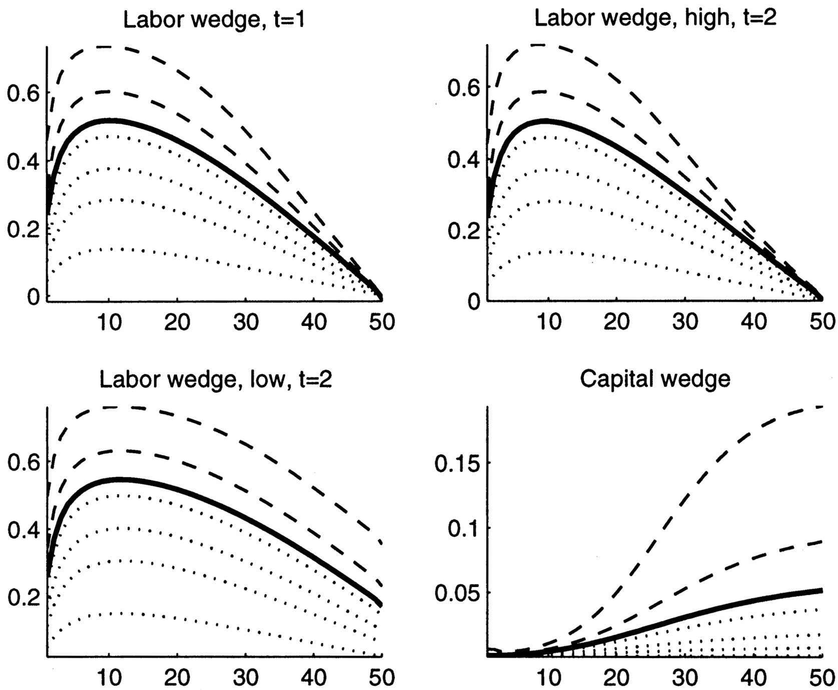

Figure 5.6
Varying Risk Aversion

The effects of higher risk aversion on the intertemporal wedge are the outcome of two opposing forces: (1) a direct effect: for a given consumption allocation, a higher risk aversion σ increases the wedge—the capital wedge results from the Inverse Euler equation by applying Jensen's inequality, which is more powerful for higher σ; (2) an indirect effect: with higher curvature in the utility function $u(c)$ it is optimal to insure more, lowering the variability of consumption across skill realizations, which reduces the capital wedge. For the cases we considered the direct effect turned out to be stronger and the capital wedge increases with risk aversion.

5.6 Effects of Changing Elasticity of Labor Supply

We further investigate the properties of the optimum by considering three modifications of the disutility of labor. Figure 5.7 shows the results. Our benchmark case, as before, is $v(l) = -l^2$ (plotted in bold in the figure). We also display two more inelastic cases: $v(l) = -l^3$ and $v(l) = -l^4$ (plotted with dashed lines).

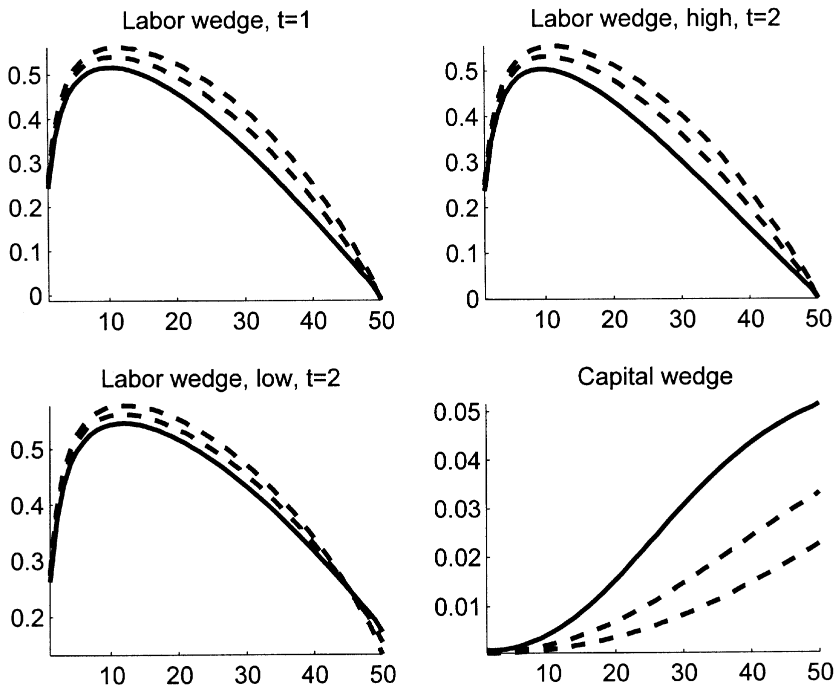

Figure 5.7
Changing Elasticity of Labor

Regarding the effect on labor distortions, intuitively, there are two opposing forces. On the one hand, as labor becomes more inelastic, wedges introduce smaller inefficiencies. Thus, redistribution or insurance is cheaper. On the other hand, since our exercises hold constant the skill distribution, when labor supply is more inelastic the distribution of earned income is more equal. Hence, redistribution or insurance are less valuable. Thus, combining both effects, there is less uncertainty or inequality in consumption, but marginal wedges may go either up or down. In our simulations it seems that the first effect dominated and the labor wedges increased when the elasticity of labor was reduced.

The distortion on capital unambiguously goes down since consumption becomes less variable.

5.7 Exploring Nonseparable Utility

We now consider a modification to the case of non-separable utility between consumption and labor. When the utility is nonseparable, the analytical Inverse Euler results that ensured a positive intertemporal wedge may no longer hold. Indeed, the effects of nonseparable utility on the intertemporal wedge are largely unexplored.

5.7.1 Building on a Baseline Case We start with the specification of the utility function that can be directly comparable with our baseline specification

$$u(c,l)=\frac{(ce^{-l^2})^{1-\sigma}}{1-\sigma}.$$

Here, the baseline case with separable utility is equivalent to $\sigma = 1$. When $\sigma < 1$ risk aversion is lower than in our baseline and consumption and work effort are substitutes in the sense that $u_{cl} < 0$, that is, an increase in labor decreases the marginal utility of consumption. When $\sigma > 1$ the reverse is true, risk aversion is higher and consumption and labor are complements, in that $u_{cl} > 0$. For both reasons, the latter case is considered to be the empirically relevant one.

We first consider $\sigma < 1$ cases. Figure 5.8 shows the schedules for $\sigma =$ 1, 0.9, 0.7, 0.65. The baseline with $\sigma = 1$ is plotted as a dotted line. Lower σ correspond to the lower lines on the graph.

We notice that a lower σ pushes the whole schedule of labor distortions down. Intuitively, with lower risk aversion it is not optimal to redistribute or insure as much as before: The economy moves along the equality-efficiency tradeoff towards efficiency.

The results for capital taxation are more interesting. First, a lower σ is associated with a uniformly lower schedule of capital distortions. Second, lower σ introduces a non-monotonicity in the schedule of capital distortions, so that agents with intermediate skills have lower capital distortion than those with higher or lower skills. Finally, for all the cases considered with $\sigma < 1$, we always find an intermediate region where the intertemporal wedge is negative.

To understand this result it is useful to think of the case without uncertainty in the second period. For this case, Atkinson and Stiglitz (1976) show that, when preferences are separable, savings should not be taxed, but that, in general, whenever preferences are non-separable some distortion is optimal. Depending on the details of the allocation and on the sign of u_{cl} this distortion may be positive or negative.

We now turn to the case with $\sigma > 1$ and consider $\sigma = 1, 2, 3$ (see figure 5.9). The baseline with $\sigma = 1$ is plotted as the dotted line. Away from the baseline, higher σ correspond to lower lines on the graph.

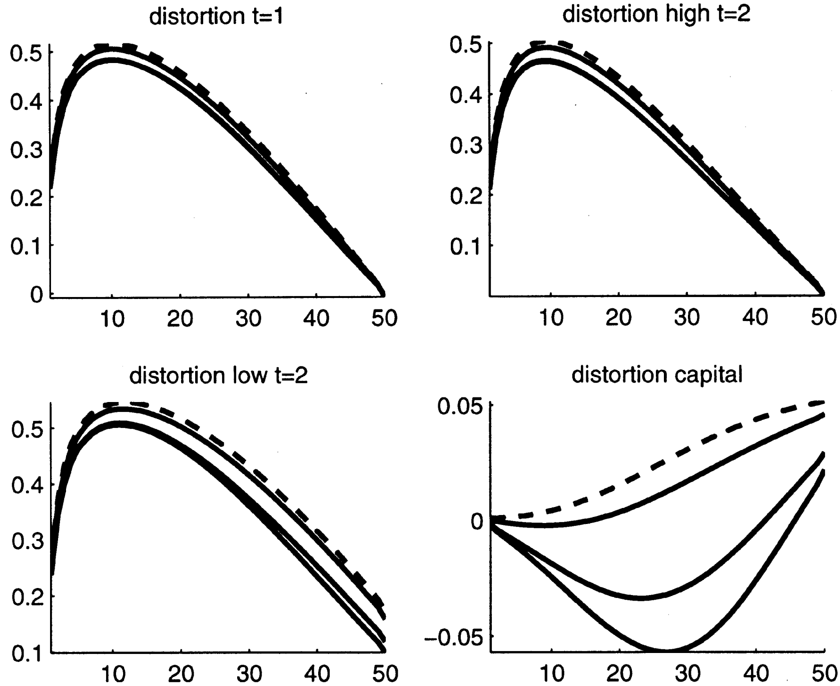

Figure 5.8
Nonseparable Utility with $\sigma \leq 1$

We notice that higher σ pushes the whole schedule of labor distortions up. The intuition is again that higher risk aversion leads to more insurance and redistribution, requiring higher distortions.

A higher σ is associated with a uniformly higher schedule of capital distortions and these are always positive. Second, higher σ may create a non-monotonicity in the schedule of capital distortions, with the highest distortions occurring for intermediate types.

It is not only the value of the σ that determines the sign of the wedge. We found that for the case where the skill shocks in the second period have an upward trend so that $\alpha_1 = 1.5$ and $\alpha_2 = 1$, i.e., an agent may experience a positive skill shock, the results are reversed. In particular, for $\sigma < 1$, we found that capital wedges were always positive, whereas for $\sigma > 1$ they were negative over some region of skills. Intuitively, the trend in skills matters because it affects the trend in labor.

We obtained similar results with the alternative specification of utility also common in macroeconomic models:

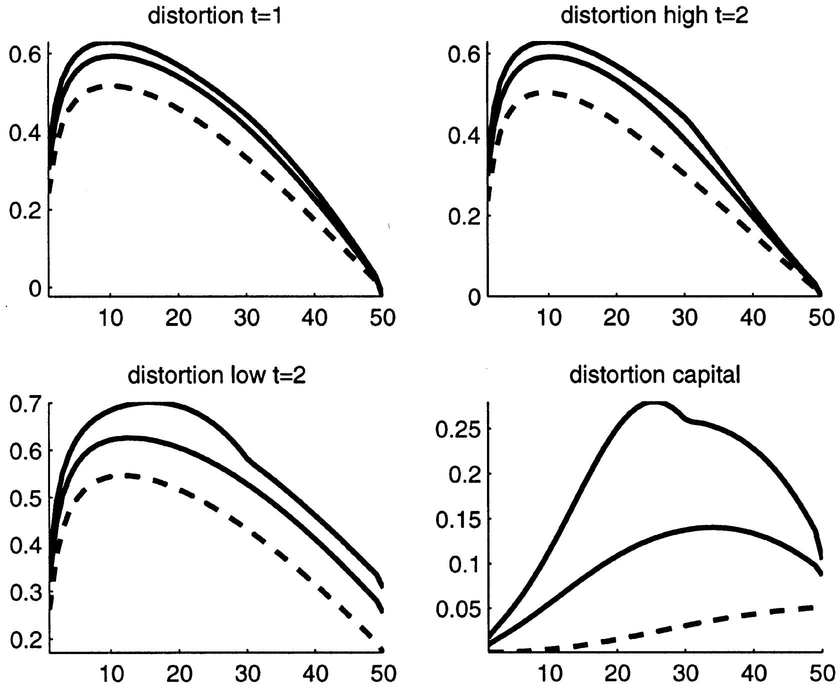

Figure 5.9
Nonseparable Utility with $\sigma \geq 1$

$$u(c,l) = \frac{(c^{1-\gamma}(L-l)^{\gamma})^{1-\sigma}}{1-\sigma}.$$

This utility function was used by Chari, Christiano, and Kehoe (1994) in their quantitative study of optimal monetary and fiscal policy.

5.8 Summarizing the Case with No Aggregate Uncertainty

The exercises above give us a comprehensive overview of how the optimal wedges depend on the parameters of the model. We now summarize what seems to be most important for the size and the shape of these wedges.

1. Labor wedges on the agent affected by an adverse shock increase with the size or the probability of that shock. However, labor wedges in other periods and labor wedges for agents unaffected by the adverse

shock are influenced only indirectly by this variable and the effects are small.

2. Higher risk aversion increases the demand for insurance and significantly increases the size of both labor wedges. However, the effect on capital wedges could be ambiguous as the uncertainty about future skills also matters.

3. Capital wedges are affected by the degree of uncertainty over future skills.

4. A lower elasticity of labor decreases the capital wedge but could have ambiguous effects on labor wedge for a given skill distribution.

5. If utility is nonseparable between consumption and labor, the capital wedge may become negative. The sign of the wedge in that case depends on whether labor is complementary or substitutable with consumption and on whether an agent expects to experience a higher or a lower shock to skills in the future.

6 Aggregate Uncertainty

In this section we explore the effects of aggregate uncertainty. In section 4.2 we showed that if agents' types are constant it is optimal to perfectly smooth labor taxes, i.e., the labor wedges are constant across states and periods. The literature on new dynamic public finance virtually has not explored implications of aggregate uncertainty.[18]

6.1 Baseline Parameterization

We use, unless otherwise noted, the same benchmark specifications as in the case with no aggregate uncertainty. Additional parameters that we have to specify are as follows. We assume that there are two aggregate states, $s = 2$. The probability of the aggregate states are symmetric: $\mu(1) = \mu(2) = 1/2$. We take the number of skills in the first period to be $N_1 = 30$. As before, skills are equispaced and uniformly distributed. We set $R_1 = 1$.

6.2 Effects of Government Expenditure Fluctuations

We now turn to analyzing the effects of government expenditures. There is a sense in which return and government expenditure shocks are similar in that they both change the amount of resources in the second period—that is, for a given amount of savings K_2 they are identical. Comparative statics in both exercises, however, are different in that they may induce different effects on savings. In the exercises that follow we assume that there are no return shocks, and $R_2(1) = R_2(2) = 1$.

6.3 Effects of Permanent Differences in G

We first consider a comparative static exercise of an increase in government expenditures. Suppose we increase $G_1 = G_2(1) = G_2(2) = 0.2$, i.e., there is no aggregate uncertainty. Figure 5.10 shows labor wedges for this case. We plot in bold the benchmark case of no government expenditures, $G_1 = G_2(1) = G_2(2) = 0$, and using thin lines the case of $G_1 = G_2(1) = G_2(2) = 0.2$ (solid lines correspond to the first period distortion; dashed lines—to the second period distortion of the low types; and dotted lines—to the second period distortion of the high types).

We see that higher G leads to higher labor wedges. Intuitively, if the wedge schedule were not changed then higher expenditure would lead to lower average consumption and higher labor. Relative differences in consumption would become larger and increase the desire for redistribution, given our constant relative risk aversion specification of preferences.

In the figure 5.11 we plot the intertemporal wedges for the case with government expenditures (thin line) and for the case of no government expenditures (bold line). As in the case of labor wedges, we see that the size of the wedge is higher in the case of government expenditures.

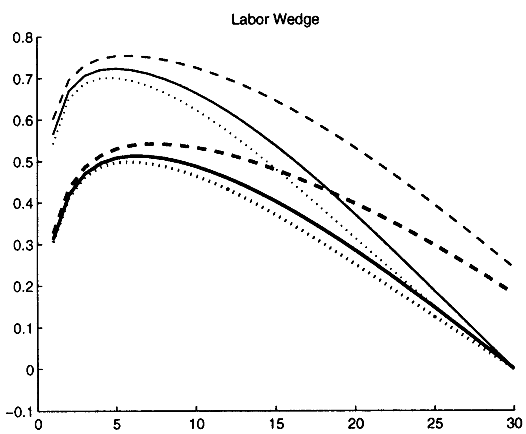

Figure 5.10
Labor Distortion

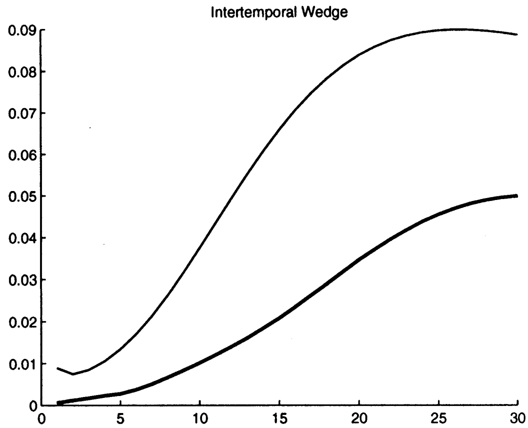

Figure 5.11
Intertemporal Distortion

We could have considered a case of transitory changes in government expenditures, i.e., keep government expenditure deterministic but make it higher or lower in the second period versus the first. This case is very similar to the one above, given our simple linear savings technology as it is the present value of government expenditures that matters, rather than the distribution of them across time.

6.4 Effects of Aggregate Shocks to Government Expenditures

We now consider the effects of stochastic shocks to government expenditures. In this specification we have $G_1 = 0.2$, $G_2(1) = 0.3$, $G_2(2) = 0.2$, and $\mu(1) = 0.7$; $\mu(2) = 0.3$. In figure 5.12 we plot labor wedges. The solid line is $\tau_{y_1}(\cdot)$ the dotted line is $\tau_{y_2}(\cdot, 1, 1)$ (i.e., high type in state 1); the dashed line is $\tau_{y_2}(\cdot, 2, 1)$ (i.e., low type in state 1); the dotted line with thick dots is $\tau_{y_2}(\cdot, 1, 2)$ (i.e., high type in state 2); the dashed line with thick dots is $\tau_{y_2}(\cdot, 2, 2)$ (i.e., low type in state 2).

The most important observation is that there is a difference in taxes across realizations of government expenditure. This contradicts one interpretation of perfect tax smoothing, which would lead one to expect wedges to remain constant across these shocks. This finding is new to both the literature on dynamic Mirrlees taxation and to the Ramsey taxation literature. For example, Ramsey models call for smoothing labor tax distortions across states of the economy. As reviewed in sub-

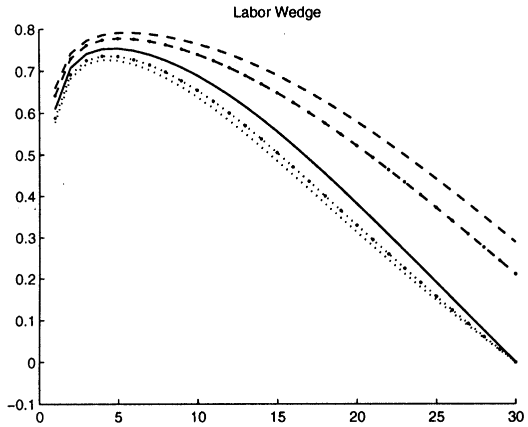

Figure 5.12
Shocks to Government Expenditure

section 4.2, without unobservable idiosyncratic shocks, tax smoothing also obtains in a Mirrleesian model.

Interestingly, the distortions do not move in the same direction for the low and high types. This is in contrast to the comparative static exercise in figure 5.10, where lower government expenditure leads to lower taxes overall. Here, instead, the spread between the distortions on the low and high types becomes smaller when government expenditures are low. Our intuition is that when government expenditure is low, resources are more abundant. As a consequence, the contribution to output from labor, the source of inequality becomes relatively smaller. Thus, insuring the new skill shocks becomes less valuable. The economy then behaves closer to the benchmark where there are no new skill shocks, where perfect tax smoothing obtains.

We now turn to figure 5.13, which shows the intertemporal distortion. In that figure, the upper (dashed) line is $\mu_1 = 0.7$, the solid line is $\mu_1 = 0.5$ and the lower (dotted) line is $\mu_1 = 0.3$.

We see that intertemporal wedge becomes higher with higher μ_1.

6.5 Effects of Rate of Return Shocks

In this section we consider the effects of shocks to returns. We consider a case in which $R_2(1) = 1$ and $R_2(2) = 4$. In figure 5.14 we plot labor distortions. We plot labor wedges as follows. The solid line is $\tau_{y_1}(\cdot)$ the

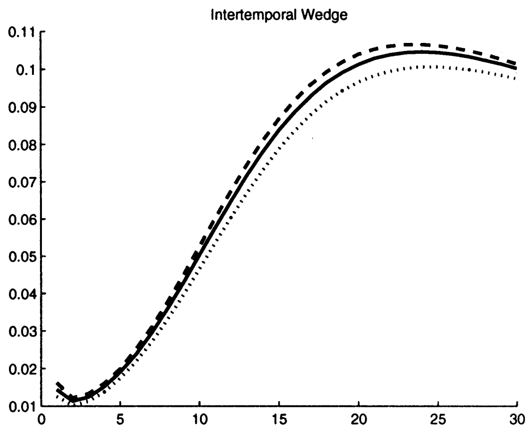

Figure 5.13
Intertemporal Distortion

Figure 5.14
Rate of Return Shocks

dotted line is $\tau_{y_2}(\cdot, 1, 1)$ (i.e., wedge for the high shock type in state 1); the dashed line is $\tau_{y_2}(\cdot, 2, 1)$ (i.e., wedge for the low type in state 1); the dotted line with thick dots is $\tau_{y_2}(\cdot, 1, 2)$ (i.e., wedge for the high type in state 2); the dashed line with thick dots is $\tau_{y_2}(\cdot, 2, 2)$ (i.e., wedge for the low type in state 2).

As in the case of government expenditure shocks, here we also observe that the spread between wedges on low and high type in a bad state are higher.

We now turn to the analysis of the behavior of the capital wedge under aggregate uncertainty. Figure 5.15 plots the intertemporal distortion τ_k for various values of the shock to the rate of return: $R_2 = 1$ (solid line – the benchmark case of no uncertainty) and $R_2(2) = 1.2, 2, 3$, and 4 (dotted lines).

We see that distortions decrease with the rate of return shock R_2. Intuitively, a higher R leads to more resources, and with more resources the planner can distribute them in a way that reduces the relative spread in consumption, making the desire for redistribution lower (given our CRRA preferences) and thus, lowering the need to distort. We also explored the effects of upwards shocks for $R_2(2) = 1, 1.2, 2, 3$, and 4 on labor distortions. Qualitatively, they are similar to the ones in figure 5.14.

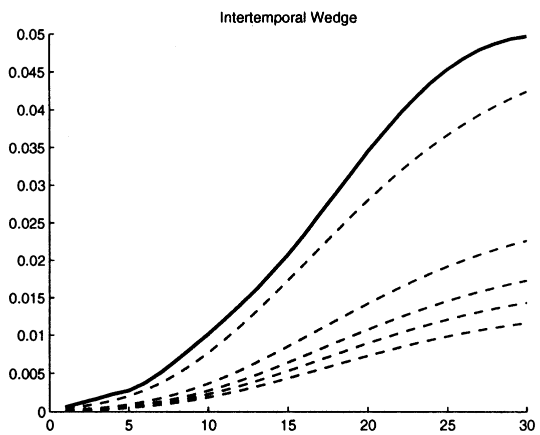

Figure 5.15
Intertemporal Distortion Varying R_2

6.6 Summary

We can now summarize the main implications of our analysis. There are two main points to take away from this section: (1) aggregate shocks lead to labor wedges differing across shocks, and (2) a positive aggregate shock (either a higher return on savings or lower realization of government expenditures) leads to lower capital wedges and to a lower spread between labor wedges.

7 Concluding Remarks

In this paper we reviewed some main results from the recent *New Dynamic Public Finance* literature. We also provided some novel explorations in the determinants of capital and labor wedges, and how these wedges respond to aggregate shocks.

We also argued that this approach not only provides a workable alternative to Ramsey models, but that it also comes with several significant advantages over its predecessor. First, while Ramsey models have provided several insights into optimal policy, their well-understood limitation regarding the ad hoc nature of tax instruments, may make interpreting their prescriptions problematic. In contrast, the main premise of the Mirrleesian approach is to model heterogeneity or uncertainty—creating a desire for insurance or redistribution—and an informational friction that prevents the first-best allocation and determines the set of feasible tax instruments endogenously. In particular, although a simple non-discriminatory lump-sum tax component is never ruled out, the optimum features distortions because these improve redistribution and insurance. Second, we also argued that this approach has novel implications for the type of dynamic policy issues that macroeconomists have been interested in: capital taxation, smoothing of labor income taxes, and the nature of the time-consistency problem. In addition, some new issues may arise directly from the focus on richer tax instruments—such as the progressivity of taxation.

In what follows we outline what we think are largely unresolved questions that we hope are explored in future research.

One remaining challenge is the quantitative exploration of the theory using calibrated models that can capture some empirically relevant features of skill dynamics—such as those studied in, for example, Storesletten, Telmer, and Yaron (2004). The main difficulty is that it is currently not tractable to solve multiple-period models with such a rich structure for skill shocks. Most current studies impose simplifying assumptions that provide illustrative insights, but remain unsuitable for quantitative purposes. One recent route around this problem is provided by Farhi and Werning (2006a) who study tax reforms in a dynamic Mirrleesian setting to evaluate the gains from distorting savings and provide a simple method which remains tractable even with rich skill processes. There is also some early progress in analyzing dynamic Mirrlees models with persistent shocks using a first-order approach in Kapicka (2005).

A quantitative analysis could also be used to address and evaluate the importance of a common challenge against the *New Dynamic Public Finance* literature: that it delivers tax systems that are "too complicated." For example, one could compare the level of welfare obtained with the fully optimal scheme to that which is attained when some elements of the tax system are simplified. For example, it may be interesting to compute the welfare losses from a tax code close to the one in the United States and other countries, or other systems with limited history dependence.

A related route is to take insights into the nature of optimal taxation from Mirrleesian models and incorporate them in a simplified fashion in Ramsey-style models, augmented with heterogeneity and idiosyncratic uncertainty regarding skills. The work by Conesa and Krueger (2005) and Smyth (2005) may be interpreted as a step in this direction. These papers compute the optimal tax schedule in a model where the tax function is arbitrarily restricted but flexibly parameterized to allow for wide range of shapes, including progressive taxation. Work along these lines, using state-of-the-art computational models, could explore other tax features, such as certain differential treatments of capital and labor income, or some forms of history dependence.

Another quantitative direction for research is to consider the implications of the new approach for classic macroeconomic questions, such as the conduct of fiscal policy over the business cycle. We only perfunctorily touched on this topic, but there is much more to be done to consider many of the issues that macroeconomists studied in the Ramsey traditions. Ideally, one could derive a rich set of quantitative predictions, similar in spirit to the quantitative Ramsey analysis in Chari, Christiano, and Kehoe (1994).

The main reason we stress the potential value of quantitative work is as follows. In our view, the approach to optimal taxation pioneered by Mirrlees (1971) and Atkinson and Stiglitz (1976) was seen as extremely promising in the '70s and early '80s, but received relatively less applied interest later. One common explanation for this is that the approach made quantitative and applied work difficult and demanding. We hope that, this time around, the recent surge in interest, combined with the more advanced quantitative techniques and computing power available today, may soon create enough progress to make solving realistic quantitative models feasible. Recent quantitative work is promising in this regard (e.g., Golosov and Tsyvinski 2006a, Farhi and Werning 2006a), but more is needed.

Another direction for future research is to relax the assumption of mechanisms operated by benevolent social planners. A relevant question in this context is whether the normative insights of the dynamic Mirrlees literature apply to the positive real-world situations where politicians care about reelection, self-enrichment or their own individual biases, and cannot commit to sequences of future policies. A related question is under what conditions markets can be better than optimal mechanisms. The potential misuse of resources and information by the government may make mechanisms less desirable relative to markets. Certain allocations resulting from anonymous market transactions cannot be achieved via centralized mechanisms. Nevertheless, centralized mechanisms may be preferable to anonymous markets because of the additional insurance they provide to risk-averse agents. Acemoglu, Golosov, and Tsyvinski (2006) approach these questions with a model that combines private information regarding individual skill types with the incentive problems associated with self-interested rulers.

Finally, we close by emphasizing that the *New Dynamic Public Finance* approach can be used to analyze a large variety of new topics, rarely explored within Ramsey settings. For instance, one recent line of research focuses on intergenerational issues. Phelan (2005) and Farhi and Werning (2007) consider how intergenerational incentives should be structured, while Farhi and Werning (2006b) and Farhi, Kocherlakota, and Werning (2005) derive implications for optimal estate taxation. This is just one example of how this approach promises more than just new answers to old questions, but also leads to new insights for a large set of unexplored questions.

Acknowledgments

For comments and suggestions we thank Daron Acemoglu, V. V. Chari, Peter Diamond, Kenneth Judd, James Mirrlees, and Michael Woodford.

Endnotes

1. However, see Diamond and Mirrlees (1978, 1986, 1995) for early work with dynamic economies with private information.

2. Judd (1999) extends the analysis to cover cases where no steady state may exist.

3. Aiyagari et al. (2002) and Werning (2005) study tax-smoothing of labor income taxes when markets are incomplete. Farhi (2005) studies capital income taxation and ownership in this context.

4. See also Diamond, Helms, and Mirrlees (1980) for an early quantitative study of models in which taxes are not linear.

5. A few papers have departed from the representative-agent setting. For example, the analysis of optimal capital taxation in Judd (1985) allowed some forms of heterogeneity.

6. One exception is Werning (2005a) who studies tax smoothing and capital taxation in a model with heterogeneous agents subject to aggregate shocks. Another one is Kocherlakota (2005) who extends the inverse Euler equation to the case of aggregate uncertainty and includes a numerical illustration of the optimum with two skill types.

7. See also Kingston (1991) and Zhu (1992) for perfect tax smoothing results within a representative agent Ramsey economy with proportional taxation.

8. See also Brito et al. (1991).

9. See Kocherlakota (2006) for another review of the literature.

10. It is straightforward to extend the model by allowing the third period to explicitly distinguish retired individuals from older workers. Indeed, if we assume no labor decision in the third period, nothing is lost by ignoring it and lumping consumption into the second period, as we implicitly do here.

11. The *Revelation Principle* guarantees that the best allocations can always be achieved by a mechanism where workers make reports about their types to the planner.

12. See Diamond (1998) and Tuomala (1990) for how the choice of the welfare function affects optimal taxes in static framework.

13. That is, we use that $\mathbb{E}[1/x] > 1/\mathbb{E}[x]$ when $Var(x) > 0$, where x in our case is the marginal utility $u'(c_2(i, j))$.

14. For example, if an agent considers changing her labor, then, in general, she also considers changing her savings. Golosov and Tsyvinski (2006a), Kocherlakota (2005) and Albanesi and Sleet (2006) showed that such *double deviations* would give an agent a higher utility than the utility from the socially optimal allocations, and therefore the optimal tax system must be enriched with additional elements in order to implement the optimal allocations.

15. A formalization of such a game and an equilibrium concept, sustainable equilibrium, is due to Chari and Kehoe (1990). They formulate a general equilibrium infinite horizon model in which private agents are competitive, and the government maximizes the welfare of the agents. Benhabib and Rustichini (1997), Klein, Krusell, and Rios–Rull (2005), Phelan and Stacchetti (2001), and Fernandez-Villaverde and Tsyvinski (2004) solve for equilibria in an infinitely lived agent version of the Ramsey model of capital taxation.

16. See also Sleet and Yeltekin (2005) who prove similar result when agents' shocks follow an *i.i.d.* process and the government is benevolent.

17. The assumption of uniformity of distribution of skills is not innocuous. Saez (2001), provides a calibrated example of distribution of skills. Diamond (1998) also uses Pareto distribution of skills. Here, we abstract from the effects of varying the skill distribution.

18. Two notable exceptions are Kocherlakota (2005) and Werning (2005a).

19. We thank Ken Judd for pointing this to us.

References

Abraham, Arpad, and Nicola Pavoni. 2003. "Efficient Allocations with Moral Hazard and Hidden Borrowing and Lending." Mimeo, University College London.

Acemoglu, Daron, Mikhail Golosov, and Aleh Tsyvinski. 2006. "Markets versus Governments: Political Economy of Mechanisms." Mimeo, MIT.

Aiyagari, S. Rao, Albert Marcet, Thomas J. Sargent, and Juha Seppälä. 2002. "Optimal Taxation without State-Contingent Debt." *Journal of Political Economy* 110(6): 1220–1254.

Albanesi, Stefania. 2006. "Optimal Taxation of Entrepreneurial Capital under Private Information." Mimeo, Columbia University.

Albanesi, Stefania, and Christoper Sleet. 2006. "Dynamic Optimal Taxation with Private Information." *Review of Economic Studies* 73: 1–30.

Arnott, Richard, and Joseph Stiglitz. 1990. "The Welfare Economics of Moral Hazard." In H. Louberge, ed., *Information and Insurance: Essays in Memory of Karl H. Borch*, 91–121. Norwell, MA: Kluwer.

Arnott, Richard, and Joseph E. Stiglitz. 1986. "Moral Hazard and Optimal Commodity Taxation." *Journal of Public Economics* 29: 1–24.

Atkinson, Anthony B., and Joseph E. Stiglitz. 1976. "The Design of Tax Structure: Direct vs. Indirect Taxation." *Journal of Public Economics* 6: 55–75.

Barro, Robert. 1979. "On the Determination of the Public Debt." *Journal of Political Economy* 87(5): 940–971.

Battaglini, Marco, and Stephen Coate. 2005. "Pareto Efficient Taxation with Stochastic Abilities." Mimeo, Cornell and Princeton.

Benhabib, Jess, and Aldo Rustichini. 1997. "Optimal Taxes without Commitment." *Journal of Economic Theory* 77: 231–259.

Bisin, Alberto, and Adriano Rampini. 2006. "Markets as Beneficial Constraints on the Government." *Journal of Public Economics* 90(4-5): 601–629.

Brito, Dagobert L., Jonathan H. Hamilton, Steven M. Slutsky, and Joseph E. Stiglitz. 1991. "Dynamic Optimal Income Taxation with Government Commitment." *Journal of Public Economics* 44(1): 15–35.

Chamley, Christophe. 1986. "Optimal Taxation of Capital in General Equilibrium." *Econometrica* 54: 607–622.

Chari, V. V., and Patrick Kehoe. 1990. "Sustainable Plans." *Journal of Political Economy* 94: 783–802.

Chari, V. V., Lawrence J. Christiano, and Patrick Kehoe. 1994. "Optimal Fiscal Policy in a Business Cycle Model." *Journal of Political Economy* 102(4): 617–652.

Conesa, Juan Carlos, and Dirk Krueger. 2005. "On the Optimal Progressivity of the Income Tax Code." *Journal of Monetary Economics* 53(7): 1425–1450.

da Costa, Carlos. 2005. "Yet Another Reason to Tax Goods." Mimeo, Getulio Vargas Foundation.

da Costa, Carlos, and Iván Werning. 2002. "Commodity Taxation and Social Insurance." Mimeo, MIT.

da Costa, Carlos, and Iván Werning. 2005. "On the Optimality of the Friedman Rule with Heterogeneous Agents and Nonlinear Income Taxation." MIT Working Paper. Cambridge, MA: MIT.

Diamond, Peter A., L. J. Helms, and James A. Mirrlees. 1980. "Optimal Taxation in a Stochastic Economy." *Journal of Public Economics* 14: 1–29.

Diamond, Peter A. 1998. "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates." *American Economic Review* 88(1): 83–95.

Diamond, Peter A., and James A. Mirrlees. 1978. "A Model of Social Insurance with Variable Retirement." *Journal of Public Economics* 10(3): 295–336.

Diamond, Peter A., and James A. Mirrlees. 1986. "Payroll-Tax Financed Social Insurance with Variable Retirement." *Scandinavian Journal of Economics* 88(1): 25–50.

Diamond, Peter A., and James A. Mirrlees. 1995. "Social Insurance with Variable Retirement and Private Saving." Mimeo, MIT.

Farhi, Emmanuel. 2005. "Capital Taxation and Ownership when Markets Are Incomplete." Mimeo, MIT.

Farhi, Emmanuel, and Iván Werning. 2006a. "Capital Taxation: Quantitative Implications of the Inverse Euler Equation." Mimeo, MIT.

Farhi, Emmanuel, and Iván Werning. 2006b. "Progressive Estate Taxation." Mimeo, MIT.

Farhi, Emmanuel, and Iván Werning. 2007. "Inequality and Social Discounting." *Journal of Political Economy,* forthcoming.

Farhi, Emmanuel, Narayana Kocherlakota, and Iván Werning. 2005. "Estate Taxes and Estate Subsidies." Work in Progress, MIT.

Fernandez-Villaverde, Jesus, and Aleh Tsyvinski. 2004. "Optimal Fiscal Policy in a Business Cycle Model without Commitment." Mimeo, Harvard University.

Golosov, Mikhail, and Aleh Tsyvinski. 2006a. "Designing Optimal Disability Insurance: A Case for Asset Testing." *Journal of Political Economy* 114(2): 257–279.

Golosov, Mikhail, and Aleh Tsyvinski. 2006b. "Optimal Taxation with Endogenous Insurance Markets." *Quarterly Journal of Economics,* forthcoming.

Golosov, Mikhail, Narayana Kocherlakota, and Aleh Tsyvinski. 2003. "Optimal Indirect and Capital Taxation." *Review of Economic Studies* 70(3): 569–587.

Greenwald, Bruce C., and Joseph E. Stiglitz. 1986. "Externalities in Economies with Imperfect Information and Incomplete Markets." *Quarterly Journal of Economics* 101: 229–264.

Judd, Kenneth L. 1985. "Redistributive Taxation in a Perfect Foresight Model." *Journal of Public Economics* 28: 59–83.

Judd, Kenneth L. 1989. "Optimal Taxation in Dynamic Stochastic Economies: Theory and Evidence." Mimeo, Stanford University.

Judd, Kenneth L. 1999. "Optimal Taxation and Spending in General Competitive Growth Models." *Journal of Public Economics* 71(1): 1–26.

Kapička, Marek. 2005. "Efficient Allocations in Dynamic Private Information Economies with Persistent Shocks: A First Order Approach." Mimeo, University of California-Santa Barbara.

Kingston, Geoffrey. 1991. "Should Marginal Tax Rates be Equalized Through Time?" *Quarterly Journal of Economics* 106(3): 911–24.

Klein, Paul, Per Krusell, and Jose-Victor Rios-Rull. 2005. "Time-Consistent Public Expenditures." Mimeo, University of Pennsylvania.

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

Kocherlakota, Narayana. 2006. "Advances in Dynamic Optimal Taxation." In *Advances in Economics and Econometrics,* Theory and Applications Ninth World Congress, Volume I, 269–299. (New York, NY: Cambridge Univ. Press).

Lucas, Robert E., Jr., and Nancy L. Stokey. 1983. "Optimal Fiscal and Monetary Policy in an Economy without Capital." *Journal of Monetary Economics* 12: 55–93.

Mirrlees, James A. 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38(2): 175–208.

Phelan, Christopher. 2005. "Opportunity and Social Mobility." *Federal Reserve Bank of Minneapolis, Research Department Staff Report* 323.

Phelan, Christopher, and Ennio Stacchetti. 2001. "Sequential Equilibria in a Ramsey Tax Model." *Econometrica* 69(6): 1491–1518.

Prescott, Edward C., and Robert Townsend. 1984. "Pareto Optima and Competitive Equilibria with Adverse Selection and Moral Hazard." *Econometrica* 52: 21–45.

Roberts, Kevin. 1984. "Theoretical Limits to Redistribution." *Review of Economic Studies* 51: 177–195.

Rogerson, William P. 1985. "Repeated Moral Hazard." *Econometrica* 53(1): 69–76.

Saez, Emmanuel. 2001. "Using Elasticities to Derive Optimal Income Tax Rates." *Review of Economic Studies* 68(1): 205–229

Shimer, Robert, and Iván Werning. 2005. "Liquidity and Insurance for the Unemployed." Mimeo, MIT.

Sleet, Christopher, and Sevin Yeltekin. 2005. "Credible Social Insurance." Mimeo, Carnegie-Mellon University.

Smyth, S. 2005. "A Balancing Act: Optimal Nonlinear Taxation of Labor and Capital in Overlapping Generation Economies." Mimeo, Harvard University.

Stiglitz, Joseph. 1987. "Pareto Efficient and Optimal Taxation and the New New Welfare Economics." In A. Auerbach and M. Feldstein, eds, *Handbook on Public Economics* 991–1042. North Holland: Elsevier Science Publishers.

Storesletten, Kjetil, Chris I. Telmer, and Amir Yaron. 2004. "Cyclical Dynamics in Idiosyncratic Labor Market Risk." *Journal of Political Economy* 112(3): 695–717.

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

Werning, Iván. 2002. "Optimal Dynamic Taxation." PhD Dissertation, University of Chicago.

Werning, Iván. 2005. "Tax Smoothing with Incomplete Markets." Mimeo, MIT.

Werning, Iván. 2007. "Optimal Fiscal Policy with Redistribution." *Quarterly Journal of Economics,* forthcoming.

Zhu, Xiaodong. 1992. "Optimal Fiscal Policy in a Stochastic Growth Model." *Journal of Economic Theory* 58: 250–89.

Appendix: Numerical Approach

In this appendix we describe the details of the numerical computations that we performed in this paper. The major conceptual difficulty with computing this class of models is that there are a large number of incentive constraints, and there is no result analogous to static models that guarantee that only local incentive compatibility constraints can bind to reduce them. Our computational strategy in this regard is as follows:

1. We start with solving several examples in which we impose all of the IC constraints. This step gives us a conjecture on what kind of constraints may bind.
2. We then impose constraints that include deviations that bind in step 1. In fact, we include a larger set that also includes constraints in the neighborhood (of reporting strategies) to the ones that bind.
3. Finally, once the optimum is computed we check that no other constraints bind.

This approach is very much like the active set approach in constrained optimization: one begins with a set of constraints that are likely to be the binding ones, one then solves the smaller problems, checking all constraints, and adding the constraints that are violated in the set of constraints that are considered for the next round (and possibly dropping some of those that were not binding) and repeat the procedure.[19]

Comment

Peter Diamond, *Massachusetts Institute of Technology and NBER*

1 Introduction

This interesting and stimulating paper (referred to as GTW) discusses four issues: when capital should not be taxed, when labor taxes should be constant over time and states of nature, the sources of concern about limited government commitment, and the methodology of modeling for tax analysis. And it contains calculated examples. I will touch on three of these issues, leaving out the complex issue of how policy feasibility and desirability are influenced by the nature of the political process in democratic states. In the macro tradition, the analysis focuses on settings with stochastic shocks. To bring a public economics perspective, I will consider the first two issues in deterministic models with heterogeneous populations. Then I will consider a stochastic model to add to the intuition about taxing savings. For clarity of presentation, I work with models with only two types of workers and assume that the binding incentive compatibility constraint is that type-A not imitate type-B. I do not consider sufficient conditions for this pattern of constraints to be correct.

2 Taxing Savings

Atkinson-Stiglitz (1976) showed that in the presence of optimal nonlinear earnings taxes, it was not optimal to also use distorting linear consumption taxes, provided that all consumer preferences are separable between goods and labor and all consumers have the same subutility function of consumption. Laroque (2005) and Kaplow (2006) have extended this result, showing that with the same preference assumptions, in the presence of any income tax function that gives rise to an equilibrium, if there are distorting consumer taxes, then a move to nondistorting consumer taxes can be done along with a permutation of the income tax that leaves every consumer with the same utility and the same labor supply, while the government collects more revenue. If labor supply is smooth with uniform transfers to all consumers (no jumps in labor supply), then this revenue gain can be used to make a Pareto improvement.

GTW explore this issue by solving a social welfare optimization with quantities as control variables and incentive compatibility constraints as well as a resource constraint or constraints (if there is uncertainty about aggregate resources). Then, they compare the MRS between first- and second-period consumptions at the optimal allocation to the MRT. The comparison allows calculation of the "wedge" between them, reflecting implicit marginal taxation of savings. They consider two other wedges—between consumption and earnings in each of the two periods, reflecting the implicit marginal taxation of earnings. They compare these two labor wedges to find conditions where earnings are marginally taxed the same in both periods. These labor wedges are also examined with aggregate uncertainty about the resource constraint in order to compare wedges across states of nature. The comparison of labor wedges across periods is really a fourth wedge—between earnings in the two periods. That is, in this four-good model there are two separate own rates of interest—in earnings and in spending.

The Atkinson-Stiglitz condition for non-use of distorting consumption taxes has naturally received a great deal of attention, particularly with the interpretation of present and future consumption goods and so the taxation of savings. That is, under these assumptions, using the vocabulary of GTW, there is no wedge between MRS and MRT for consumptions in different periods. With no wedge for intertemporal consumption, unless the implicit marginal taxation of earnings is constant over time, there is a nonzero wedge between earnings in different periods. Below I will offer a simple example of an optimal model with no wedge in intertemporal consumption but a wedge in intertemporal earnings, that is, non-constant marginal taxation of earnings.

Despite the great interest in the Atkinson-Stiglitz result, there remain arguments in favor of taxing savings with nonlinear earnings taxes. One obvious argument would be that preferences do not exhibit the separability between consumption and labor used in the theorem. Then the Corlett-Hague (1953) style analysis in a 3-good model (current work, current consumption, and future consumption) can examine whether a move towards taxing savings or towards subsidizing savings raises welfare. But we do not know much about the relevant cross-elasticities, although the commonly-used assumptions of atemporal and intertemporal separability strike me as implausible.

Another argument for taxing savings, one that is based closely on empirical observations, is due to Saez (2002). He argues that there is a positive correlation between labor skill level (wage rate) and the savings rate. In a two-period certainty setting with additive preferences, this is consistent with those with higher earnings abilities having less discount of future consumption. In terms of the conditions of the Atkinson-Stiglitz theorem, Saez preserves separability but drops the assumption that the subutility function of consumption is the same for everyone. I begin my formal analysis (echoing Diamond, 2003) by showing this result in a two-types model with labor only in the first period, illustrating the Atkinson-Stiglitz result at the same time.

Consider the following social welfare function optimization. Assume full nonlinear taxation, and two types of households, with the only binding incentive compatibility constraint being type A considering imitating type B. I do not analyze sufficient conditions for this to be the only binding constraint.

$$\text{Maximize}_{c,y} \quad \sum_i \pi_i (u[c_1(i)] + \beta_i u[c_2(i)] - v[y_1(i) / \theta_1(i)])$$

subject to:

$$G + \sum_i \pi_i (c_1(i) + R^{-1} c_2(i) - y_1(i)) \leq 0 \tag{1}$$

$$u[c_1(A)] + \beta_A u[c_2(A)] - v[y_1(A)/\theta_1(A)] \geq$$

$$u[c_1(B)] + \beta_A u[c_2(B)] - v[y_1(B)/\theta_1(A)]$$

with notation

$c_j(i)$	consumption in period j of household i
$y_j(i)$	earnings in period j of household i
$\theta_j(i)$	skill in period j of household i
π_i	number of workers of type i
β_i	discount factor of household i
R	1 plus the return to capital
G	government expenditures
λ, ψ	LaGrange multipliers

This problem has the FOCs for consumption levels:

$$(\pi_A + \psi)\, u'[c_1(A)] = \lambda \pi_A \tag{2}$$

$$(\pi_A + \psi)\, \beta_A u'[c_2(A)] = \lambda \pi_A R^{-1} \tag{3}$$

$$(\pi_B - \psi)\, u'[c_1(B)] = \lambda \pi_B \tag{4}$$

$$(\pi_B \beta_B - \psi \beta_A)\, u'[c_2(B)] = \lambda \pi_B R^{-1} \tag{5}$$

Taking the ratio of FOCs for A, there is no tax on savings on the high type:

$$\frac{u'[c_1(A)]}{u'[c_2(A)]} = \beta_A R. \tag{6}$$

This is the familiar no-marginal-taxation condition at the very top of the earnings distribution.

Now let us turn to type B. Taking the ratio of FOCs we have

$$\frac{(\pi_B - \psi)u'[c_1(B)]}{\left(\pi_B - \psi \dfrac{\beta_A}{\beta_B}\right)u'[c_2(B)]} = \beta_B R. \tag{7}$$

The plausible case is that high earners have a lower discount of future consumption, $\beta_A < \beta_B$, resulting (with $\pi_B - \psi > 0$) in

$$\frac{u'[c_1(B)]}{u'[c_2(B)]} < \beta_B R. \tag{8}$$

That is, type-B would save if that were possible at zero taxation of savings, so there is implicit marginal taxation of savings. If and only if $\beta_A = \beta_B$ does this imply no taxation of savings for type B. Saez does his analysis with linear taxation of savings and concludes that since higher earners have higher savings rates, taxing savings is part of the optimum.

The GTW exploration of the taxation of savings focuses on uncertainty about future earnings as a source of the desirability of taxation of savings. It is true that people are uncertain about future earnings. It is also true that people differ in discount rates. The case for not taxing savings does not survive either issue with plausible characterizations.

3 Earnings Tax Smoothing

With uncertainty about future earnings, different workers will realize different age-earnings profiles and this uncertainty can require varying implicit taxes on earnings over time (over worker ages). In contrast, GTW show tax smoothing when everyone has the same age-earnings profile and the disutility of labor is a power function. A failure of tax smoothing also comes without uncertainty if we allow different age-earnings profiles for different workers. In this example, there is no wedge on the intertemporal consumption decision. However, there are different consumption-earnings wedges in the two periods and so a wedge on the intertemporal earnings decision.

With the same notation as above, consider a two-types model with two periods of earnings and the only binding incentive compatibility constraint that type-A not want to imitate type-B, with that imitation done for the entire life.

$$\text{Maximize}_{c,y} \qquad \Sigma\pi_i (u\,[c_1(i)] + \beta u[c_2(i)] - v[y_1(i)/\theta_1(i)] - \beta v[y_2(i)/\theta_2(i)])$$

subject to:

$$G + \Sigma\pi_i(c_1(i) + R^{-1}c_2(i) - y_1(i) - R^{-1}y_2(i)) \leq 0 \tag{9}$$

$$u[c_1(A)] + \beta u[c_2(A)] - v[y_1(A)/\theta_1(A)] - \beta v[y_2(A)/\theta_2(A)] \geq$$

$$u[c_1(B)] + \beta u[c_2(B)] - v[y_1(B)/\theta_1(A)] - \beta v[y_2(B)/\theta_2(A)]$$

From the FOCs for consumption levels, there is no tax on savings:

$$\frac{u'[c_1(B)]}{u'[c_2(B)]} = \beta R = \frac{u'[c_1(A)]}{u'[c_2(A)]} \tag{10}$$

Now consider the FOCs for earnings:

$$(\pi_A + \psi)\, v'[y_1(A)/\theta_1(A)]/\theta_1(A) = \lambda\pi_A \tag{11}$$

$$(\pi_A + \psi)\, \beta v'[y_2(A)/\theta_2(A)]/\theta_2(A) = \lambda\pi_A R^{-1} \tag{12}$$

$$\pi_B v'[y_1(B)/\theta_1(B)]/\theta_1(B) - \psi v'[y_1(B)/\theta_1(A)]/\theta_1(A) = \lambda\pi_B \tag{13}$$

$$\pi_B \beta v'[y_2(B)/\theta_2(B)]/\theta_2(B) - \psi\beta v'[y_2(B)/\theta_2(A)]/\theta_2(A) = \lambda\pi_B R^{-1} \tag{14}$$

Taking a ratio of FOCs, there is no intertemporal earnings wedge for the high type, consistent with no-marginal-taxation of the highest type on all margins:

$$\frac{v'[y_1(A)/\theta_1(A)]/\theta_1(A)}{v'[y_2(A)/\theta_2(A)/\theta_2(A)} = \beta R \tag{15}$$

Turning to type B, let us define Δ as the wedge:

$$\frac{v'[y_1(B)/\theta_1(B)]/\theta_1(B)}{v'[y_2(B)/\theta_2(B)]/\theta_2(B)} + \Delta = \beta R \tag{16}$$

If Δ is negative, then the first period marginal disutility of earning is larger than the discounted second period marginal disutility.

From the ratio of FOCs the sign of Δ depends on the difference in intertemporal MRS for type B and for type A if imitating type B:

$$\Delta = \frac{\pi_B v'[y_1(B)/\theta_1(B)]/\theta_1(B) - \psi v'[y_1(B)/\theta_1(A)]/\theta_1(A)}{\pi_B v'[y_2(B)/\theta_2(B)]/\theta_2(B) - \psi v'[y_2(B)/\theta_2(A)]/\theta_2(A)} - \frac{v'[y_1(B)/\theta_1(B)]/\theta_1(B)}{v'[y_2(B)/\theta_2(B)]/\theta_2(B)} \tag{17}$$

with

$$\pi_B - \psi \frac{v'[y_2(B)/\theta_2(A)]}{v'[y_2(B)/\theta_2(B)]} \frac{\theta_2(B)}{\theta_2(A)} > 0$$

and

$$\psi > 0,$$

the sign of Δ is the same as that of

$$\frac{v'[y_2(B)/\theta_2(A)]}{v'[y_2(B)/\theta_2(B)]} \frac{\theta_2(B)}{\theta_2(A)} - \frac{v'[y_1(B)/\theta_1(A)]}{v'[y_1(B)/\theta_1(B)]} \frac{\theta_1(B)}{\theta_1(A)}. \tag{18}$$

If $v'[z] = z^{\alpha}$, then the sign of Δ is the same as that of

$$\frac{(y_2(B)/\theta_2(A))^{\alpha}}{(y_2(B)/\theta_2(B))^{\alpha}} \frac{\theta_2(B)}{\theta_2(A)} - \frac{(y_1(B)/\theta_1(A))^{\alpha}}{(y_1(B)/\theta_1(B))^{\alpha}} \frac{\theta_1(B)}{\theta_1(A)} \tag{19}$$

or, simplifying, that of

$$\left(\frac{\theta_2(B)}{\theta_2(A)}\right)^{\alpha+1} - \left(\frac{\theta_1(B)}{\theta_1(A)}\right)^{\alpha+1}. \tag{20}$$

Thus with power function disutility of labor and the same age-earnings profile for both types, we have tax smoothing (as in Werning 2005). But tax smoothing requires the same age-earnings profile for everyone. If higher earners have steeper age-earnings profiles

$$\left(\frac{\theta_2(A)}{\theta_1(A)}\right) > \left(\frac{\theta_2(B)}{\theta_1(B)}\right)$$

then Δ is negative and there is heavier marginal taxation of second-period earnings, and a wedge in the intertemporal earnings tradeoff. Without a power function, there may not be tax smoothing even with the same age-earnings profile.

4 Taxing Savings with Uncertainty

GTW explore the case for taxing savings in models with uncertainty about future productivity. I will present a simple model of that and then contrast the route to taxing savings in this model to one with fewer government controls.

With the same notation as above, consider a one-type model with uncertainty about second-period skill, but not first period skill. This is a simpler version of GTW analysis. Let π_i now stand for the probability of having skill i in the second period. We continue to assume that the only binding incentive compatibility constraint is that type-A not want to imitate type-B, which now refers only to the second period.

$$\text{Maximize}_{c,y} \qquad u[c_1] - v[y_1/\theta_1] + \Sigma\pi_i(\beta u[c_2(i)] - \beta v[y_2(i)/\theta_2(i)])$$

subject to:

$$G + c_1 - y_1 + \Sigma\pi_i(R^{-1}c_2(i) - R^{-1}y_2(i)) \le 0$$

$$\beta u[c_2(A)] - \beta v[y_2(A)/\theta_2(A)] \ge \beta u[c_2(B)] - \beta v[y_2(B)/\theta_2(A)] \tag{21}$$

This problem has the FOCs for consumption levels:

$$u'[c_1] = \lambda \tag{22}$$

$$(\pi_A + \psi)\,\beta u'[c_2(A)] = \lambda\pi_A R^{-1} \tag{23}$$

$$(\pi_B - \psi)\,\beta u'[c_2(B)] = \lambda\pi_B R^{-1} \tag{24}$$

Adding the last two equations and taking a ratio to the first equation, we have

$$\frac{u'[c_1]}{(\pi_A+\psi)u'[c_2(A)]+(\pi_B-\psi)u'[c_2(B)]}=\beta R. \tag{25}$$

In contrast, without a wedge, the individual would see a gain from savings if

$$\frac{u'[c_1]}{\pi_A u'[c_2(A)]+\pi_B u'[c_2(B)]}<\beta R. \tag{26}$$

Thus we have implicit marginal taxation of savings provided $u'[c_2(A)] < u'[c_2(B)]$, as follows from the need to have $c_2(A) > c_2(B)$, to induce type A not to imitate type B. The underlying argument does not need the additive structure of preferences, provided that preferences are such that keeping $c_2(A)$ enough larger than $c_2(B)$ to just induce the higher labor supply implies a lower marginal utility of consumption at the higher consumption level. That is, consider the condition:

$$u[c, y/\theta] = u[c', y'/\theta'] \text{ and } c > c' \text{ implies} \tag{27}$$

$$\frac{\partial u[c, y/\theta]}{\partial c} < \frac{\partial u[c', y'/\theta']}{\partial c}.$$

Then, the argument above goes through—if the binding incentive compatibility constraint is that the high skill worker not imitate the low skill worker, then the optimum has a positive wedge on intertemporal consumption. This parallels the result that Mirrlees and I have found in the special case that labor is a zero-one variable and the low skill person does not work (Diamond and Mirrlees 1978, 1986, 2000). The insight, paralleling the argument through the inverse Euler condition, is that when less future work with lower future consumption results in a higher marginal utility of consumption (and so a greater incentive to save), making savings less available eases the incentive compatibility constraint. Additivity makes this argument easy to make, but the underlying condition is plausible and has much greater generality.

To see this argument I go through the same steps as above. The optimization becomes:

$$\text{Maximize}_{c,y} \quad u[c_1, y_1/\theta_1] + \Sigma\pi_i\beta u[c_2(i), y_2(i)/\theta_2(i)]$$

subject to:

$$G + c_1 - y_1 + \Sigma\pi_i(R^{-1}c_2(i) - R^{-1}y_2(i)) \le 0 \tag{28}$$

$$\beta u[c_2(A), y_2(A)/\theta_2(A)] \ge \beta u[c_2(B), y_2(B)/\theta_2(A)]$$

This problem has the FOCs for consumption levels:

$$u_c[c_1, y_1/\theta_1] = \lambda \tag{29}$$

$$(\pi_A + \psi)\,\beta u_c[c_2(A), y_2(A)/\theta_2(A)] = \lambda\pi_A R^{-1} \tag{30}$$

$$\pi_B\beta u_c[c_2(B), y_2(B)/\theta_2(B)] - \psi\beta u_c[c_2(B, y_2(B)/\theta_2(A))] = \lambda\pi_B R^{-1} \tag{31}$$

Adding the last two equations and taking a ratio to the first equation, we have

$$\frac{u_c[c_1, y_1/\theta_1]}{(\pi_A + \psi)u_c[c_2(A), y_2(A)/\theta_2(A)] + \pi_B u_c[c_2(B), y_2(B)/\theta_2(B)] - \psi u_c[c_2(B), y_2(B)/\theta_2(A)]} = \beta R \tag{32}$$

In contrast, without a wedge, the individual would see a gain from savings if

$$\frac{u_c[c_1, y_1/\theta_1]}{\pi_A u_c[c_2(A), y_2(A)/\theta_2(A)] + \pi_B u_c[c_2(B), y_2(B)/\theta_2(B)]} < \beta R. \tag{33}$$

Thus the sign of the wedge depends on the sign of

$$\psi(u_c[c_2(A), y_2(A)/\theta_2(A)] - u_c[c_2(B), y_2(B)/\theta_2(A)]), \tag{34}$$

which is signed by the condition above. Thus in a setting where everyone is the same in the first period, a plausible condition is sufficient for a positive intertemporal consumption wedge. The insight, paralleling the argument through the inverse Euler condition, is that with this condition, less future work and lower future consumption will result in a higher marginal utility of consumption and a greater incentive to save (unless the condition is not satisfied and the impact of hours worked on the marginal utility of consumption overcomes the higher level of consumption). Easing the incentive compatibility constraint then comes from making the return to saving smaller. Additivity makes this argument easy to make, but the underlying argument has much greater generality.

GTW explore a class of nonseparable period utility functions in their numerical results. They work with the utility function $u[c, y/\theta] = (ce^{-(y/\theta)^2})^{1-\sigma}/(1 - \sigma)$. And they have many first-period productivity levels, not just one. This utility function satisfies the condition above that at equal utilities, marginal utility of consumption is higher at the consumption-labor pair that has higher consumption and labor. Their finding of a negative wedge at some skill levels comes from a direct impact of nonseparability on the desired wedge, as can be seen in the optimization in a model with first period variation and no conditional uncertainty about second period productivities.

$$\text{Maximize}_{c,y} \quad \Sigma\pi_i\,(u[c_1(i), y_1(i)/\theta_1(i)] + \beta u[c_2(i), y_2(i)/\theta_2(i)])$$

subject to:

$$G + \Sigma\pi_i(c_1(i) + R^{-1}c_2(i) - y_1(i) - R^{-1}y_2(i)) \leq 0 \tag{35}$$

$$u[c_1(A), y_1(A)/\theta_1(A)] + \beta u[c_2(A), y_2(A)/\theta_2(A)] \geq$$

$$u[c_1(B), y_1(B)/\theta_1(A)] + \beta u[c_2(B), y_2(B)/\theta_2(A)]$$

This problem has the FOCs for consumption levels:

$$(\pi_A + \psi)\, u_c\,[c_1(A), y_1(A)/\theta_1(A)] = \lambda\pi_A \tag{36}$$

$$\pi_B u_c[c_1(B), y_1(B)/\theta_1(B)] - \psi u_c[c_1(B), y_1(B)/\theta_1(A)] = \lambda\pi_B \tag{37}$$

$$(\pi_A + \psi)\, \beta u_c[c_2(A), y_2(A)/\theta_2(A)] = \lambda\pi_A R^{-1} \tag{38}$$

$$\pi_B\beta u_c[c_2(B), y_2(B)/\theta_2(B)] - \psi\beta u_c[c_2(B, y_2(B)/\theta_2(A))] = \lambda\pi_B R^{-1} \tag{39}$$

While there is no tax on savings for the high type, for the low type, we have

$$\frac{\pi_B u_c[c_1(B), y_1(B)/\theta_1(B)] - \psi u_c[c_1(B), y_1(B)/\theta_1(A)]}{\pi_B u_c[c_2(B), y_2(B)/\theta_2(B)] - \psi u_c[c_2(B, y_2(B)/\theta_2(A))]} = \beta R. \tag{40}$$

Thus the sign of the wedge,

$$\frac{u_c[c_1(B), y_1(B)/\theta_1(B)]}{u_c[c_2(B), y_2(B)/\theta_2(B)]} - \beta R$$

depends on that of

$$\frac{u_c[c_2(B), y_2(B)/\theta_2(B)]}{u_c[c_1(B), y_1(B)/\theta_1(B)]} - \frac{u_c[c_2(B), y_2(B)/\theta_2(A)]}{u_c[c_1(B), y_1(B)/\theta_1(A)]}. \tag{41}$$

Thus there can be a negative wedge for a suitable impact of additional labor in both periods on the intertemporal consumption MRS. The

GTW example with nonseparable utility and second-period uncertainty has both of these elements in it, providing both positive and negative pushes on the wedge.[1]

I have followed GTW in examining individual marginal incentives at the point of the optimal allocation assuming full government control (full observability of consumption and earnings). A similar insight comes from considering the same model except that while the government can observe savings it can not observe who is saving, implying linear taxation of savings. This decrease in observability lowers social welfare since the incentive compatibility constraint becomes more restrictive when the potential imitator can simply modify savings. A parallel result is then that the optimum includes taxation of savings, not subsidization. That is, one can see the same underlying mechanism—that savings adjustment makes the incentive compatibility constraint harder to meet and so one should discourage savings in this slightly different setting.

First, consider the individual savings problems (1) if planning to produce the output level of type-A when type-A and (2) if planning to produce the type-B output even if type-A. Note that what was previously consumption is now the net-of-tax wage. Denote the net-of-tax return on savings by Q. Define the indirect utility-of-consumption functions:

$$V_A[c_1, c_2(A), c_2(B), Q] \equiv \mathrm{Max}_s\{u\,[c_1 - s] + \pi_A \beta u[c_2(A) + Qs] + \pi_B \beta u[c_2(B) + Qs]\} \tag{42}$$

$$V_B[c_1, c_2(B), Q] \equiv \mathrm{Max}_s\{u[c_1 - s] + \beta u[c_2(B) + Qs]\}. \tag{43}$$

Note that the optimal savings levels, s^*_i, depend on the same variables as the indirect utility functions V_i. With these preferences (and much more generally) since $c_2(A) > c_2(B)$, we have $s^*_B > s^*_A$.

The social welfare maximization now becomes

$$\mathrm{Maximize}_{c,y} \qquad V_A[c_1, c_2(A), c_2(B), Q] - v[y_1/\theta_1] - \Sigma \pi_i \beta v[y_2(i)/\theta_2(i)]$$

subject to:

$$G + \Sigma \pi_i (c_1 + R^{-1}c_2(i) - y_1 - R^{-1}y_2(i)) \le s^*_A (1 - QR^{-1}) \tag{44}$$

$$V_A[c_1, c_2(A), c_2(B), Q] - \pi_A \beta v[y_2(A)/\theta_2(A)] \ge$$

$$V_B[c_1, c_2(B), Q] - \pi_A \beta v[y_2(B)/\theta_2(A)]$$

The actual collection of wealth tax revenue is irrelevant and we could have considered a constraint on Q consistent with there being no savings. After some manipulation we can sign the tax on capital income:

$$\text{sign}\,(R - Q) = \text{sign}\,(s_A^* - s_B^*) \tag{45}$$

Thus, there is a tax on savings since there would be an increase in savings if a type-A decided to imitate a type-B. (See Diamond and Mirrlees, 1982 for a special case.)

To explore tax smoothing despite an age structure of workers that prevents its optimality for a single cohort, one could examine OLG models with an assumption that taxes are period-specific and cannot be age-specific, or how age-specific taxes change over time.

5 Ramsey vs. Mirrlees

In contrasting Ramsey and Mirrlees approaches, GTW draws three distinctions. The first is that the Ramsey approach has a representative agent while the Mirrlees approach has a heterogeneous population. Since income distribution matters, this aspect of the Ramsey approach implies that Ramsey models can generate insight into influences relevant for tax policy but should not be viewed as generating answers to what taxes should be. But then I think that is true generally of models. As Alfred Marshall put it (1948, page 366):

> "it [is] necessary for man with his limited powers to go step by step; breaking up a complex question, studying one bit at a time, and at last combining his partial solutions into a more or less complete solution of the whole riddle. ... The more the issue is thus narrowed, the more exactly can it be handled: but also the less closely does it correspond to real life. Each exact and firm handling of a narrow issue, however, helps towards treating broader issues, in which that narrow issue is contained, more exactly than would otherwise have been possible. With each step ... exact discussions can be made less abstract, realistic discussions can be made less inexact than was possible at an earlier stage."

I view a "realistic discussion" as best drawing intuitively on multiple models of different aspects of a question. This is very different from taking literally the answer generated by a single model, even one viewed as the best available single model. This is especially true when the best available model is visibly highly limited in key dimensions, as is the case when a representative agent model is being analyzed for normative tax analysis.

A second distinction they draw is between linear taxes and nonlinear taxes. Since some taxes are linear in practice, it seems worthwhile to analyze how to set linear taxes as well. Since it is often the case that linear taxes operate in the presence of nonlinear ones, it is important to learn about that interaction. But not all linear taxes are in a setting where there are nonlinear taxes, making a separate analysis also worthwhile. In Massachusetts it is not constitutional to have progressive taxation of a single kind of income, apart from an exempt amount. Some would love to see the same restriction in the U.S. constitution. More generally, political economy considerations may call for restrictions in the taxes considered. I wonder if the very minor distinctions in income taxation by age of the worker in current U.S. law are not a reflection of the difficulty in setting so many tax parameters as would be needed with different income taxes for each age of a worker (or pairs of ages for a working couple). Or maybe this is just the lag of practice behind theory—as we saw in the roughly two decade lag in the United States in collecting tolls only one way on some bridges and tunnels.

The third distinction drawn by GTW is between a given, restricted set of tax tools, referred to as an ad hoc restriction, and deriving the set of tax tools from an underlying technology, asymmetric information in the Mirrlees case. I think this distinction is overdrawn. First, if we assume that for some transactions asymmetric information extends to the parties engaged in transactions, then taxation of a transaction might vary with the size of the transaction but cannot vary with the presence of other transactions. Then, nonlinear taxation based on total earnings is not feasible. Assuming that without this constraint there would be higher taxation of larger transactions, and that such taxation can be prevented by repeated transactions, then we are left with linear taxation, derived, not assumed. Second, there is the issue of administrative costs, which are assumed to be zero for observables in the Mirrlees model. We can recast asymmetric information as assuming that the administrative cost is infinite for what are otherwise labeled non-observables. This can be a helpful recasting. We could track the identity for each purchase of gasoline the way we do each payment of earnings. But that would be expensive (but becoming less so, particularly if we do not allow purchases for cash). If expensive enough, gasoline purchase should be subject to linear taxes, as they are. Having a more basic model (deriving what tax structure might otherwise be assumed) is not necessarily a virtue if the basic model has critical incompleteness.

In GTW there are two periods with a stochastic change in worker skill between the two periods. This allows taxes to be set differently in each period. But if skills evolve more rapidly than taxes are set (because of administrative costs, perhaps) then the modeling needs to recognize an explosion of types depending on all the stochastic realizations of opportunities that might occur within a year. Plausibly we are in the same basic position as with the assumption of a complete set of markets—no one can list all the states that might occur. So we can not envision trading on all of them, even apart from the cost in today's resources of preparing in this way for distant and/or low probability events, which would not be worthwhile. Just as incomplete markets are a reality, so too incomplete use of incentives is a reality. I see no reason to believe that assuming such a reality is necessarily worse than deriving it when trying to model something as complex as tax policy.

6 Concluding Remarks

It is good to have macroeconomists looking at the same issues as public finance economists. In the spirit of encouraging further complementary analysis, let me say that there is a great deal of current interest in annuities and taxation. This might appeal to macroeconomists as well. After all, as Benjamin Franklin wrote (in a letter to M. Leroy, 1789):

> "Our Constitution is in actual operation; everything appears to promise that it will last; but in this world nothing is certain but death and taxes."

Acknowledgments

I am indebted to Iván Werning for very clarifying comments, and to the National Science Foundation for financial support under grant SES 0239 080.

Endnote

1. Nonseparability over time in the utility of consumption is also plausible. Mirrlees and I (1986) explored an extreme (Leontieff) case of intertemporal nonseparability and (2000) a standard-of-living model.

References

Atkinson, A. B., and J. E. Stiglitz. 1976. "The Design of Tax Structure: Direct versus Indirect Taxation." *Journal of Public Economics* 6: 55–75.

Corlett, W. J., and D. C. Hague. 1953. "Complementarity and the Excess Burden of Taxation." *Review of Economic Studies* 21(1): 21–30.

Diamond, P. 2003. "Taxation, Incomplete Markets, and Social Security." Cambridge, MA: MIT Press.

Diamond, P., and J. Mirrlees. 1978. "A Model of Social Insurance with Variable Retirement." *Journal of Public Economics* 10: 295–336.

Diamond, P., and J. Mirrlees. 1986. "Payroll-Tax Financed Social Insurance with Variable Retirement." *Scandinavian Journal of Economics* 88(1): 25–50.

Diamond, P., and J. Mirrlees. 2000. "Adjusting One's Standard of Living: Two Period Models." In P. J. Hammond and G. D. Myles, eds., *Incentives, Organization, and Public Economics, Papers in Honour of Sir James Mirrlees*. Oxford: Oxford University Press.

Diamond, P., and J. Mirrlees. 1982. "Social Insurance with Variable Retirement and Private Saving." MIT Working Paper no. 296. 1995, unpublished. Forthcoming in *Journal of Public Economics*.

Kaplow, L. 2006. "On the Undesirability of Commodity Taxation Even When IncomeTaxation Is Not Optimal." *Journal of Public Economics* 90: 1235.

Laroque, G. 2005. "Indirect Taxation is Superfluous under Separability and Taste Homogeneity: A Simple Proof." *Economics Letters* 87: 141–144.

Marshall, A. 1948. *Principles of Economics*, eighth edition. New York: The Macmillan Company.

Saez, E. 2002. "The Desirability of Commodity Taxation under Non-Linear Income Taxation and Heterogeneous Tastes." *Journal of Public Economics* 83: 217–230.

Werning, I. 2005. "Tax Smoothing with Redistribution." Federal Reserve Bank of Minneapolis, Staff Report 365.

Comment

Kenneth L. Judd, *Hoover Institution and NBER*

Professors Golosov, Tsyvinski, and Werning have given us an excellent overview of recent work applying the Mirrlees (1971) approach for income taxation to questions in the theory of taxation of dynamic and stochastic environments. I am delighted to see this renewed interest in optimal taxation problems. The work discussed in this paper shows us that there is great value in this effort and also how much is left to be done.

My comments will focus on three issues. First, I will comment on the relationship between this work and the earlier literature. Second, I want to discuss a possibly heretical interpretation of Mirrlees work. Third, I will discuss the problems facing future work.

This literature has worked to emphasize the difference between the dynamic Mirrlees literature and the Ramsey literature. In particular, these papers often interpret any difference between a marginal rate of substitution and the corresponding marginal rate of transformation as a tax. However, the dynamic Mirrlees approach is not strictly comparable to the Ramsey approach. In the Ramsey approach, as executed in, for example, Atkinson and Stiglitz (1976), Diamond and Mirrlees (1971), and Judd (1985, 1999) assume that a full set of private markets exist and that prices are determined competitively. Even Mirrlees (1971) assumes that workers are paid their marginal product, implicitly assuming that there is no market power in labor markets. In these analyses, taxes are then chosen to distort market outcome so as to accomplish a reallocation of resources desired by the government. In the dynamic Mirrlees approach outlined in this paper, there are no private markets for insurance and government policy is used for both conventional purposes of raising revenue for government expenditures and redistribution, as well as to replace, or at least offer a substitute for, the missing private markets.

This point is acknowledged in this paper. The authors are often careful to refer to distortions as wedges, staying away from the question of what they signify. Section 4.5 correctly argues that in many cases private markets will attain a constrained Pareto allocation, and that these private outcomes will have many of the same wedges often called taxes in the dynamic Mirrlees literature. If the government does not enjoy an advantage in either transaction costs or information, then no government policy can attain a Pareto superior allocation.

This does not mean that the dynamic Mirrlees approach as executed so far has no value. The point here is that we should do as Mirrlees did, assume that private markets work, and then find the policy that best achieves the goal taking into account the presence of a private market. I suspect that this is a much more difficult problem, explaining why this path has not been taken, but the insights in the work summarized in this paper will help us tackle the more complex problem.

This paper makes the common assertion that Mirrlees endogenized the tax instruments by basing his analysis on an informational friction; more specifically, Mirrlees assumed that the government could observe income but could not observe either hours or wages. This is argued to be superior to "starting with an exogenously restricted set of tax instruments." I disagree with this characterization of Mirrlees (1971). In fact, wages and hours are not only observable but are often used by the government. Many workers punch a time clock, recording when a worker begins his work and when he finishes, and his income is the product of the measured hours and a wage rate known to both worker and employer. If wages and hours could not be observed then we could have neither minimum wage laws nor laws regarding overtime pay.

Of course, wages and hours would be difficult to measure for many individuals, and impossible for some occupations such as professors. However, ignoring the wage and hours information that could be obtained cheaply is particularly odd in any analysis, such as Mirrlees (1971), where the objective is to shift money to the poor since they are the ones more likely to have jobs with easily observed wages and hours.

For these reasons, I do not view Mirrlees' analysis as an explanation why we have income taxation instead of, say, lump sum taxes. We do not need information economics to understand why taxes need to be different for people with different abilities to pay. The key accomplishment in Mirrlees (1971) is that he did not restrict the functional form of the tax policy. He made the exogenous assumption that taxes depended only on income but avoided any further simplification such as linearity. The asymmetric information story is useful as a way to motivate the search for the optimal nonlinear tax schedule, and is a story that may apply in some tax problems and mechanism design problems, but we should not take it literally in these tax models.

I conjecture that the commodity tax literature can be similarly motivated. That literature, typified by Diamond and Mirrlees, assumes that different goods are taxed at different rates but that for each good all individuals pay the same constant marginal tax rate. If the government can observe only transactions, not final consumption, and cannot keep track of each individual's participation in each transaction, then any nonlinearity in the tax system would be a source of arbitrage profits. Therefore, it is likely that the only feasible tax system would have constant tax rates. In fact, most countries have a hybrid system where they do not attempt to measure each individual transaction except in the case of the labor and capital markets where the monitoring costs are moderate.

This reinterpretation is important because it frees us from unnecessary constraints on the models we look at. There is currently a kind of orthodoxy that tries to draw a sharp line between models with exogenous and endogenous institutions, arguing that the latter is obviously better. However, a closer examination of the problem, such as in this tax case, reveals shades of gray. It is not clear which is better: An analysis that exogenously specifies a set of policy instruments corresponding to the ones we see used, or using false assumptions about informational costs in order to derive an endogenous set of instruments. Tax problems like the ones examined in this paper quickly become extremely complex. Demanding analyses with fully endogenous sets of instruments will severely limit the range of problems we can examine.

The models discussed in this paper are obviously limited in many ways. In particular, there are too few periods in the dynamic dimension and there is usually no capital accumulation. There is great potential in this literature but only if we address the mathematical difficulties. We must give up focusing on simple problems that can be solved analytically or characterized in simple ways and exploit computational tools if we are to attain quantitatively substantive results. This won't be easy. For example, the numerical approach used in this paper is indicative of the challenges that we face when we move beyond the simple models. In particular, the optimal tax problem becomes multidimensional in some cases forcing the authors to consider far more incentive constraints than is necessary in the usual one-dimensional models. This is because the single-crossing property that is heavily exploited in the one-dimensional literature has no analogue for even two-dimensional problems. Therefore, if there are N types of taxpayers, we need to examine N^2 incentive constraints instead of N.

Judd and Su (2006) have examined this problem in more complex cases and find cases far more challenging than the ones in this paper, and argue that the multidimensional optimal tax problem is generally far more difficult. They show that the solution to an optimal taxation problem will generally not satisfy the linear independence constraint qualification, a fact that greatly increases the difficulty of solving these problems numerically. Fortunately, the last decade has seen many advances in the field of mathematical programming with equilibrium constraints which can be applied to these problems.

Again, I congratulate the authors for their "users guide" to an approach that can potentially provide major insights into the design of rational public policy and encourage other young researchers to follow their lead.

References

Atkinson, A. B., and J. E. Stiglitz. 1976. "The Design of Tax Structure: Direct vs. Indirect Taxation." *Journal of Public Economics* 6: 55–75.

Diamond, Peter, and James A. Mirrlees. 1971. "Optimal Taxation and Public Production." *American Economic Review* 61: 8–27, 175–208.

Judd, Kenneth L. 1985. "Redistributive Taxation in a Perfect Foresight Model." *Journal of Public Economics* 28: 59–83.

Judd, Kenneth L. 1999. "Optimal Taxation and Spending in General Competitive Growth Models." *Journal of Public Economics* 71(1): 1–26.

Judd, Kenneth L., and Che-Lin Su. 2006. "Optimal Income Taxation with Multidimensional Taxpayer Types." Hoover Institution Working Paper.

Mirrlees, James A. 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38(2): 175–208.

Discussion

Iván Werning began by saying that he agreed with most of what the discussants had said. He noted that part of the discussants' comments had focused on bringing new issues to the table. He and his coauthors felt that this was exactly what was nice about their approach to the tax problem, namely that it could address issues that could not have been addressed before using the traditional Ramsey approach. Werning observed that their approach provided scope for making normative assessments on the effects of policies related to unemployment, complementing the positive analysis from the previous day's discussion on unemployment in Europe.

Werning agreed that the optimal tax systems that emerge from the class of models they studied were in some cases quite complex. With respect to this issue, he felt that there was room for a middle ground. In their view, it was essential to bring heterogeneity and skill shocks into the models. In such models, it turned out to be convenient analytically to start by studying the case where the government is only restricted by the informational friction and not in addition by restrictions on the set of tax instruments. He suggested that restrictions on tax instruments should be considered, but only after the basic models were well understood. He also noted that in some cases, the tax systems that emerged from their approach were reasonably simple, citing recent work on disability insurance by Mikhail Golosov and Aleh Tsyvinski as well as work by himself and Robert Shimer.

Golosov said that they were sympathetic to Kenneth Judd's comment that it was important to think about the interaction of private market arrangements and government policies. He said that this was the reason why they had deliberately used the term "wedges" rather than taxes in the paper. However, he emphasized that there are many circumstances where even if markets are perfectly functioning they would fail to yield efficient outcomes due to externalities.

Greg Mankiw asked Peter Diamond what the evidence was for the statement he had made that high type people are more patient. Diamond responded that the assumptions on preferences that are made in these models imply that high skilled people have higher earnings and that people who discount the future less heavily have higher savings rates. Given this, he said, the statement follows from the empirical correlation between savings rates and earnings. Mankiw responded that this correlation may be due to consumption smoothing. Diamond thought that it was unreasonable to think that consumption smoothing explained the entire correlation.

James Poterba remarked that the paper had potential implications for the design of the tax period. He observed that many people had argued in favor of a lifetime income tax. He noted that such a tax seemed to dilute the information on what happens period by period. Poterba asked if the paper was pushing in the opposite direction by advocating that the government should exploit high frequency information. Werning responded that some of their results were supportive of tax smoothing but that temporary shocks to individuals generally did move the optimal tax system away from a completely smooth tax. He conjectured that it might be possible that a lifetime income tax accompanied with side programs like unemployment insurance to deal with temporary shocks might be close to what the theory suggests is optimal.

Kenneth Rogoff remarked that the discussants had emphasized the importance of knowing how robust the results of the paper were along several dimensions. He noted that another important dimension to generalize the model was the international dimension. Rogoff felt that this was especially important in the context of a world in which both financial and human capital were increasingly mobile.

Daron Acemoglu remarked that the Mirrlees approach to optimal taxation was not so much in the business of writing exact models that could make precise predictions, but rather concerned with understanding general principles. He felt that the real power of the Mirrlees approach was that it was making an explicit effort to understand what the constraints on taxes are. He noted that even though the Ramsey approach often yielded nice insights, the question about why lump sum taxes were ruled out always remained. He noted that in the dynamic setting, lump sum taxes sneak in through the back door in that the optimal tax mimics a lump sum tax. Golosov agreed with Acemoglu's assessment.

Peter Diamond said that while Werning had stressed the role of shocks and Kenneth Judd had talked about insurance markets, his own comments stressed the role of predictable differences between people. He emphasized that there were many predictable differences between people and that in these cases what insurance markets cannot do comes to the fore. He noted that the conclusions of optimal tax theory were likely to change once it was taken into account that the adjustments made by workers in response to shocks are in practice not always smooth in the number of hours worked. Diamond also remarked that it was important to recognize incompleteness when analyzing taxes. In the context of taxes, he thought it was unrealistic to think that policies could be contingent on a full set of types. However, he thought it was important to know what the optimal policy would be if policies could be contingent on a full set of types as a first step to thinking about what to do with fewer powers.