New Dynamic Public Finance:
A User's Guide*

Mikhail Golosov M.I.T.
Aleh Tsyvinski Harvard University
Iván Werning M.I.T.

This draft: March, 2006

Abstract

This paper presents a simple dynamic Mirrleesian model. There are two main goals for this paper: (i) to review some recent results and contrast the Mirrlees approach with the Ramsey framework in a dynamic setting; and (ii) to present new numerical results for a flexible two-period economy featuring aggregate shocks.

*For comments on an early draft of this paper we thank Daron Acemoglu, V.V. Chari, Kenneth Judd and Michael Woodford.

1 Introduction

New Dynamic Public Finance is a recent literature that extends the static Mirrlees (1971) framework to dynamic settings.[1] This approach addresses a much broader set of issues related to dynamic policy than its static counterpart and does not rely on exogenously specified tax instruments as in Ramsey approach.

We show that *New Dynamic Public Finance* delivers three results that contrast with predictions from a Ramsey approach. First, it is optimal to introduce a positive distortion in savings. This wedge improves the provision of incentives over time by implicitly discouraging savings (Diamond and Mirrlees, 1978; Rogerson, 1985; Golosov, Kocherlakota, and Tsyvinski, 2003). This contrasts with the important Chamley-Judd (Judd, 1985; Chamley, 1986) result obtained in Ramsey models that capital should go untaxed, at least in the long run.[2] Second, when agents skill evolve stochastically, their marginal labor income tax rates are affected by aggregate shocks. Thus, a perfect version of labor tax smoothing prevalent in Ramsey models (Barro, 1979; Lucas and Stokey, 1983; Judd, 1989; Zhu, 1992; Chari, Christiano, and Kehoe, 1994) does not hold in dynamic Mirrlees models with uncertain and evolving skills.[3] However, it is optimal to smooth labor distortions when skills are heterogenous but constant (Werning, 2005b). Finally, the nature of time consistency problems in dynamic Mirrleesian models is very different from those arising within Ramsey setups. A benevolent government without full commitment cannot refrain from exploiting the information that it has collected at previous dates. The time consistency problem is, essentially, about learning agents' private information, rather than taxing sunk capital. Even if the government were to control all capital accumulation in the economy, a time consistency problem would arise in a dynamic Mirrlees model.

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 Ramsey. Our workhorse model is a two-period setup with and without aggregate uncertainty, regarding government purchases or rates of returns on savings. The

[1] However, see Diamond and Mirrlees (1978, 1986, 1995) for important 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, Marcet, Sargent, and Seppälä (2002) and Werning (2005a) study tax-smoothing of labor income taxes when markets are incomplete. Farhi (2005) studies capital income taxation and ownership in this context.

model is flexible enough to illustrate some key results. Moreover, its tractability allows us to explore some new issues. The aim of the paper is to comprehensively explore the structure of distortions and its dependence on parameters within our dynamic Mirrlees economy. Papers by Albanesi and Sleet (2006), Golosov and Tsyvinski (2006) and Kocherlakota (2005b) have some of the insights that we derive here but our model allows a broader overview of the issue.[4] Although some of our work is numerical, the 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.

Theoretically, we know that presence of private information regarding skills and the stochastic evolution of skills introduces distortions in the marginal decisions of agents. We focus on two types of wedges. The first wedge is a consumption-labor wedge that we would often call a labor wedge. This wedge is a ratio of the marginal utility of consumption of an agent to a marginal utility of labor. The second wedge arises because of the stochastic nature of the problem. We call that wedge a capital or intertemporal wedge. This is a wedge between marginal rate of substitution between periods (ratio of marginal utilities) and the return on savings. Our focus in this paper is distinctively on wedges rather than on taxes implementing them, although we do devote a section to discussing the latter.

1.2 Ramsey and Mirrlees approaches

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

This approach studies the problem of choosing taxes within a given set of available tax instruments. Usually, to avoid the first-best, it is assumed that taxation must be proportional. Lump-sum taxation, in particular, is prohibited. The benevolent government sets taxes so as to finance its expenditures and maximize the representative agent's utility.

If lump-sum taxes were allowed then the first welfare theorem would apply, and the

[4] An interesting paper by Diamond, Helms, and Mirrlees (1980) is 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.

unconstrained 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 and why they take such a particular form. As such, it is often recognized that Ramsey representative agent models do not deliver a theoretical foundation for distortionary taxation. Distortions are assumed, and their overall level are largely determined exogenously by the level of government expenditure.

The Mirrlees approach to optimal taxation is built on a different foundation. Rather than starting with a restricted set of tax instruments as in Ramsey, Mirrlees (1971) assumed that an informational friction endogenously restricted the set of taxes that implement the optimal allocation. In these models workers are heterogenous with respect to their skills or productivity. Importantly, worker skills and work effort are not observed by the government. The private information creates a tradeoff between insurance and incentives, making perfect insurance impractical. Even when tax instruments are not unduly constrained, distortions generally arise at 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 skills then a simple lump-sum tax—that is, an income tax with no slope—would 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 marginal tax rates is the desire to redistribute or insure skill draws. Thus, taxes are affected by the distribution of skills and risk aversion, among other things.

1.3 Numerical results

We begin by summarizing the finding from our numerical simulations for 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, labor wedges are higher for agents who receive an adverse skill shock. We find that the labor wedge in the first period or the labor wedge in the second for those not suffering the adverse shock period are largely unaffected by the size or probability of the adverse shock; these parameters affect such agents only indirectly through the ex-ante incentive compatibility constraints.

We also find that higher risk aversion leads to higher labor wedges because it leads to higher desire for the planner to redistribute or insure agents. As for the elasticity of labor supply we show two opposing effects that affect 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.

We find that the two key determinants of the size of the capital wedges are the size of the adverse future shock and its probability. We found that a higher elasticity of labor decreases the savings wedge because it decreased the desire to redistribute.

We derive some novel predictions for capital wedges when preferences over consumption and labor are nonseparable. Most of the theoretical results in dynamic Mirrleesian taxation are derived assuming additive separability between consumption and labor. In particular, the analytical Inverse Euler results, which ensures a positive capital wedge is proven only for the case of separable utility. Indeed, the effects of nonseparable utility on the intertemporal wedge are largely unexplored. Here we feel the gap by showing the effects of nonseparability of utility. Most importantly, we find that when utility is nonseparable, the capital wedge may become negative. The sign of the wedge depends on whether labor and consumption are complements or substitutes and on whether an agent expects to experience an upward or downward shock to skills in the future.

We now discuss our numerical findings for the case with aggregate uncertainty. Most of these findings are novel, since aggregate shocks have not been extensively explored with the Mirrleesian approach. One exception is Kocherlakota (2005b) who extends the inverse Euler equation to the case of aggregate uncertainty and also considers a numerical illustration of his tax implementation in a model with two skill types.

An important result in the representative agent Ramsey framework, due to Barro (1979) and Lucas and Stokey (1983), is that tax rates on labor income should be smoothed across time and states.[6] As shown by Werning (2005b), this important benchmark does not rely on the Ramsey framework, and extends to situations with heterogenous agents subject to linear taxation or nonlinear taxation. In the our setup this result obtains as long as all the idiosyncratic uncertainty for skills is resolved in the first period. In our numerical exercises we consider the case with aggregate uncertainty while allowing individual skills to evolve stochastically.

[6] See also Zhu (1992) for a perfect tax smoothing result within a representative agent Ramsey economy with proportional taxation.

There are two main implications of aggregate uncertainty. First, we find that labor wedges vary across aggregate shocks. Thus, perfect tax smoothing, where the wedges for each skill type are invariant to aggregate states, does not hold. Tax rates vary because individual skill shocks and aggregate shocks are linked through the incentive constraints. Second, 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. Formally, a lower government expenditure is equivalent to a higher endowment in the hands of the government. Intuitively, these extra resources, in government hands, reduce the relative importance of income inequality from the second period skill shocks. As a result, insuring the second period shocks becomes less valuable, so the optimal allocation behaves more like that of an economy where there is no second period skill uncertainty, where perfect tax smoothing obtains, so the spread in the wedge across skill types is reduced.

2 An Overview of the Literature

The dynamic Mirrleesian literature builds on the seminal work of Mirrlees (1971), Diamond and Mirrlees (1978), Atkinson and Stiglitz (1976) and Stiglitz (1987).[7] These authors laid down a foundation for the analysis of 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 aggregate shocks and the stochastic evolution of individual skills.

Werning (2002) and Golosov, Kocherlakota, and Tsyvinski (2003) incorporated Mirrleesian framework into the a standard neoclassical growth model. Werning (2002) on 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, Kocherlakota, and Tsyvinski (2003) showed that it is optimal to distort savings in 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. Golosov and Tsyvinski (2005) and da Costa (2005) further extended the analysis by considering economies where not only skills but also financial transactions such as borrowing and lending are not observable. In such settings they show that non-linear distortions of savings is not feasible. The government may uniformly influence the rate of return by taxing observable capital stock. Al-

[7] See also Brito, Hamilton, Slutsky, and Stiglitz (1991).

though the optimal savings distortion still remains positive, Golosov and Tsyvinski (2005) show examples when it is optimal to subsidize capital. 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. Kocherlakota (2005a) provides a comprehensive review of the literature.

Unlike taxation of savings, less work has been done for 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 interesting 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 in wealth. The intuition for their result is that the need to redistribute resources among the future generations implies that relatively rich, high skilled parents should face higher distortions on bequests than poor, low skilled ones. This equalizes opportunities for the generation which is born in the next period. Albanesi (2006) considers optimal taxation of the entrepreneurs. In her set up an entrepreneur exhorts 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 to a monetary model with a continuum of heterogeneous agents with privately observed skills, where they prove the optimality of Friedman rule, so that the optimal inflationary tax is zero.

The analysis of the optimal taxation in response to aggregate shocks has traditionally been primarily studied in macro-oriented Ramsey literature. Werning (2002, 2005b) reevaluated the results on tax smoothing in a model with private information where agents skills are constant over a worker's lifetime. We further explore the extent of tax smoothing in response to the aggregate shocks in a general model in which agents' skills are stochastic in Section 6.

Some papers, for example Albanesi and Sleet (2006), Kocherlakota (2005b) and Golosov and Tsyvinski (2006), 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 (2005) 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.

Following theoretical advances, several authors carry out a quantitative analysis 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 (2006) 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 does not hold in such an environment. One of the early contributors was Roberts (1984) who studies an economy where individual have constant skills which are private information. Bisin and Rampini (2006) study a two period version of this problem. Sleet and Yeltekin (2005) and Acemoglu, Golosov, and Tsyvinski (2006) show conditions under which even the simplest versions of the Revelation Principle can be applied along the equilibrium path. We discuss these issues in Section 4.

3 A Two-Period Mirrleesian Economy

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

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 should represent 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. However, with no labor decision in the third period, nothing is lost by lumping consumption into the second period, as we 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, so that by an informal appeal to the law of large numbers, there is no aggregate uncertainty.

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 populations relative 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

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

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

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 come 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 incentives to misreport his type. Thus the allocations must satisfy the

following incentive constraint

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

for all alternative feasible reporting strategies i_r and $j_r(j,s)$.[8] If one assumes that the support of skills does not shift then it is possible to parcel out 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.[9] 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^*) 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

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

[8] 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.

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

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 derive implications of the intertemporal distortion, or implicit tax on capital. We first work with an important benchmark with no skill shocks in the second period. For this case we recover Atkinson and Stiglitz's (1976) classical uniform taxation result, which implies no capital taxation in our context. 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 skills of agents are fixed over time, and there is no aggregate uncertainty. Specifically, assume that $N_2(i) = 1$ for $\forall i$ and that $\theta_1(i) = \theta_2(i,j) = \theta(i)$. In this case the constrained efficient problem simplifies to:

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

subject to the incentive compatibility constraint that for $\forall i' \in \{1, ..., N_1\}$, 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,

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

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 important 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, $N_2(i) = 1$. Alternatively, consider the case in which skills evolve deterministically, i.e. $\pi(j|i) = 1$ for some i, j. 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

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

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 the context of the two type example above, for the two cases we outlined, $c_2(1) = c_2(2) = \bar{c}_2$, and the Inverse Euler equation would then reduce to the standard Euler equation (9).

We can now derive an important proposition that shows that, if skills are stochastic, i.e., if the probability of a change in skill is interior, than the Euler equation is

distorted.

Proposition 3 *Suppose that for some i, there exists j such that $0 < \pi(j|i) < 1$. 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, 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

An important result in 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 important benchmark does not rely on the Ramsey framework, and extends to situations with heterogenous agents subject to linear taxation or nonlinear taxation (Werning, 2005b). In our setup this result obtains as long as all the idiosyncratic

uncertainty for skills is resolved in the first period, so that $\theta_2(j,i) = \theta_1(i) = \theta(i)$. We can then write the allocation entirely of the first period skill shock and the second period aggregate shock. The incentive constraints then only require truthful revelation of the first period skill,

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

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

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

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 canceling terms then leads to

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

which proves that perfect tax smoothing is optimal in this case. We summarize this

result in the next proposition, due to Werning (2005b).

Proposition 4 *When idiosyncratic uncertainty for skills is resolved 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 (2005b), even without additional uncertainty in the second period, if the skill distribution varies marginal taxes should also vary. 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 the optimal allocations with tax instruments. The general goal is to find taxes such that the resulting competitive equilibrium yields the socially optimal 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. The relationships between wedges and taxes may be 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.[10]

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

[10]For example, if an agent considers changing her labor, then, in general, she also considers changing her savings. Golosov and Tsyvinski (2006), 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.

$\tau_2(y_1, y_2, k, s)$ can depend on labor supply in both first and second period, as well as agents' wealth. In competitive equilibrium, agent i solves

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

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 this 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 particular 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 agent's 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. The rate at which the capital stock is taxed depends on the whole history of labor supplies by an agent, and thus is highly history dependent. The tax rate is defined by

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

It turns out that an interesting implication of this implementation is that, at the optimum, taxes on capital average out to zero and raise no revenue. Although the average net return on savings is unaffected, it does induce savings distortions by making this return risky: capital taxes are higher in those states where consumption is lower (see equation (11)). Since net returns are positively correlated with consumption—thus, negatively correlated with marginal utility—this makes saving less attractive. However, it is important to emphasize that the intertemporal wedge is positive for each agent.

In some applications the number of shocks that agents face is small and allow for simple decentralizations. Golosov and Tsyvinski (2006) 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 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).[11]

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.

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 sustain-

[11] 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, Krusell, and Rios-Rull (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.

ability 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, the *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.[12] The fact that 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 environment Golosov and Tsyvinski (2005) follow Prescott and Townsend (1984) and Atkeson and Lucas (1992) to show that allocations provided by competitive markets are constrained efficient and the first welfare theorem holds. Intuitively, the argument for constrained efficiency of the competitive equilibrium is as follows. Consider an economy populated by competitive firms. In the absence of governmental policy, firms and agents can write contracts that provide agents with insurance. The competitive nature of the insurance markets would lead to agents receiving the constrained efficient allocation. They conclude that, even in the presence of private information, markets can provide optimal insurance if consumption is observable. This result, however, does not mean that unconstrained efficient allocation can be achieved. Individual insurance contracts between agents and firms would feature exactly the same wedges as the social planner's problem. This implementation again highlights that it is important to focus on wedges rather than the taxes implementing them. In this paper we do not model explicitly the

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

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 (2005) show why markets might 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 (2006). A recent paper by Farhi and Werning (2006) takes a different approach by studying partial tax reforms—that capture the savings distortions implied by the Inverse Euler equation—that are possible in a general model.

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 periods 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. Specifically, the number of shocks $N_2(i) = 2$ 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 of a higher skill would always prefer to claim to be of a lower type to receive the same consumption as before the deviation but work less. The optimal allocation with unobservable types balances two objectives of the social planner: providing insurance and respecting incentive compatibility constraints.

The optimal allocation for the benchmark case with unobservable types is shown in Figure 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.

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.

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

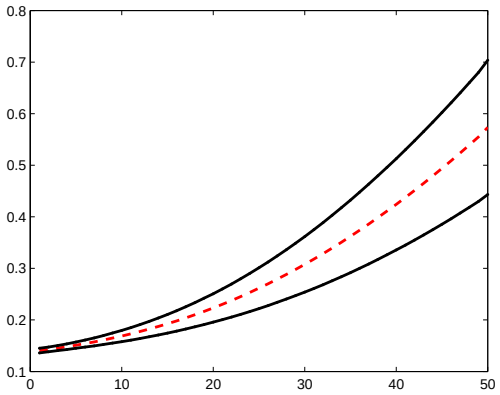


Figure 1: Consumption allocation. Middle dotted line shows first period consumption; outer solid lines are second period consumption.

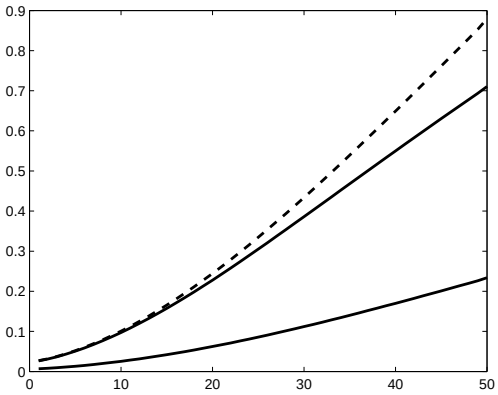


Figure 2: Effective labor allocation. Dashed line is for first period. Solid lines are for second period, top is high shock, bottom low shock.

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

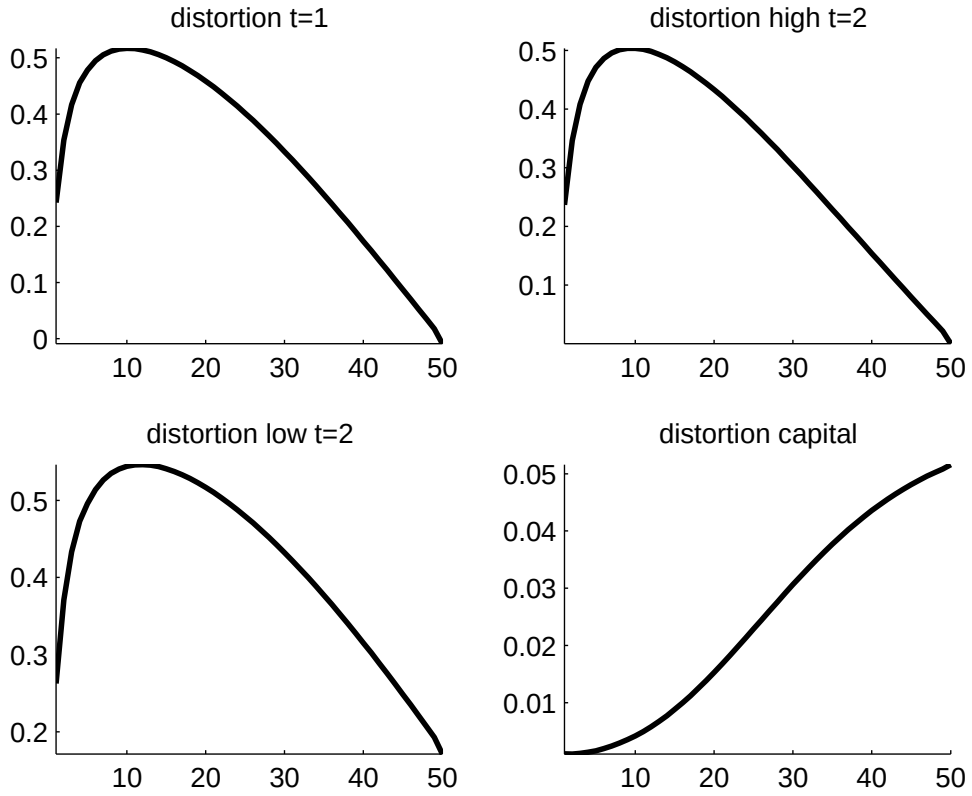


Figure 3: Benchmark implicit marginal tax rates.

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 we plot in red the results for $\alpha_2 = 0.05$. The other lines show $\alpha_2 = 0.1$, 0.3, 0.5, 0.6, 0.8, 0.9 and 0.95 respectively.

We now describe 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

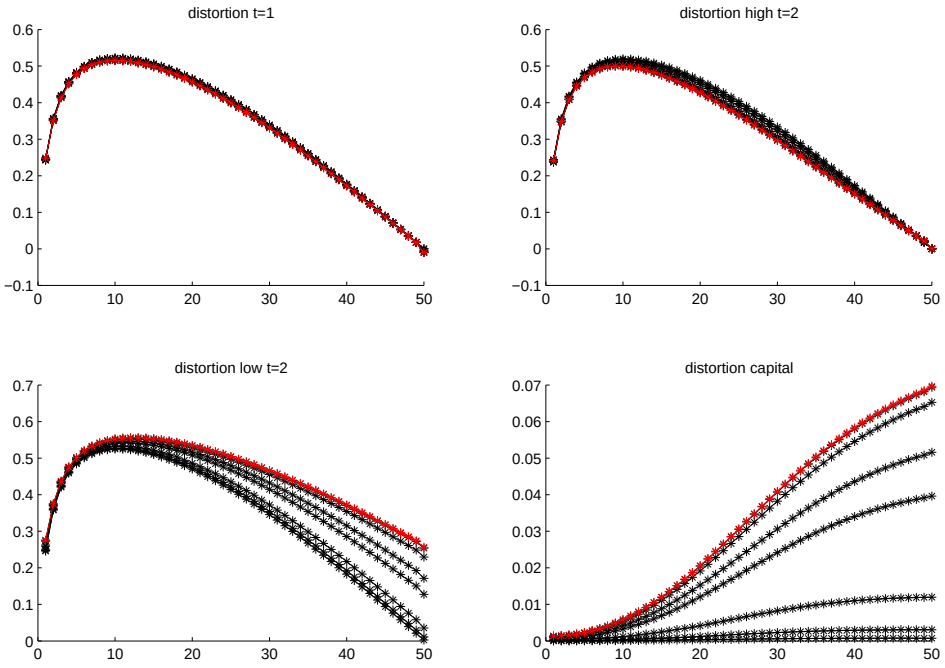


Figure 4: Varying α_2.

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 with $\alpha_2 = 0.05$ with the bottom black line with $\alpha_2 = 0.95$. The results are intuitive as an increase in α_2 makes the informational frictions smaller and allows us 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 the 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 interesting as 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 red the benchmark case where $\pi_2(2|\cdot) = 0.5$. In blue, $\pi_2(2|\cdot) = 0.1$; in black, $\pi_2(2|\cdot) = 0.3$; in yellow, $\pi_2(2|\cdot) = 0.7$.; in green, $\pi_2(2|\cdot) = 0.9$.

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 motive for the social planner. Specifically, we change the risk aversion parameter σ in the utility function. The results are shown

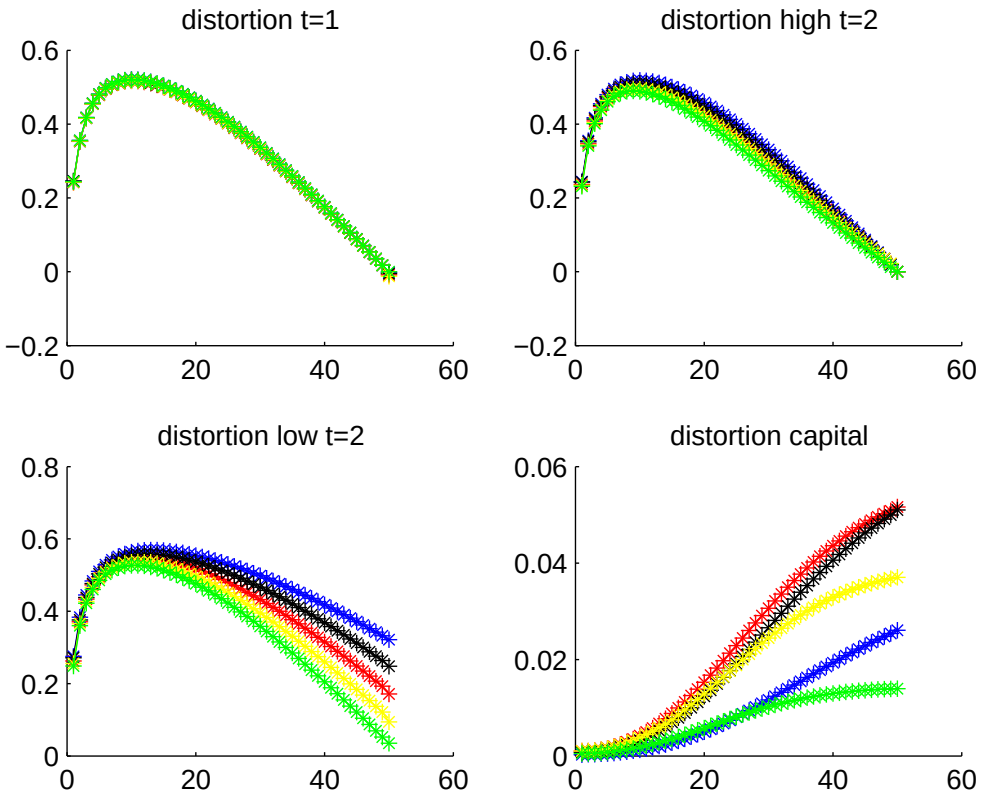


Figure 5: Varying the probability of skill drop $\pi_2(2|\cdot)$.

in Figure 6. Our benchmark case of logarithmic utility $\sigma = 1$ is shown in red. In black we plot lower risk aversions: $\sigma = 0.8$, 0.5, 0.3 and 0.1. In blue we plot higher risk aversions: $\sigma = 1.5$ and 3. The immediate observation is that a higher degree of risk aversion leads to uniformly higher distortions. The intuition is again rather simple. We know that if $\sigma = 0$, so that utility is linear in consumption and an agent is risk neutral, private information about the skill would not affect the optimal allocation and the unconstrained allocation in which all wedges are equal to zero can be obtained. The higher is risk aversion, the higher is the desire of the social planner to redistribute and insure agents. Therefore, all distortions rise.

The effects of higher risk aversion on the intertemporal wedge are very interesting. Intuitively, there are two forces: (1) for a given distribution of consumption in both periods, a higher risk aversion σ increases the wedge directly (recall that the capital wedge was derived by applying Jensen's inequality to the inverse Euler equation. Here, loosely speaking, Jensen's inequality is more powerful for higher is σ); (2) on the other hand, with higher risk aversion it is optimal to insure more, which would

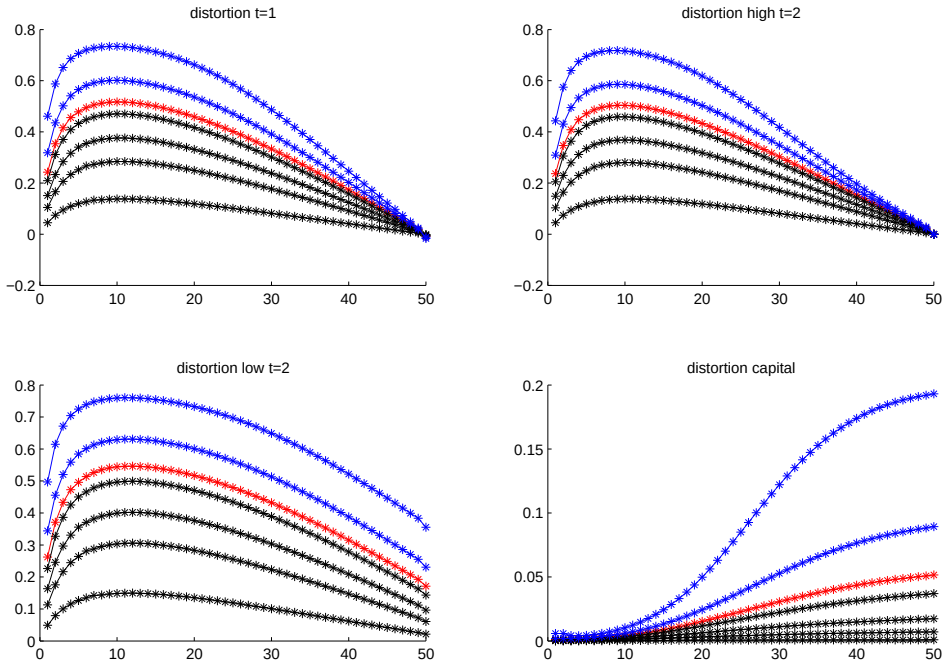


Figure 6: Varying Risk Aversion

reduce the wedge — this is an indirect effect. For the parametrization we considered the direct effect is stronger.

We conclude that higher risk aversion through increasing the desire to redistribute among agents has significant effects on the size of the wedges.

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 red in the figure). We also display two more inelastic cases: $v(l) = -l^3$ (plotted in blue), and $v(l) = -l^4$ (plotted in black).

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.

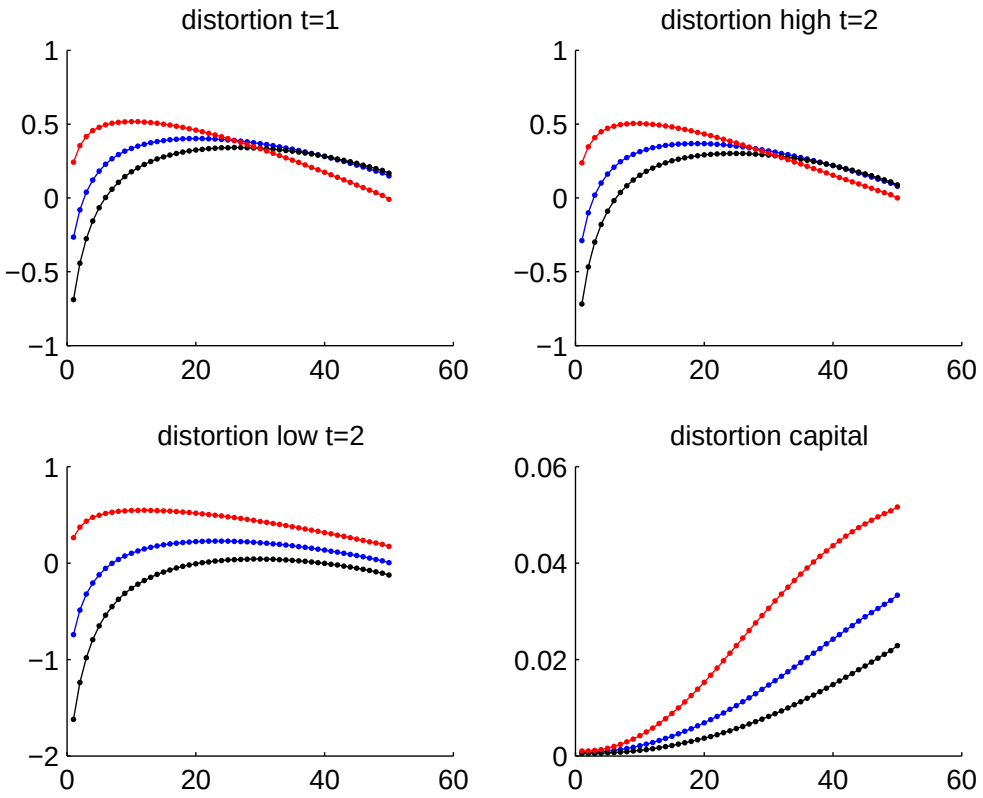


Figure 7: Changing elasticity of labor.

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 ensure a positive intertemporal wedge do not 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. The figure 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.

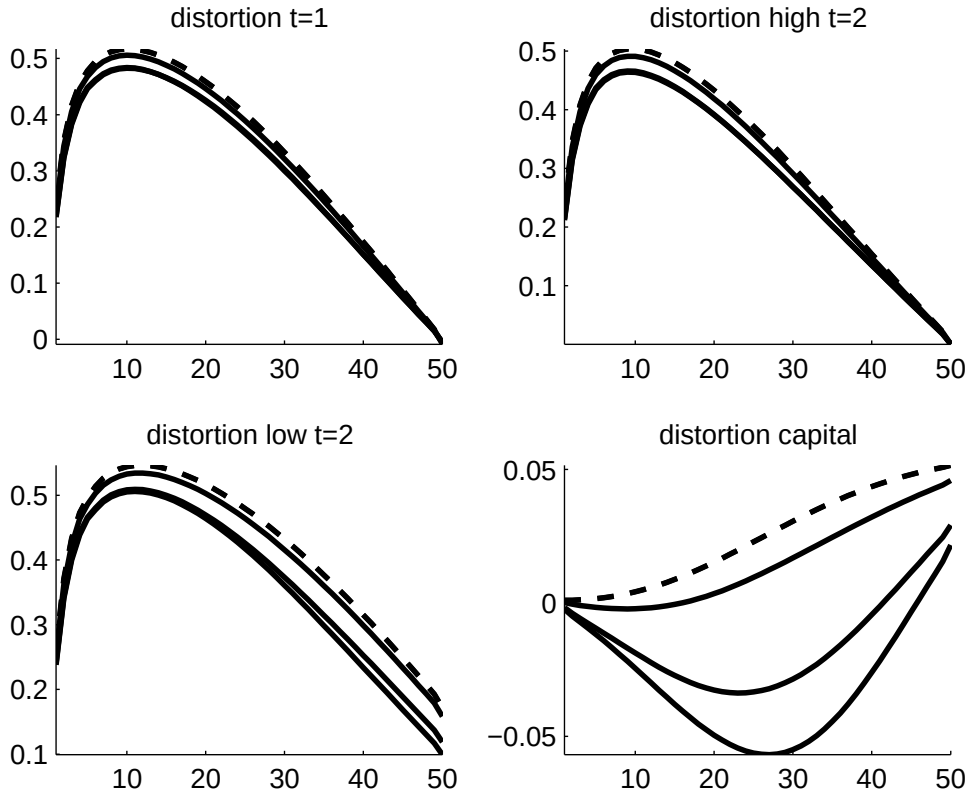


Figure 8: Nonseparable utility with $\sigma \leq 1$.

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. The baseline with $\sigma = 1$ is plotted as the dotted line. Away from the baseline, higher σ correspond to lower lines on the graph.

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 first plot the results for $\sigma = 2$. The capital wedge is negative in a range. We now plot the results for $\sigma = 0.65$ and see that the capital wedge is positive.

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

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

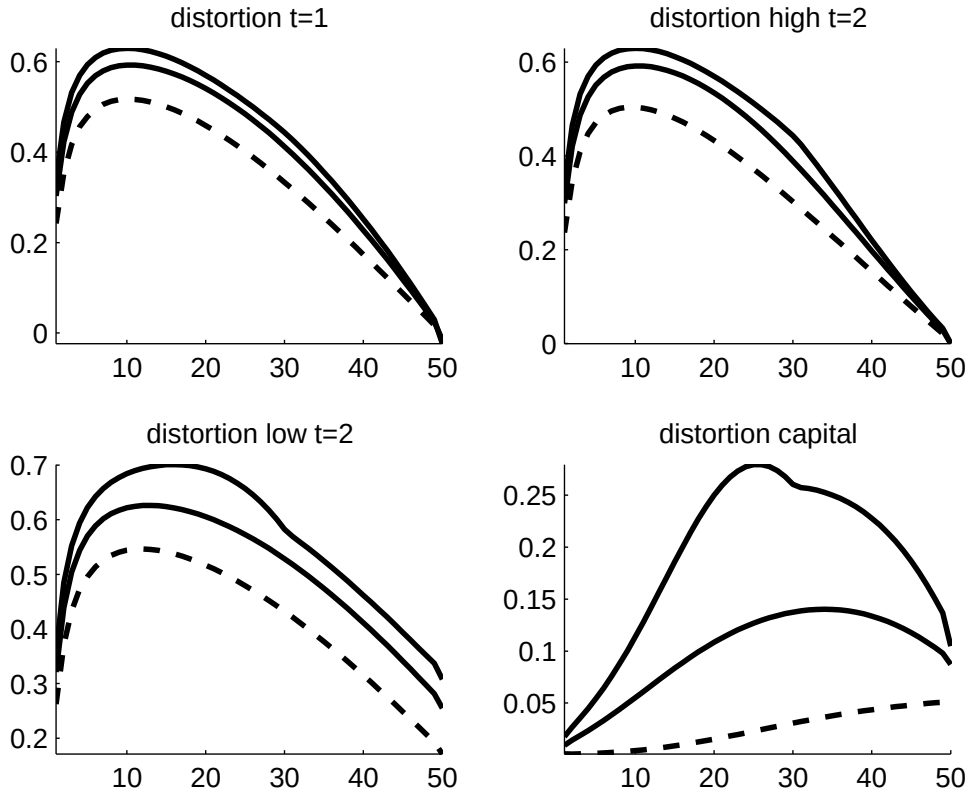


Figure 9: Nonseparable utility with $\sigma \geq 1$.

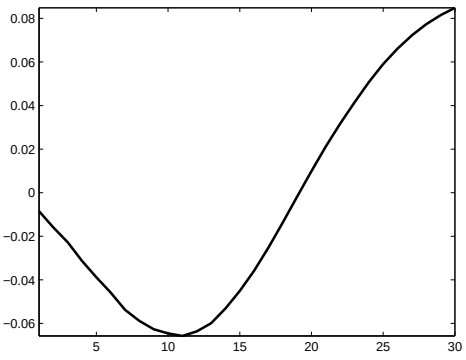


Figure 10: Nonseparable utility.

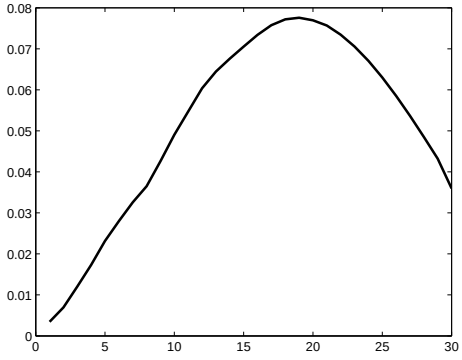


Figure 11: Nonseparable utility.

This utility function was used by Chari, Christiano, and Kehoe (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 both return and government expenditure shocks are similar in that they both shock 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 increases in government expenditure. Suppose we increase $G_1 = G_2(1) = G_2(2) = 0.2$, i.e., there is no aggregate uncertainty. Figure 12 shows labor wedges for this case. We plot in black the benchmark case of no government expenditures, $G_1 = G_2(1) = G_2(2) = 0$,and in red the case of $G_1 = G_2(1) = G_2(2) = 0.2$. We see that higher G leads to significantly higher labor

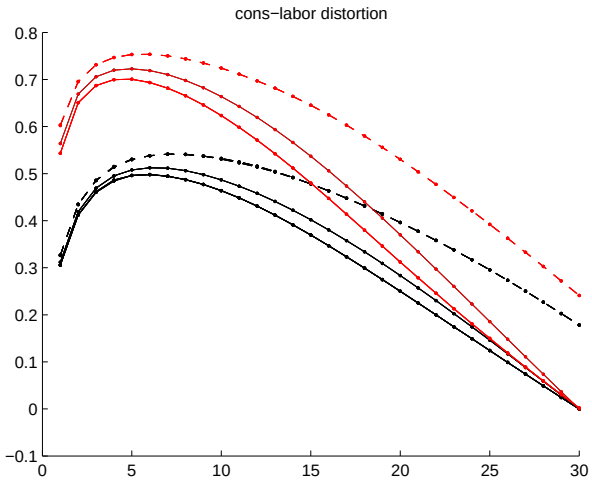


Figure 12: Labor Distortion

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 13 we plot the intertemporal wedges for our case of government expenditures (in red) and for the case of no government expenditures (in black). 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.

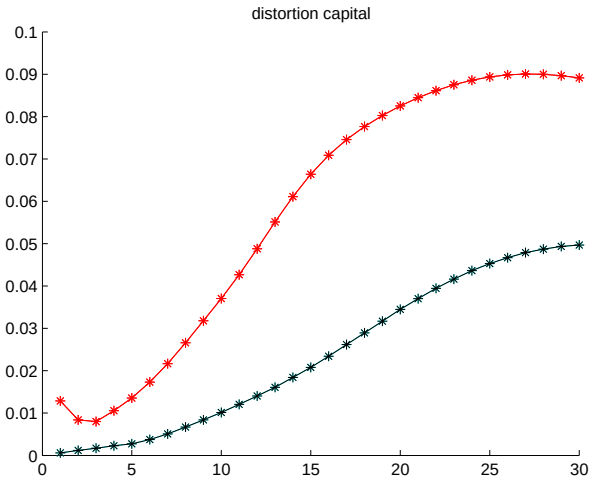


Figure 13: Intertemporal Distortion

We could have always 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 14 we plot labor wedges. The black line is τ_{y_1} ; the blue solid line is $\tau_{y_{1_1}}(.,1)$ (i.e. high type in state 1); the blue dashed line is $\tau_{y_{1_1}}(.,2)$ (i.e. low type in state 1); the red solid line is $\tau_{y_{1_2}}(.,1)$ (i.e. high type in state 2); the red dashed line is $\tau_{y_{1_2}}(.,2)$ (i.e. low type in state 2)

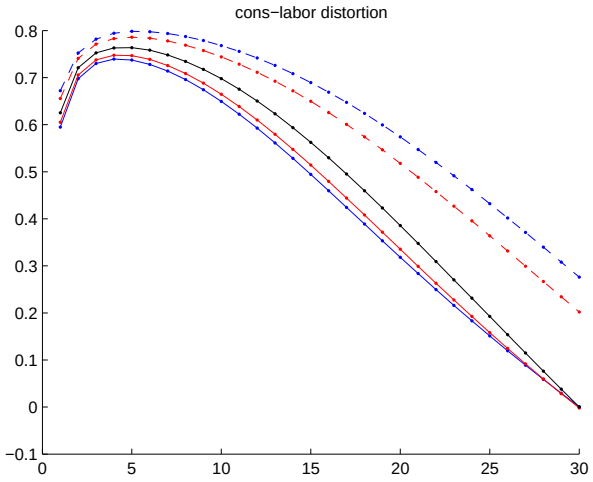


Figure 14: Shocks to government expenditure

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 calls 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 12, 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 15 that shows the intertemporal distortion. in that figure, the upper line is $\mu_1 = 0.7$, the solid line is $\mu_1 = 0.5$ and the lower line is $\mu_1 = 0.3$.

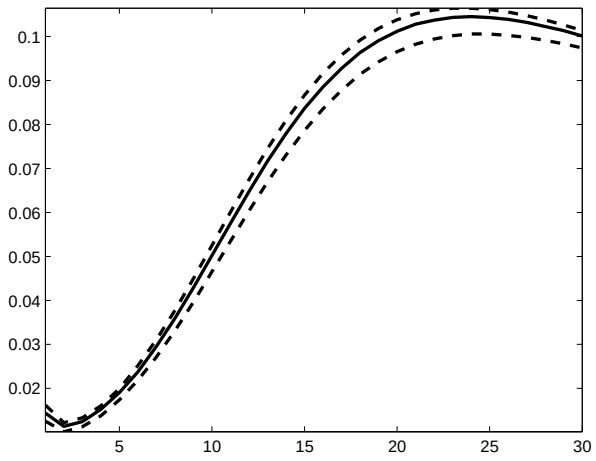


Figure 15: Intertemporal distortion

We see that intertemporal wedge becomes higher the higher μ_1 is, indicating a higher 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 16 we plot labor distortions. We plot labor wedges as follows. The black line is τ_{y_1}; the blue solid line is $\tau_{y_{1_1}}(\cdot, 1)$ (i.e., wedge for the high shock type in state 1); the blue dashed line is $\tau_{y_{1_1}}(\cdot, 2)$ (i.e., wedge for the low type in state 1); the red solid line is $\tau_{y_{1_2}}(\cdot, 1)$ (i.e., wedge for the high type in state 2); the red dashed line is $\tau_{y_{1_2}}(\cdot, 2)$ (i.e., wedge for the low type in state 2).

As in the case of government expenditure shocks, here we also observe that the spread between wedges on low and high type in a bad state are higher, 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 17 plots the intertemporal distortion τ_k for various values of the shock to the rate of return $R_2 = 1$, 1.2, 2, 3 and 4. The red line shows $R_2 = 4$ while the cyan line shows $R_2 = 1$ (i.e. a case with no uncertainty).

We see that distortions decrease with the rate of return shock R_2. Intuitively, a higher R leads to more resources, with more resources one could 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

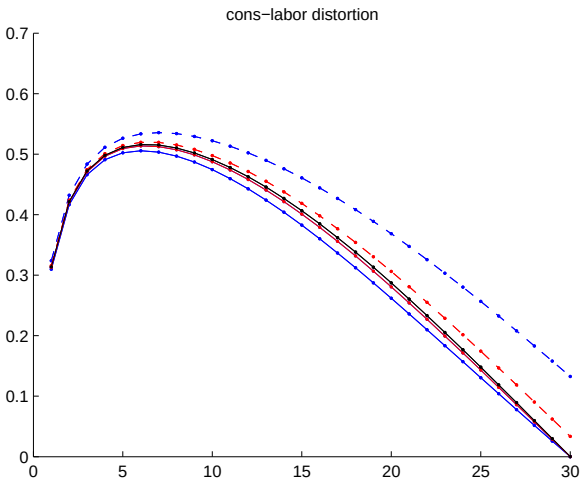


Figure 16: Rate of Return Shocks

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 Conclusions

In this paper we reviewed some main results of the recent literature on *New Dynamic Public Finance* in a tractable two-period model. In addition, we explored how capital and labor wedges, are affected by the model's parameters and how they respond to aggregate shocks.

We argued that this dynamic Mirrlees literature may be an important alternative to Ramsey models of taxation. Ramsey models have developed many important insights into optimal policy. However, as is well understood, their limitation is the ad hoc nature of the tax instruments assumed. The main premise of Mirrleesian optimal taxation literature is to model heterogeneity or uncertainty, which gives rise for a desire to redistribute or insure. An informational friction then prevents the first-best

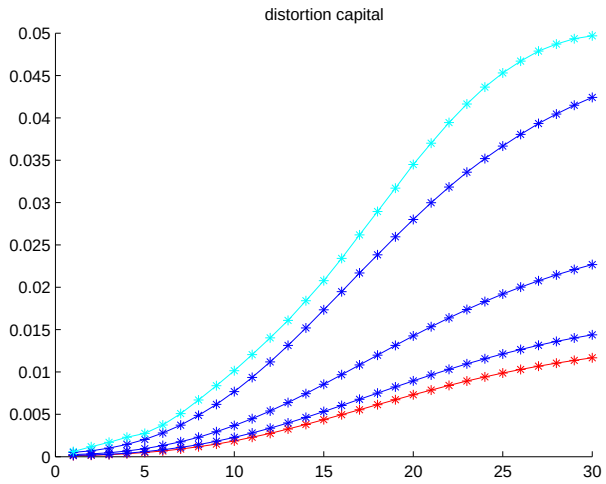


Figure 17: Intertemporal distortion varying R_2.

allocation and endogenously determines the feasible tax instruments.

We also argued that, in addition to having more solid theoretical foundations, Mirrleesian models have novel implications for the dynamic policy issues that Macroeconomists have been interested in: capital taxation, the smoothing of labor income taxes, and the nature of the time consistency problem. In addition, some new issues directly arise 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.

Given the numerical focus of this paper we first discuss what we think the challenges are in this area. It is important to analyze quantitative implications of the theory in more standard multiperiod calibrated model. The goal is to explore a plausible parameterized model with a realistically calibrated skill dynamics such as in, for example, Storesletten, Telmer, and Yaron (2004). The main difficulty is that there are virtually no methods for solving multiple period models with such a general structure of skill shocks. One interesting recent route is work by Farhi and Werning (2006) who study partial reforms in a dynamic Mirrlees setting to evaluate the gains from distorting savings. They provide a simple method that is tractable even with rich skill dynamics. There is also some preliminary progress in analyzing dynamic Mirrlees models with persistent shocks using a first-order approach in Kapicka (2005).

An ideal goal would be to derive a rich set of quantitative predictions similar in spirit to what a quantitative Ramsey model such as Chari, Christiano, and Kehoe (1994) deliver. Within a quantitative model one can also address a common criticism

of *New Dynamic Public Finance* literature that it delivers tax schedules that are "too complicated". For example, one could compare welfare of the fully optimal scheme to ones where some elements of the tax code are simplified. It would be interesting to compute welfare losses of a tax system comprised of linear taxes on capital and nonlinear labor income tax, or a tax system that has a limited dependence on history.

The main reason we stress the importance 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 interest later. One possible explanation is that the approach was difficult to apply quantitatively. We hope that we now will be able to to solve much more complicated dynamic models and use them to guide policy. Some policy relevant quantitative work is already emerging from this framework (see for example, disability insurance in Golosov and Tsyvinski, 2006), but more is needed.

Another route to take is to take some of the insights in the nature of optimal taxes from dynamic Mirrlees models and include them in the Ramsey style models of optimal taxation. The papers by Conesa and Krueger (2005) and Smyth (2005) may be interpreted as one step towards that goal. These papers compute the optimal tax system in a model where the tax function is still exogenously given but is flexibly parameterized to allow for a variety of taxes including progressive taxes, uniform lump sum taxes, and various exemptions. We think more work can be done in that area to use state-of-the-art computational models to consider general tax systems perhaps including joint conditioning on capital and labor incomes or some form of history dependence. This may be fruitful area of research to those who are also interested in the design of optimal Social Security systems where history dependent taxes are more natural to arise. Another important area of research in this quantitative spirit is to consider implications of *New Dynamic Public Finance* to 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 delivered in the Ramsey traditions.

The second main direction of future research that we see as important is to move away from the assumption of the mechanisms run by a fictitious benevolent social planner. A relevant and important question in this context is whether the insights of the dynamic Mirrlees literature apply to real world situations where politicians care about reelection, self-enrichment or their own individual biases, and cannot commit to sequences of future policies. To this end, Acemoglu, Golosov, and Tsyvinski (2006) consider both the informational constraints on tax instruments and the incen-

tive problems associated with politicians are taken into account. A related question is under what conditions markets can be better than optimal mechanisms. The potential misuse of resources and information by the government makes mechanisms less desirable relative to markets than in the standard mechanism design approach. 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. We think of this question as one of the central issues that need to be addressed.

Finally, we want to point out that *New Dynamic Public Finance* can be used to analyze a larger variety of new topics. One interesting venue of research is to consider intergenerational issues. Phelan (2005) and Farhi and Werning (2005) consider how intergenerational incentives should be structured. Farhi and Werning (2005) and Farhi, Kocherlakota, and Werning (2005) derive implications for optimal estate taxation. This is just one example of how the study of New Dynamic Public Finance models promise more than new answers to old questions, but can also lead to insights for a large set of new 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.[13]

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

References

Abraham, A., and N. Pavoni, 2003. "Efficient Allocations with Moral Hazard and Hidden Borrowing and Lending." .

Acemoglu, Daron, Mikhail Golosov, and Aleh Tsyvinski, 2006. "Markets versus Governments: Political Economy of Mechanisms." mimeo, M.I.T.

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

Albanesi, Stefania, 2006. "Optimal Taxation of Entrepreneurial Capital Under Private Information." mimeo, Columbia University.

Albanesi, Stefania, and Christoper Sleet, 2006. "Dynamic Optimal Taxation with Private Information." *Review of Economic Studies*. forthcoming.

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

Arnott, R., and J. Stiglitz, 1990. "The Welfare Economics of Moral Hazard." in H. Louberge (ed.), *Information and Insurance: Essays in Memory of Karl H. Borch* . pp. 91–121, Norwell, MA: Kluwer.

Atkeson, Andrew, and Robert E., Jr. Lucas, 1992. "On Efficient Distribution with Private Information." *Review of Economic Studies*. v59 (3): 427–453.

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

Barro, Robert, 1979. "On the Determination of the Public Debt." *Journal of Political Economy*. 87 (5): 940–71.

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

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

Bisin, Alberto, and Adriano Rampini, 2006. "Markets as Beneficial Constraints on the Government." *Journal of Public Economics*. forthcoming.

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

Chamley, Christophe, 1986. "Optimal Taxation of Capital in General Equilibrium." *Econometrica*. 54: 607–22.

Chari, V.V., Lawrence J. Christiano, and Patrick J. Kehoe, 1994. "Optimal Fiscal Policy in a Business Cycle Model." *Journal of Political Economy*. 102 (4): 617–52.

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

Conesa, Juan Carlos, and Dirk Krueger, 2005. "On the Optimal Progressivity of the Income Tax Code." *Journal of Monetary Economics*. forthcoming.

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

da Costa, Carlos, and Iván Werning, 2005. "On the Optimality of the Friedman Rule with Heterogeneous Agents and Nonlinear Income Taxation." working paper, MIT.

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

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

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

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

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

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

Farhi, Emmanuel, Narayana Kocherlakota, and Iván Werning, 2005. work in progress.

Farhi, Emmanuel, and Iván Werning, 2005. "Inequality, Social Discounting and Progressive Estate Taxation." NBER Working Papers 11408.

Farhi, Emmanuel, and Ivan Werning, 2006. "Capital Taxation: Quantitative Implications of the Inverse Euler equation." mimeo.

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

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

Golosov, M., and A. Tsyvinski, 2005. "Optimal Taxation with Endogenous Insurance Markets." *NBER working paper #11185*.

Golosov, M., and A. Tsyvinski, 2006. "Designing Optimal Disability Insurance: A Case for Asset Testing." *Journal of Political Economy*. forthcoming.

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

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

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

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

Kapicka, Marek, 2005. "Optimal Taxation with Human Capital Accumulation." University of California-Santa Barbara.

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

Kocherlakota, Narayana, 2005a. "Advances in Dynamic Optimal Taxation." mimeo.

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

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

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

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

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

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

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

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

Saez, Emmanuel, 2001. "Using Elasticities to Derive Optimal Income Tax Rates." *Review of Economic Studies*. 68 (1): 205–29.

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

Sleet, Christopher, and Sevin Yeltekin, 2005. "Credible Social Insurance." mimeo.

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

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

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

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

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

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

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

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