New Dynamic Public Finance:

A User's Guide*

Mikhail Golosov
M.I.T.

Aleh Tsyvinski
Harvard University

Iván Werning
M.I.T.

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

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, 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 [Werning, 2005a]. 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 the three implications of the dynamic Mirrlees literature that differ from those of Ramsey's.

[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 [2005b] study tax-smoothing of labor income taxes when markets are incomplete. Farhi [2005] studies capital income taxation and ownership in this context.

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 [2005b] 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 the 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.

[4] See also Diamond et al. [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.

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 and incentives, making perfect insurance impractical. 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 the 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 when utility is nonseparable. 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. One exception is Kocherlakota [2005b] who extends the inverse Euler equation to the case of aggregate uncertainty and includes a numerical illustration of the optimum with two skill types.

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].[6] As shown by Werning [2005a], this notion does

[6] See also Kingston [1991] and Zhu [1992] for perfect tax smoothing results within a representative agent Ramsey economy with proportional taxation.

not depend on the representative-agent assumption, as it extends to economies with heterogenous agents subject to linear or nonlinear taxation. Thus, in the 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.

To understand this result, it helps to relate a favorable aggregate shock—e.g. a lower government expenditure shock—to a higher aggregate endowment of available goods. Intuitively, in both cases the extra resources will be consumed, thus, they reduce the relative importance of income inequality from the second period skill shocks. As a result, insuring the second period shocks becomes less critical and the solution behaves more like that of an economy without idiosyncratic skill uncertainty in the second period. In the latter case perfect tax smoothing is optimal for each first-period skill type. The smaller spread in wedges can be interpreted as moving towards this perfect tax smoothing ideal.

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].[7,8] These authors laid down the foundation for analyzing optimal non-linear taxation with heterogeneous agents and private information. Many of the more recent result 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,

[7] See also Brito et al. [1991].

[8] See Kocherlakota [2005a] for another review of the literature.

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 [2005b] extended this result to the economy with aggregate shocks. We discuss these results in Section 4. Werning [2002], Shimer and Werning [2005], Abraham and Pavoni [2003] and Kocherlakota [2005b] 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], 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 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 [2005] 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 a 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 a continuum of heterogeneous agents with privately observed skills, where they prove the optimality of Friedman rule, that the optimal inflationary tax is zero.

The analysis of the optimal taxation in response to aggregate shocks has traditionally been studied in macro-oriented Ramsey literature. Werning [2002, 2005a] reevaluated the results on tax smoothing in a model with private information regarding heterogeneous skills. In his setup, all idiosyncratic uncertainty is revealed in the first period. 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 idiosyncratic shocks are also present in the second period.

Some papers, for example Albanesi and Sleet [2006], Kocherlakota [2005b] 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] shows 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 the 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 [2006] 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 the incentive of the government is 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 Acemogly, 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.

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. 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.

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 $\theta_2(i, j)$ where $j = 1, 2, \ldots, N_2(i)$ with probability $\pi_2(j|i)$ is the conditional probability distribution for skill type j in the second period, given skill type i in the first period. We start by assuming that the aggregate shock does not affect the distribution of the population's skills π.

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

We consider two types of shocks in the second period: (i) a shock to the rate of return; and (ii) 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 now functions of s. The probability of state s is denoted by $\mu(s)$.

The resource constraints are

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

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

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:

$$\sum_i \Big(c_1(i) - y_1(i) + \frac{1}{R} \sum_j \left[c_2(i,j) - y_2(i,j) \right] \pi_2(i,j) \Big) \pi_1(i) \leq R_1 K_1 - G_1 - \frac{1}{R} G_2. \tag{3}$$

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 describe 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, as well as decides on the aggregate investments in the first period. 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

$$
\begin{aligned}
&u(c_1(i)) + v\left(\frac{y_1(i)}{\theta_1(i)}\right) + \beta \sum_{s,j} \left[u(c_2(i,j,s)) + v\left(\frac{y_2(i,j,s)}{\theta_2(i,j,s)}\right)\right] \pi_2(i|j)\mu(s) \geq \\
&\quad u(c_1(i_r)) + v\left(\frac{y_1(i_r)}{\theta_1(i)}\right) + \beta \sum_{s,j} \left[u(c_2(i_r, j_r(j,s), s)) + v\left(\frac{y_2(i_r, j_r(j,s), s)}{\theta_2(i,j)}\right)\right] \pi_2(j|i)\mu(s), \quad (4)
\end{aligned}
$$

for all alternative feasible reporting strategies i_r and $j_r(j,s)$.[9] If one assumes that the support of skills does not shift, then it is possible to divide the incentive constraints into simpler first and second period incentive constraints, where only one-shot deviations are considered. For our numerical work, however, it is important to allow the support of the skill distribution to shift.

In our applications we will concentrate on maximizing a utilitarian social welfare function.[10] The *constrained efficient planning problem* maximizes expected discounted utility

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

subject to the resource constraints in (1) and (2) and the incentive constraints in (4). Let (c^*, y^*, k^*)

[9] A powerful *Revelation Principle* guarantees that the best allocations can always be achieved by a mechanism where workers makes report about their types to the planner.

[10] See Diamond [1998] and Tuomala [1990] how choice of the welfare function affects optimal taxes in static framework.

denote the solution to this problem. To understand the implications of these allocation 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

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

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

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

The intertemporal wedge for type i is

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

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:

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

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

$$u(c_1(i)) + v\Big(\frac{y_1(i)}{\theta(i)}\Big) + \beta \Big[u(c_2(i)) + v\Big(\frac{y_2(i)}{\theta_2(i)}\Big) \Big] \geq \\ u(c_1(i_r)) + v\Big(\frac{y_1(i_r)}{\theta(i)}\Big) + \beta \Big[u(c_2(i_r)) + v\Big(\frac{y_2(i_r)}{\theta(i)}\Big) \Big],$$

and subject to the feasibility constraint,

$$\sum_i \left[c_1(i) - y_1(i) + \frac{\beta}{R} \sum_j \big(c_2(i) - y_2(i) \big) \right] \pi_1(i) \leq 0.$$

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 $c_1(i) + c_2(i)$ 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*

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

Note that if $\beta = R$ 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 nor the 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

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

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 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:*

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

There are two cases for which this condition reduces to a standard Euler equation. Both cases involve situations where there is no uncertainty in second period consumption, after conditioning on the first period shock. The first case is when there is no heterogeneity in skills in the second period, i.e., for some i, $N_2(i) = 1$. In the two-type example above, if $\theta_2(1) = \theta_2(2)$, $c_2(1) = c_2(2) = \bar{c}_2$, and the condition becomes

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

which is the standard Euler equation that must hold for a consumer who optimizes savings without distortions. The same is true if $N_2(i) > 1$ but skills evolve deterministically, i.e. $\pi(j|i) = 1$ for some i, j. The second case is when there is no private information. Suppose that, for some i, skills $\theta_2(i, j)$ are observable. Then the planner can ensure that full insurance is achieved $c_2(i, j) = c_2(i, j') = \bar{c}_2$ for all j and j' following such j. In both examples $c_2(1) = c_2(2) = \bar{c}_2$, and the Inverse Euler equation would then reduce to the standard Euler equation (9).

Whenever skills are private information and stochastic, then 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).[11]

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 constrained efficient allocation satisfies:*

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

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. We use the phrase "implicitly save" here to indicate that all savings are controlled by the planner here. A reader of this intuition should think about such "implicit savings" as perturbations of the optimal allocation.

The Inverse Euler Equation can be extended to the case of aggregate uncertainty [Kocherlakota,

[11] 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 $u'(c_2(i, j))$. In our case $Var(c_2(i, j)|i) > 0$ is guaranteed as long as the solution does not hit corners, since bunching all agents in the second period is never optimal.

2005b]. At the optimum

$$\frac{1}{u'(c_1(i))} = \frac{1}{\beta E\left[R(s)\left[\sum_j \pi(j|i)\left[u'(c_2(i,j,s))\right]^{-1}\right]^{-1}\right]}$$

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

$$u'(c_1) = \beta E\left[R(s)u'(c_2(s))\right]$$

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:

$$u'(c_1(i)) < \beta \sum \sum \mu(s)\pi(j|i)R(s)u'(c_2(i,j,s)).$$

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, 2005a], that is, where all the 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,

$$\begin{aligned} u(c_1(i)) + v\left(\frac{y_1(i)}{\theta_1(i)}\right) + \beta \sum_s \left[u(c_2(i,s)) + v\left(\frac{y_2(i,s)}{\theta_2(i)}\right)\right]\mu(s) \geq \\ u(c_1(i_r)) + v\left(\frac{y_1(i_r)}{\theta_1(i)}\right) + \beta \sum_s \left[u(c_2(i_r,s)) + v\left(\frac{y_2(i_r,s)}{\theta_2(i,s)}\right)\right]\mu(s) \end{aligned} \tag{10}$$

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

$$\sum_{i,i_r,s}\Big\{(1+\psi(i,i_r))\Big[u(c_1(i))+v\Big(\frac{y_1(i)}{\theta_1(i)}\Big)+\beta\Big(u(c_2(i,s))+v\Big(\frac{y_2(i,s)}{\theta_2(i,s)}\Big)\Big)\Big]$$
$$-\psi(i,i_r)\Big[u(c_1(i_r))+v\Big(\frac{y_1(i_r)}{\theta_1(i)}\Big)+\beta\Big(u(c_2(i_r,s))+v\Big(\frac{y_2(i_r,s)}{\theta_2(i,s)}\Big)\Big)\Big]\Big\}\mu(s)\pi_1(i)$$

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

$$u'(c_1(i))\gamma^c(i) = \lambda_1\pi(i) \qquad \beta u'(c_2(i,s))\gamma^c(i) = \lambda_2(s)\pi(i)$$
$$-\frac{1}{\theta(i)}v'\Big(\frac{y_1(i)}{\theta(i)}\Big)\gamma^y(i) = \lambda_1\pi(i) \qquad -\frac{1}{\theta(i)}v'\Big(\frac{y_2(i,s)}{\theta(i)}\Big)\gamma^y(i) = \lambda_2(s)\pi(i)$$

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

$$\gamma^c(i) \equiv \pi(i) + \sum_{i'}(\psi(i,i') - \psi(i',i))$$
$$\gamma^y(i) \equiv \pi(i) + \sum_{i'}\Big(\psi(i,i') - \psi(i',i)\frac{\theta(i)}{\theta(i')}\Big)$$

for notational convenience. Combining and cancelling terms then leads to

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

which proves that perfect tax smoothing is optimal in this case. We summarize this result in the next proposition, derived by Werning [2005a] 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 $\tau_1 = \tau_2(s) = \bar{\tau}$.*

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 [2005a], even if idiosyncratic uncertainty is resolved in the first period, but the distribution of skills varies across periods or aggregate states, then the 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. We consider the case in which idiosyncratic uncertainty continues to evolve after the first period, with skill shocks in the second period.

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 relationships 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.[12]

In the context of our economy, taxes in the first period $\tau_1(y_1)$ can depend only on the observable labor supply of agents in that periods, and taxes in the second period $\tau_2(y_1, y_2, k, s)$ can depend on labor supply in both first and second period, as well as agents' wealth. In competitive equilibrium, agent i solves

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

[12] For example, if an agent considers changing her labor, then, in general, she also considers changing her savings. Golosov and Tsyvinski [2006a], Kocherlakota [2005b] and Albanesi and Sleet [2006] showed that such *double deviations* would give an agent a higher utility that 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.

subject to

$$c_1(i) + k(i) \leq y_1(i) - \tau_1(y_1(i))$$

$$c_2(i,j) \leq y_2(i,j) + R(s)k(i) - \tau_2(y_1(i), y_2(i,j,s), k(i), s)$$

We say that a tax system implements the socially optimal allocation $\{(c_1^*(i),\ y_1^*(i),\ c_2^*(i,j),\ y_2^*(i,j,s)\}$ if this allocation solves the agent's problem, given $\tau_1(y_1(i))$ and $\tau_2(y_1(i), y_2(i,j,s), k(i), s)$.

Generally, an optimal allocation may be implementable by various tax systems so $\tau_1(y_1(i))$ and $\tau_2(y_1(i), y_2(i,j,s), k(i), s)$ 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 $\{(c_1^*(i),\ y_1^*(i),\ c_2^*(i,j),\ y_2^*(i,j,s)\}$ 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 actual tax systems employed in the US 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 $\tau_t(y_t, k_t)$ 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 [2005b] 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)

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

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 $R(s)(1 - \tilde{\tau}_k(i,j,s))$ 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 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. A Chamley-Judd [Judd, 1985, Chamley, 1986] result states that capital should not 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 leads to the necessity of the analysis of time consistent policy as a game between a policy maker (government) and a continuum of economic agents (consumers).[13]

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 differently from 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 time inconsistency in Ramsey model. In 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. As well as Roberts [1984], a recent important paper by Bisin and Rampini [2006] considers the problem of mechanism design without commitment in a two-period setting. Bisin and Rampini extend Roberts's analysis and show how the presence of anonymous markets acts as an additional constraint on the government, ameliorating the commitment problem.

[13] A formalization of such 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 et al. [2005] and 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.

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.[14] 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 substantially 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 a 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 desirability and extent of optimal taxation and social insurance policies.

We assumed that individual asset trades and, therefore, agents' consumption, is 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. Intuitively, efficiency results, even in the absence of governmental policy, when firms and agents can write contracts that provide agents with insurance. The competitive nature of the 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 insur-

[14] See also Sleet and Yeltekin [2005] who prove similar result when agents' shocks follow an i.i.d process and the government is benevolent.

ance 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], Arnott and Stiglitz [1990], Greenwald and Stiglitz [1986], Golosov and Tsyvinski [2006b] explore why markets may fail in the presence of asymmetric information.

5 Numerical Exercises

We now perform 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 still possible. In contrast, most work on Mirrleesian models focused on either partial characterization 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 [2006] takes 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.

While we can only conjecture whether the results of our two-period model can be extended to a more general multi-period setup, we are confident that many insights developed here would hold true in a more general model.

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: (i) the allocation for consumption is scaled by the same factor; (ii) the allocation of labor is unaffected; and (iii) marginal taxes rates are unaffected. This result is relevant for thinking about balanced growth in an extension of the model to an indefinite 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.

Below we 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 the first period, skills are distributed uniformly. Individual skills in the first period, $\theta_1(i)$, are equally spaced in the interval $[\underline{\theta}_1, \bar{\theta}_1]$. The probability of realization of each skill are equal to $\pi_1(i) = 1/N_1$ for all i. We choose baseline parameters to be $\underline{\theta}_1 = 0.1$, $\bar{\theta}_1 = 1$ and $N_1 = 50$. Here, a relatively large number of skills allows us to closely approximate a continuous distribution of skills such as in Mirrlees [1971]. 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 $\pi_2(1|i) = \pi_2(2|i) = 1/2$. The agent learns his skill in the second period only at time $t = 2$. We chose the above parametrization of skills to allow a stark characterization of the main forces determining the optimum. The assumption of uniformity of distribution of skills is not innocuous. Saez [2001], a state of the art treatment of static Mirrlees models, 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.

We choose the utility function to be power utility. The utility of consumption is $u(c) = \frac{c^{1-\sigma}}{1-\sigma}$. As our baseline we take $\sigma = 1$, so that $u(c) = \log(c)$. The utility of labor is given by $v(l) = -l^\alpha$; as our benchmark we set $\alpha = 2$. The choice of the baseline utility to be separable is motivated by the fact that most of the theoretical results in the dynamic Mirrlees literature are derived for the case of separable utility functions. Most importantly, the Inverse Euler equation and the optimality of a positive intertemporal wedge are derived only for separable utility functions. In the sections that follow, we provide a numerical characterization of the optimum for the utility function more common in macroeconomic literature on optimal taxation, a nonseparable utility function consistent with a balanced growth path.

In the sections that follow, 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 labelled as follows:
   (a) "Distortion t=1" is a consumption-labor wedge in period 1– τ_{y_1};
   (b) "Distortion high t=2" is a 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 a consumption-labor wedge in period 2 for an agent with a low skill shock – $\tau_{y_2}(i, 2)$;
   (d) "Distortion capital" is an intertemporal (capital) wedge – $\tau_k(i)$;

5.1 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 optimal insurance. The agent's consumption would be equalized across time and 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 as before the deviation but to 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 1 and Figure 2. There is no bunching in either period: agents of different skill 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 the 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.

[**Insert figure 1 and 2 here**]

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 3.

[**Insert figure 3 here**]

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: $\tau_{y_1}(\bar{\theta}_1) = \tau_{y_2}(\bar{\theta}_1, 1) = 0$. This result confirms a familiar "no distortion 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 $\{\tau_{y_1}(\underline{\theta}_1), \tau_{y_2}(\underline{\theta}_1, 1), \tau_{y_2}(\underline{\theta}_1, 1)\}$ 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 from the second period. It follows from the Inverse Euler optimality condition that the intertemporal distortion required is smaller. The intuition is that it is costly to the planner to elicit large effort from those agents and the planner chooses to effectively insure them. To illustrate the intuition, note that Figure 1 shows that consumption uncertainty in the second period increases with the first period shock.

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 ($\alpha_2 = 0.95$) in which agent's skill barely changes from previous period. In Figure 4 the bold line corresponds to the benchmark case of $\alpha_2 = 0.5$; dashed lines correspond to $\alpha_2 = 0.6$, 0.8, 0.9 and 0.95 while dotted lines correspond to $\alpha_2 = 0.3$, 0.1 and 0.05 respectively.

[Insert figure 4 here]

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. This is a surprising result because 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). The labor wedge schedule does become steeper as α_2 increases, i.e., when downward drops are smaller. 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 red line for $\alpha_2 = 0.05$ with the bottom black line for $\alpha_2 = 0.95$. The results are intuitive as an increase in α_2 makes the informational frictions smaller and allows the planner to distort agents' decisions less to provide optimal distortion and redistribution.

To summarize the discussion above, we conclude that the size of the second period shock has significant effects on labor wedges of only the agents who experience that shock and only in that period, while these agents' previous labor decisions and the labor decisions of agents not experiencing the adverse shocks are not affected by the shock. 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 shocks that an agent may experience in the future.

Effects of the probability of second period shocks and uncertainty

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

[Insert figure 5 here]

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 $\pi_2(2|\cdot)$ 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. We can see from the red line that the wedge is the highest when uncertainty about skills is the highest: at the symmetric baseline case with $\pi_2(2|\cdot) = 0.5$. Intuitively, the reason is that the uncertainty about next period's skill is maximized at $\pi_2(2|\cdot) = 0.5$. We conclude that 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.

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 which is a primary objective for the social planner. Specifically, we change the risk aversion parameter σ in the utility function. The results are shown in Figure 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.

[**Insert figure 6 here**]

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.

The effects of higher risk aversion on the intertemporal wedge are the outcome of two opposing forces: (i) 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; (ii) 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.

Effects of changing elasticity of labor supply

We further investigate the properties of the optimum by considering three modification of the disutility of labor. Figure 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).

[**Insert figure 7 here**]

We see that the effect on labor distortions is ambiguous. 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.

The distortion on capital unambiguously goes down. The intuition is that consumption becomes less variable (as argued above) and that the Jensen inequality argument applied to the Inverse Euler equation is less powerful.

5.2 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.

5.2.1 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 widely considered to be the empirically relevant one.

We first consider $\sigma < 1$ cases. Figure 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.

[Insert figure 8 here]

We notice that 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, 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 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.

[Insert figure 9 here]

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.

To show that it is not only the value of the σ that determines the sign of the wedge, we now turn to the case where the skill shocks in the second period have an upward trend so that $\alpha_1 = 1.5$ and $\alpha_2 = 1$, that is an agent may experience a positive skill shock. The results in this case are reversed. 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:

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

This utility function was used by Chari et al. [1994] in their quantitative study of optimal monetary and fiscal policy.

5.3 Summarizing the Case with No Aggregate Uncertainty

The exercises above give us a comprehensive overview of how the optimal allocations and 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 of on capital wedges may be ambiguous as the uncertainty about future skills also matters.

3. Capital wedges are affected by the size of the adverse wedge and by the uncertainty over future skills.

4. Higher elasticity of labor decreases the capital wedge but may have ambiguous effects on labor wedge.

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 on the optimal allocations. 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 main result of this section is to show numerically that if agents' types change over time, the labor wedge smoothing result may no longer hold. This is a novel prediction of this paper.

The literature on new dynamic public finance virtually has not explored implications of aggregate uncertainty on the optimal allocations. A notable exception is Kocherlakota [2005b] who derives a version of the Inverse Euler Equation for the economy with aggregate shocks and explores some quantitative implications of a version of the model with two types of agents.

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 $R_1 = 1$.

6.1 Effects of Government Expenditure Fluctuations

We now turn to analyzing the effects of government expenditures on optimal allocations. 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 $R_2(1) = R_2(2) = 1$.

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 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).

[**Insert figure 10 here**]

We see that higher G leads to significantly 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. The intuition also parallels the case in which there is a shock to the rate of return. Here, an increase in government consumption leads to the planner needing to extract a larger amount of resources from the economy than in the absence of government purchases.

In the Figure 11 we plot the intertemporal wedges for our case of 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. A minor point is that introduction of government expenditures may lead to nonmonotonicity in the capital wedge schedule especially at the lower levels of skills.

[**Insert figure 11 here**]

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 as it is the present value of the government expenditures that matters rather than the distribution of them across time.

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 12 we plot labor wedges. The solid line is τ_{y_1}; the dotted line is $\tau_{y_{1\_1}}(.,1)$ (i.e., high type in state 1); the dashed line is $\tau_{y_{1\_1}}(.,2)$ (i.e., low type in state 1); the dotted line with thick dots is $\tau_{y_{1\_2}}(.,1)$ (i.e., high type in state 2); the dashed line with thick dots is $\tau_{y_{1\_2}}(.,2)$ (i.e., low type in state 2).

[Insert figure 12 here]

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 subsection 4.2, with fixed types, 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 10, where lower government expenditure leads to lower taxes overall. Here, instead, the spread between the distortions on the low and high types become smaller when government expenditures are low. Our intuition is that when government expenditure is low, resources are more abundant. As a consequence output from labor becomes relatively less important. 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 13 that 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 $\mu_1 = 0.3$.

[Insert figure 13 here]

We see that intertemporal wedge becomes higher with higher μ_1, indicating a larger informational distortion.

6.2 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 $R_2(2) = 4$, i.e., there is an upward shock to the return on savings technology. In Figure 14 we plot labor distortions. We plot labor wedges as follows. The solid line is τ_{y_1}; the dotted line is $\tau_{y_{1\_1}}(\cdot, 1)$ (i.e., wedge for the high shock type in state 1); the dashed line is $\tau_{y_{1\_1}}(\cdot, 2)$ (i.e., wedge for the low type in state 1); the dotted line with thick dots is $\tau_{y_{1\_2}}(\cdot, 1)$ (i.e., wedge for the high type in state 2); the dashed line with thick dots is $\tau_{y_{1\_2}}(\cdot, 2)$ (i.e., wedge for the low type in state 2).

[**Insert figure 14 here**]

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, indicating that in that state the informational friction is higher.

We now turn to the analysis of the behavior of the capital wedge under aggregate uncertainty. Figure 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 $R_2 = 1.2$, 2, 3 and 4 (dotted lines).

[**Insert figure 15 here**]

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 $R_2 = 1$, 1.2, 2, 3 and 4 on labor distortions. Qualitatively, they are similar to the ones in the picture above.

6.3 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 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 et al. [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 [2006] who study partial 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 U.S. and other countries, comprised of linear tax on capital and nonlinear labor income tax, 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, non-discriminatory lump-sum taxation, and various exemptions. 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 et al. [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, 2006], 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 *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 [2005] consider how intergenerational incentives should be structured, while Farhi and Werning [2005] and Farhi et al. [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.

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.[15]

[15]We thank Ken Judd for pointing this to us.

References

A. Abraham and N. Pavoni. Efficient allocations with moral hazard and hidden borrowing and lending. 2003.

Daron Acemoglu, Mikhail Golosov, and Aleh Tsyvinski. Markets versus governments: Political economy of mechanisms. mimeo, M.I.T., 2006.

S. Rao Aiyagari, Albert Marcet, Thomas J. Sargent, and Juha Seppälä. Optimal taxation without state-contingent debt. *Journal of Political Economy*, 110(6), 2002.

Stefania Albanesi. Optimal taxation of entrepreneurial capital under private information. mimeo, Columbia University, 2006.

Stefania Albanesi and Christoper Sleet. Dynamic optimal taxation with private information. *Review of Economic Studies*, 2006. forthcoming.

R. Arnott and J. Stiglitz. The welfare economics of moral hazard. In H. Louberge, editor, *Information and Insurance: Essays in Memory of Karl H. Borch*, pages 91–121. Norwell, MA: Kluwer, 1990.

R. Arnott and J.E. Stiglitz. Moral hazard and optimal commodity taxation. *Journal of Public Economics*, 29:1–24, 1986.

A.B. Atkinson and J.E. Stiglitz. The design of tax structure: Direct vs. indirect taxation. *Journal of Public Economics*, 6:55–75, 1976.

Robert Barro. On the determination of the public debt. *Journal of Political Economy*, 87(5): 940–71, 1979.

Marco Battaglini and Stephen Coate. Pareto efficient taxation with stochastic abilities. Cornell and Princeton mimeo., 2005.

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

Alberto Bisin and Adriano Rampini. Markets as beneficial constraints on the government. *Journal of Public Economics*, forthcoming, 2006.

Dagobert L. Brito, Jonathan H. Hamilton, Steven M. Slutsky, and Joseph E. Stiglitz. Dynamic optimal income taxation with government commitment. *Journal of Public Economics*, 44(1): 15–35, 1991.

Christophe Chamley. Optimal taxation of capital in general equilibrium. *Econometrica*, 54:607–22, 1986.

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

V.V. Chari, Lawrence J. Christiano, and Patrick J. Kehoe. Optimal fiscal policy in a business cycle model. *Journal of Political Economy*, 102(4):617–52, 1994.

Juan Carlos Conesa and Dirk Krueger. On the optimal progressivity of the income tax code. *Journal of Monetary Economics*, forthcoming, 2005.

Carlos da Costa. Yet another reason to tax goods. mimeo, Getulio Vargas Foundation, 2005.

Carlos da Costa and Iván Werning. Commodity taxation and social insurance. mimeo, MIT, 2002.

Carlos da Costa and Iván Werning. On the optimality of the friedman rule with heterogeneous agents and nonlinear income taxation. working paper, MIT, 2005.

P.A. Diamond, L.J. Helms, and J.A. Mirrlees. Optimal taxation in a stochastic economy. *Journal of Public Economics*, 14:1–29, 1980.

Peter A. Diamond. Optimal income taxation: An example with a u-shaped pattern of optimal marginal tax rates. *American Economic Review*, 88(1):83–95, 1998.

Peter A. Diamond and James A. Mirrlees. A model of social insurance with variable retirement. *Journal of Public Economics*, 10(3), 1978.

Peter A. Diamond and James A. Mirrlees. Payroll-tax financed social insurance with variable retirement. *Scandinavian Journal of Economics*, 88(1), 1986.

Peter A. Diamond and James A. Mirrlees. Social insurance with variable retirement and private saving. mimeo, MIT, 1995.

Emmanuel Farhi. Capital taxation and ownership when markets are incomplete. Mimeo, MIT, 2005.

Emmanuel Farhi and Iván Werning. Inequality, social discounting and progressive estate taxation. NBER Working Papers 11408, 2005.

Emmanuel Farhi and Ivan Werning. Capital taxation: Quantitative implications of the inverse euler equation. mimeo, 2006.

Emmanuel Farhi, Narayana Kocherlakota, and Iván Werning. Estate taxes and estate subsidies. work in progress, 2005.

Jesus Fernandez-Villaverde and Aleh Tsyvinski. Optimal fiscal policy in a business cycle model without commitment. mimeo, 2004.

M. Golosov and A. Tsyvinski. Optimal taxation with endogenous insurance markets. *Quarterly Journal of Economics*, 2006b. forthcoming.

M. Golosov and A. Tsyvinski. Designing optimal disability insurance: A case for asset testing. *Journal of Political Economy*, 2006a.

Mikhail Golosov, Narayana Kocherlakota, and Aleh Tsyvinski. Optimal indirect and capital taxation. *Review of Economic Studies*, 70(3):569–587, 2003.

B.C. Greenwald and J.E. Stiglitz. Externalities in economies with imperfect information and incomplete markets. *Quarterly Journal of Economics*, 101:229–264, 1986.

Kenneth L. Judd. Redistributive taxation in a perfect foresight model. *Journal of Public Economics*, 28:59–83, 1985.

Kenneth L. Judd. Optimal taxation in dynamic stochastic economies: Theory and evidence. mimeo, 1989.

Kenneth L. Judd. Optimal taxation and spending in general competitive growth models. *Journal of Public Economics*, 71(1):1–26, 1999.

Marek Kapicka. Optimal taxation with human capital accumulation. University of California-Santa Barbara, 2005.

Geoffrey Kingston. Should marginal tax rates be equalized through time? *Quarterly Journal of Economics*, 106(3):911–24, 1991.

Paul Klein, Per Krusell, and Jose-Victor Rios-Rull. Time-consitent public expenditures. Mimeo, University of Pennsylvania, 2005.

Narayana Kocherlakota. Advances in dynamic optimal taxation. mimeo, 2005a.

Narayana R. Kocherlakota. Zero expected wealth taxes: A mirrlees approach to dynamic optimal taxation. *Econometrica*, 73(5):1587–1621, 2005b.

Robert E., Jr. Lucas and Nancy L. Stokey. Optimal fiscal and monetary policy in an economy without capital. *Journal of Monetary Economics*, 12:55–93, 1983.

James A. Mirrlees. An exploration in the theory of optimum income taxation. *Review of Economic Studies*, 38(2):175–208, 1971.

Christopher Phelan. Opportunity and social mobility. *Federal Reserve Bank of Minneapolis, Research Department Staff Report*, 323, 2005.

Christopher Phelan and Ennio Stacchetti. Sequential equilibria in a ramsey tax model. *Econometrica*, 69(6):1491–1518, 2001.

Edward C. Prescott and Robert Townsend. Pareto optima and competitive equilibria with adverse selection and moral hazard. *Econometrica*, 52:21–45, 1984.

Kevin Roberts. Theoretical limits to redistribution. *Review of Economic Studies*, 51:177–195, 1984.

William P. Rogerson. Repeated moral hazard. *Econometrica*, 53(1):69–76, 1985.

Emmanuel Saez. Using elasticities to derive optimal income tax rates. *Review of Economic Studies*, 68(1):205–29, 2001.

Robert Shimer and Iván Werning. Liquidity and insurance for the unemployed. mimeo, MIT, 2005.

Christopher Sleet and Sevin Yeltekin. Credible social insurance. mimeo, 2005.

S Smyth. A balancing act: Optimal nonlinear taxation of labor and capital in overlapping generation economies. mimeo, 2005.

Joseph Stiglitz. Pareto efficient and optimal taxation and the new new welfare economics. In A. Auerbach and M. Feldstein, editors, *Handbook on Public Economics*, pages 991–1042. North Holland: Elsevier Science Publishers, 1987.

Kjetil Storesletten, Chris I. Telmer, and Amir Yaron. Cyclical dynamics in idiosyncratic labor market risk. *Journal of Political Economy*, 112(3):695–717, 2004.

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

Iván Werning. Optimal dynamic taxation. PhD Dissertation, University of Chicago, 2002.

Iván Werning. Tax smoothing with redistribution. Federal Reserve Bank of Minneapolis, Staff Report 365, 2005a.

Iván Werning. Tax smoothing with incomplete markets. Mimeo, mit, 2005b.

Xiaodong Zhu. Optimal fiscal policy in a stochastic growth model. *Journal of Economic Theory*, 58:250–89, 1992.

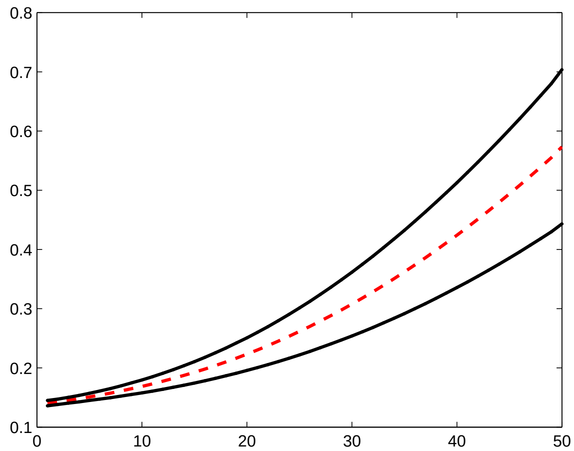

Figure 1: Consumption allocation. Middle dotted line shows first period consumption; outer solid lines are second period consumption.

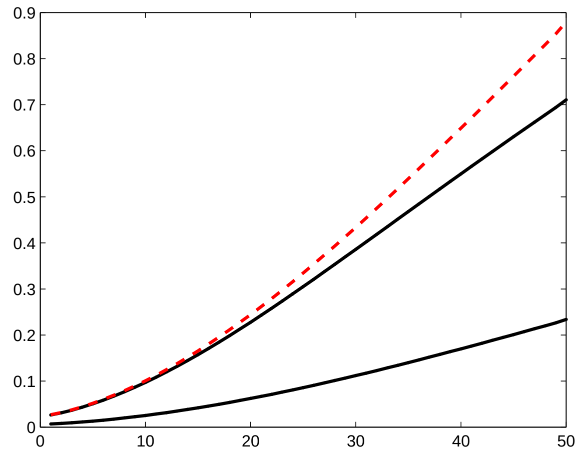

Figure 2: Effective labor allocation. Dashed line is for first period. Solid lines are for second period, top is high shock, bottom low shock.

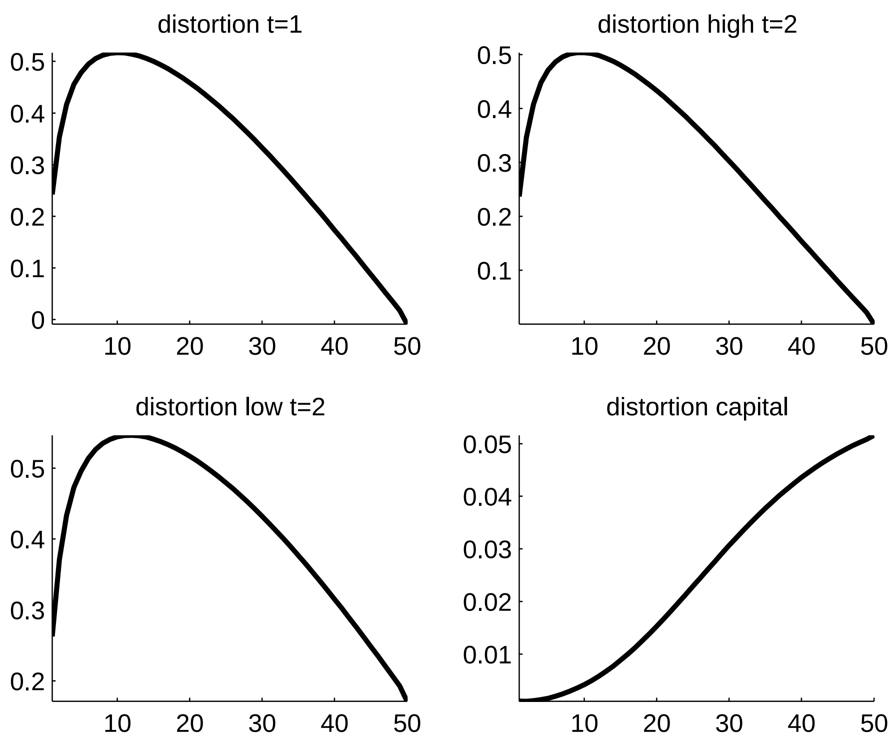

Figure 3: Benchmark implicit marginal tax rates.

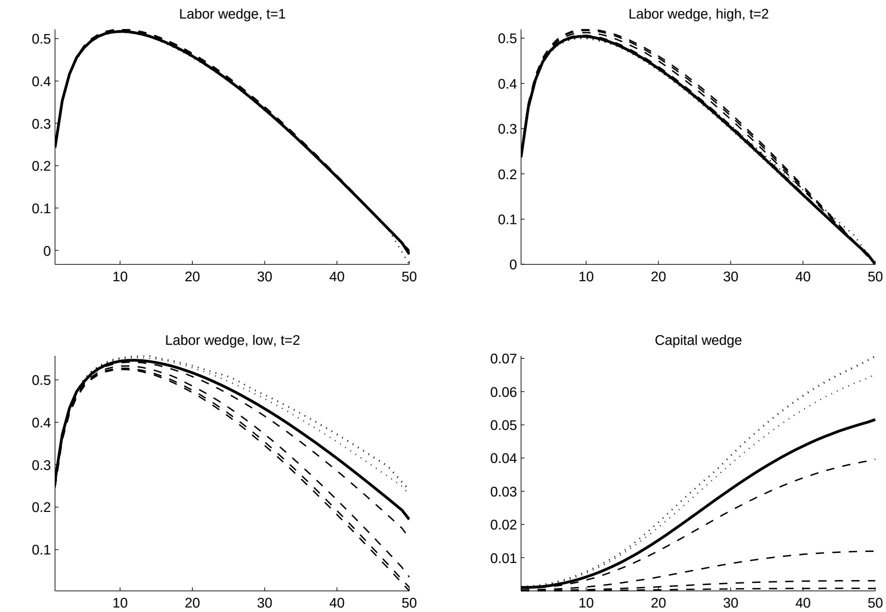

Figure 4: Varying α_2.

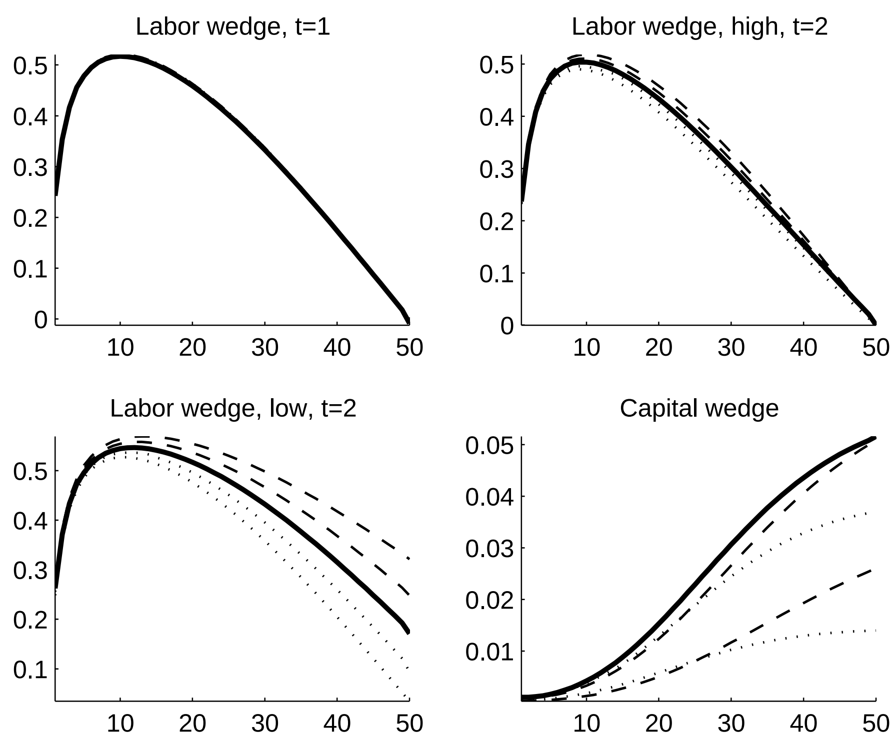

Figure 5: Varying the probability of skill drop $\pi_2(2|\cdot)$.

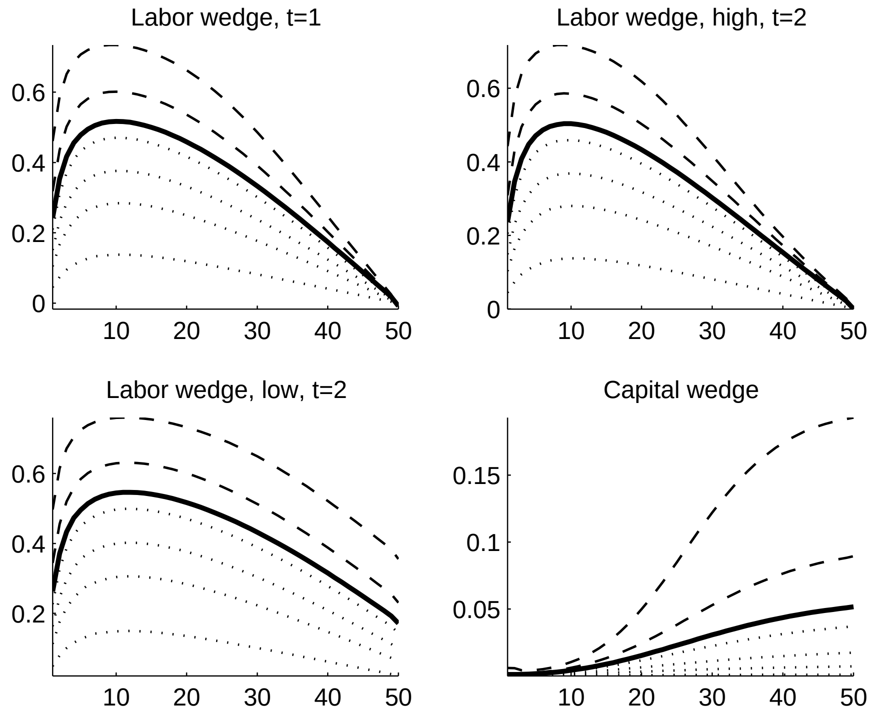

Figure 6: Varying Risk Aversion

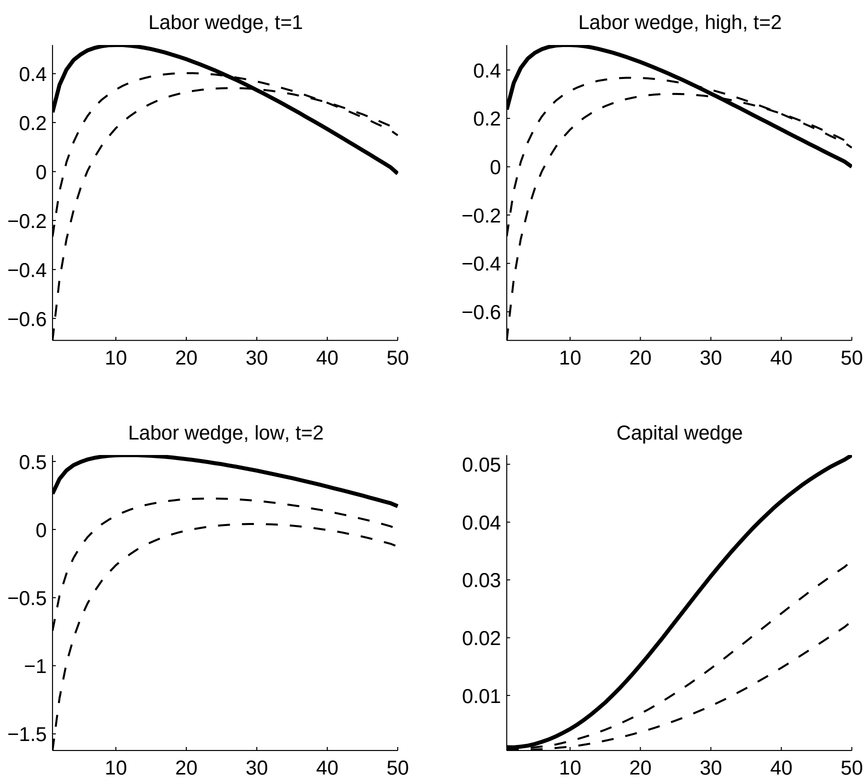

Figure 7: Changing elasticity of labor.

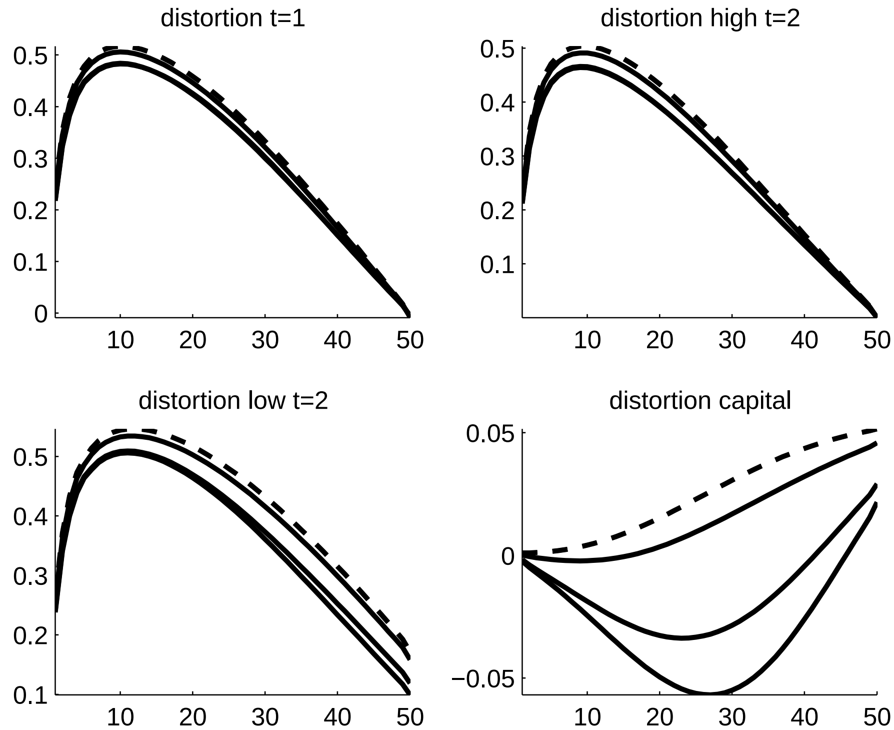

Figure 8: Nonseparable utility with $\sigma \leq 1$.

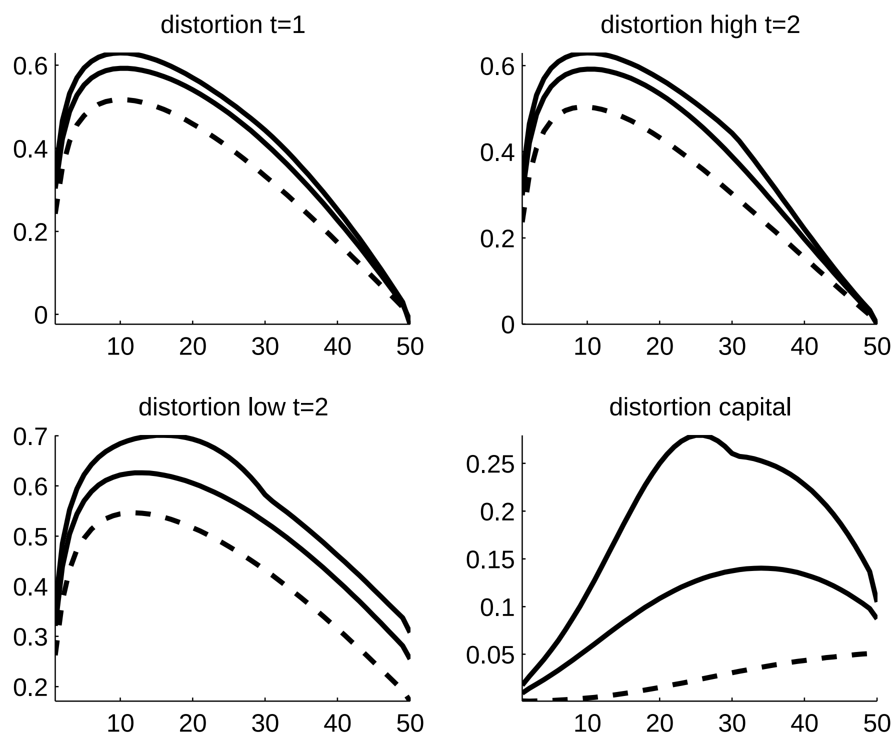

Figure 9: Nonseparable utility with $\sigma \geq 1$.

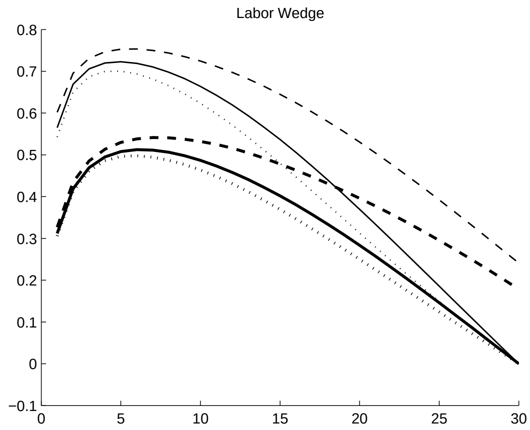

Figure 10: Labor Distortion

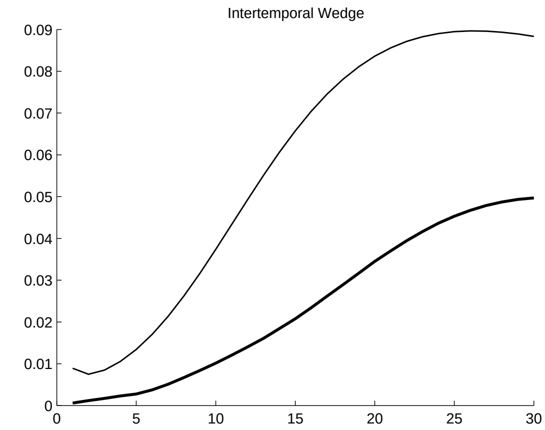

Figure 11: Intertemporal Distortion

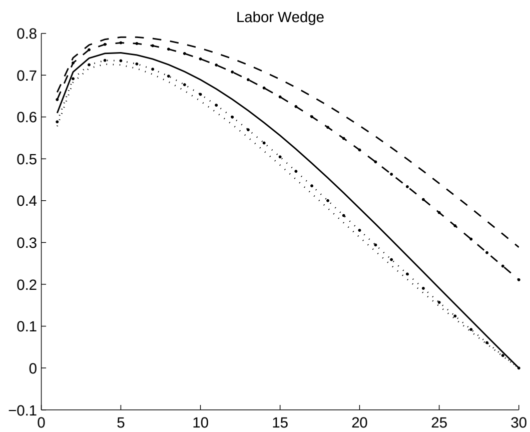

Figure 12: Shocks to government expenditure

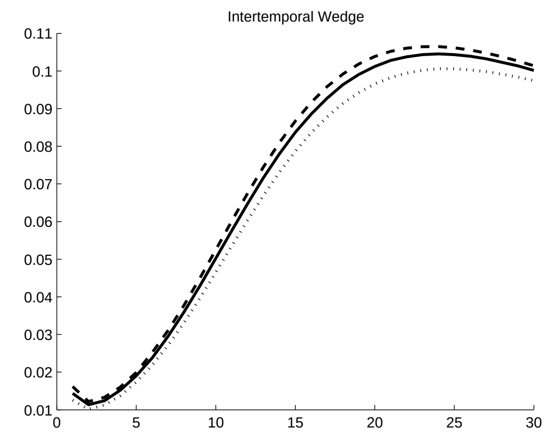

Figure 13: Intertemporal distortion

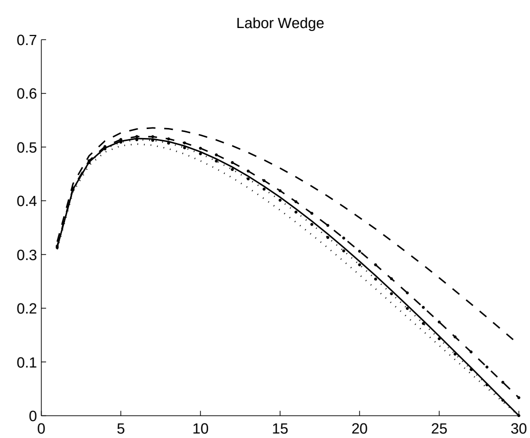

Figure 14: Rate of Return Shocks

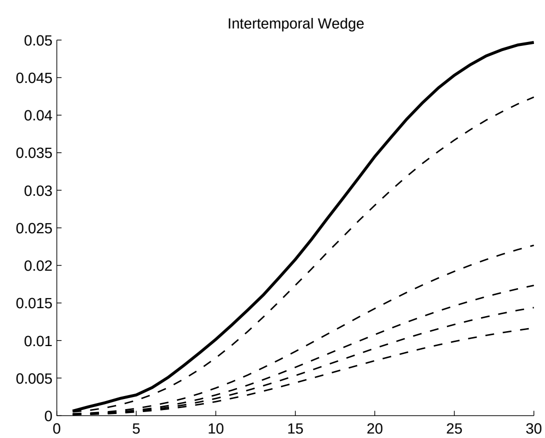

Figure 15: Intertemporal distortion varying R_2.