MIKHAIL GOLOSOV
MAXIM TROSHKIN
ALEH TSYVINSKI

Optimal Taxation: Merging Micro and Macro Approaches

This paper argues that the large body of research that follows Mirrlees approach to optimal taxation has been developing in two directions, referred to as the micro and macro literatures. We review the two literatures and argue that both deliver important insights that are often complementary to each other. We argue that merging the micro and macro approaches can prove beneficial to our understanding of the nature of efficient redistribution and social insurance and can deliver implementable policy recommendations.

EFFICIENT PROVISION OF social insurance and efficient redistribution of resources among individuals are some of the most important and challenging questions in macroeconomics and public finance. A seminal contribution of Mirrlees (1971) is the starting point for the modern approach to answering these questions. A trade-off between efficiency and insurance or equity is inherent to this approach and is a key determinant of the optimal policy.

In this paper, we argue that the large body of research that follows Mirrlees approach has been developing in two quite separate directions—referred to in this paper as the *micro* and *macro approaches*. We argue that merging the two directions

can help develop new insights into optimal taxation and ultimately into the nature of efficient social insurance and redistribution policies.

We start with what we call the *micro approach* to optimal taxation. It originates with Mirrlees (1971, 1976, 1986)[1] and is more recently carried out primarily by public finance economists such as Diamond (1998) and Saez (2001). The micro approach is generally static.[2] That is, there is no uncertainty about future shocks and individuals in the modeled environment make no savings decisions. Crucially, individuals are assumed to be heterogeneous with respect to their productivities or skills, while the government does not directly observe workers' skills and work efforts. Unobservable skills create an information friction. The key trade-off in these optimal taxation environments is between offering insurance—or, alternatively, redistributing resources—and providing correct incentives to work.

The micro approach proceeds by characterizing optimal distortions that directly translate into optimal taxes in static environments. One advantage of the literature exercising this approach is then a clear connection between the parameters of the optimal tax policy in the model and empirical data. A strong feature of the micro approach is that if one believes its static environment to be relevant then concrete policy recommendations for tax code reforms can be made. In Section 1, we illustrate within a simple static model the approach of micro literature and the main insights it offers as well as its limitations.

Many important classical questions in public economics and macroeconomics are, however, inherently dynamic. Workers' skills change stochastically over time and the question of designing optimal taxation policy has an important dynamic dimension. For instance, to be able to explore the optimal taxation of savings in the presence of stochastic shocks, a dynamic framework is necessary. Many other macroeconomic and public finance problems are intrinsically dynamic as well: How to design optimal social insurance? How should labor income and consumption be taxed over the life cycle? Should the government tax bequests? Should education be subsidized?

The *macro approach* to optimal taxation extends the static framework of Mirrlees (1971) to dynamic environments to be able to address questions such as the ones above. A more recent strand of this literature—which we refer to as the *New Dynamic Public Finance*[3]—develops new insights about optimal taxation in dynamic settings.[4] The macro approach typically assumes rich dynamic structure. Uncertainty about future shocks plays a central role—stochastically evolving productivities are the essence of dynamics in the model.[5] This literature offers both a framework for the

1. See also, among numerous other studies, Sadka (1976), Seade (1977), and Tuomala (1990).

2. An important exception is Diamond and Mirrlees (1978).

3. For surveys of this part of the macro literature see Golosov, Tsyvinski, and Werning (2006) and Kocherlakota (2010).

4. For earlier contributions see, for example, Diamond and Mirrlees (1978), Atkinson and Stiglitz (1976), and Stiglitz (1987).

5. The micro approach can also be used to study dynamic issues such as optimal taxation of capital, but only in the environments in which productivities do not change. For example, in Atkinson and Stiglitz

analysis of many challenging dynamic taxation questions and a range of applications for this framework.

Although recently the macro literature has been making significant progress, as any literature, it still leaves many important questions unanswered. First, only partial characterizations of optimal allocations are available in general. Once the dynamics are added to the model, obtaining its solution becomes complex. Second, optimal taxes that implement the optimal allocations depend on the particulars of implementation. In addition, it is important for the macro literature to be explicit about how private insurance markets operate. The macro approach addresses efficient provision of social insurance and hence the insights and the policy prescriptions of the dynamic macro literature depend on the availability of private insurance.

A key outstanding issue is thus the development of concrete, data-based policy implications of dynamic public finance. Banks and Diamond (2008) argue in the "Mirrlees Review" for the importance of the Mirrlees approach, both static and dynamic, as a guide to policy.[6] By appealing to recent results in Golosov, Troshkin, and Tsyvinski (2010b), we argue in this paper that progress can be made by merging the micro and macro approaches to deliver implementable policy prescriptions. Importantly, we show that considering dynamic models significantly changes optimal policy prescriptions based on the static micro approach.

The rest of the paper is organized as follows. In Section 1, we use a simple model to illustrate the micro approach and review some of the main insights it offers. In Section 2, we do the same for the macro approach. We argue that the approaches of both literatures deliver valuable insights, many of which complement each other. Section 3 suggests directions to merge the micro and macro approaches and reviews recent results in this area. We argue that merging the two approaches can help make progress in our understanding of optimal taxation and ultimately of the nature of efficient redistribution and social insurance policies as well as provide policy relevant results. To make the exposition more concrete, throughout Sections 1, 2, and 3, we discuss the results of quantitative studies based on empirical data and realistic parameter values. In Section 4, we review related literature on political economy and taxation. Section 5 concludes.

1. MICRO APPROACH

In this section, we use a static optimal taxation model, based on the environment in Mirrlees (1971), to illustrate the approach of micro literature, the insights it offers, and its drawbacks. We start by presenting the basics of the static setup. Next, we

(1976), one can interpret an environment with many consumption goods as that of many periods. However, as unobservable skills remain constant, the model is essentially static.

6. Commissioned by the Institute for Fiscal Studies, the Review is the successor to the influential "Meade Report" (Meade 1978) and is an authoritative summary of the current state of tax theory as it relates to policy.

analyze the main insights it offers into what determines optimal marginal tax rates. Then, we examine how those insights extend to generalized static settings and how they connect to empirical data. We review the results of several numerical simulation studies based on empirical data and realistic parameter values. Finally, we point out the main limitations of the micro approach.

1.1 Static Setup

Consider a static economy populated by a continuum of agents of unit mass. Each agent derives utility from a single consumption good and disutility from work effort according to $U(c, l)$, where $c \in \mathbb{R}_+$ denotes the agent's consumption of the single consumption good and $l \in \mathbb{R}_+$ denotes the work effort of the agent. Assume that $U : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ is strictly concave in c, strictly convex in l, and twice continuously differentiable.

The agents in this economy are heterogeneous. Each agent has a type $\theta \in \Theta \equiv [\underline{\theta}, \bar{\theta}]$, where $\underline{\theta} > 0$ and $\bar{\theta} \le \infty$, drawn from a distribution $F(\theta)$ with density $f(\theta)$. From the point of view of an individual agent, $f(\theta)$ represents *ex ante* probability of being type θ. Alternatively, $f(\theta)$ can be interpreted at the aggregate level as the measure of agents of type θ, assuming the law of large numbers holds.

An agent of type θ, who supplies l units of effort, produces $y = \theta l$ units of output of the consumption good. Thus, one can think of type, θ, as representing productivity or skill. The following information friction is present. The type, θ, of an agent as well as his effort supply, l, are private information, that is, they are known only to the agent. Output, y, and consumption, c, are public information, that is, observable by all.

An *allocation* in this economy is (c, y), where

$$c : \Theta \to \mathbb{R}_+,$$
$$y : \Theta \to \mathbb{R}_+.$$

Aggregate *feasibility* requires that aggregate consumption does not exceed aggregate output:

$$\int c(\theta)dF(\theta) \le \int y(\theta)dF(\theta), \tag{1}$$

where $c(\theta)$ and $y(\theta)$ are consumption and output, respectively, of an agent of type θ.

This economy has a benevolent government that can *ex ante* choose a tax system and fully commit to it. The social objective is to maximize social welfare G, where G is a real-valued increasing and concave function of individual utilities. The government then chooses taxes $T(y)$ optimally, that is, to achieve the social objective subject to the aggregate feasibility.[7]

---

7. In applications, the government can be required to also finance government revenue $\bar{G} \ge 0$ so that the aggregate feasibility is $\int c(\theta)dF(\theta) + \bar{G} \le \int y(\theta)dF(\theta)$.

One approach to analyzing this environment is well known since the seminal work of Mirrlees (1971).[8] It in turn builds on the foundation provided by the mechanism design theory pioneered by Hurwicz (1960, 1972).[9] The approach is to realize that the solution to the government's problem is equivalent to the solution to a mechanism design problem. In the mechanism design problem, all agents report their types to a fictitious social planner who allocates feasible consumption and output subject to incentive compatibility; that is, the planner chooses feasible $c(\theta)$ and $y(\theta)$ so that no agent has incentives to lie about his type.

The solution is then a two-step procedure. In the first step, appealing to the revelation principle of the mechanism design, an *optimal allocation* is found as a solution to the mechanism design problem. In the mechanism design problem, the planner receives reports $\sigma(\theta)$: $\Theta \rightarrow \Theta$ from the agents about their types (i.e., each agent makes a report about his own type) and allocates feasible consumption and output $\{c(\theta), y(\theta)\}_{\theta \in \Theta}$ as functions of the agents' reports. *Incentive compatibility* constraint ensures that no agent finds it beneficial to lie about his type:

$$U(c(\theta), y(\theta)/\theta) \geq U(c(\theta'), y(\theta')/\theta) \qquad \text{for all } \theta, \theta'. \tag{2}$$

The optimal—or constrained efficient—allocations thus solve the planner's problem of maximizing the social welfare function:

$$\max_{\{c(\theta), y(\theta)\}_{\theta \in \Theta}} \int G(U(c(\theta), y(\theta)/\theta)) dF(\theta) \tag{3}$$

subject to the aggregate feasibility constraint (1) and the incentive compatibility constraint (2). Let $\{c^*(\theta), y^*(\theta)\}_{\theta \in \Theta}$ denote a solution to this problem.

The second step is *implementation*, that is, characterization of optimal taxes $T(y)$ that decentralize—or implement—an optimal allocation. In this static setting, finding taxes that implement an optimal allocation is straightforward. Define a marginal distortion, or a wedge, $\tau'(\theta)$ by

$$1 - \tau'(\theta) = \frac{-U_l(c^*(\theta), y^*(\theta)/\theta)}{\theta U_c(c^*(\theta), y^*(\theta)/\theta)}, \tag{4}$$

where U_c and U_l denote partial derivatives of the utility function with respect to c and l, respectively, and $\{c^*(\theta), y^*(\theta)\}_{\theta \in \Theta}$ is the optimal allocation. That is, $\tau'(\theta)$ is a measure of how distorted individual agent's decisions are in the optimal allocation versus what they normally would be in a full information *ex ante* optimum.[10] To find the optimal taxes $T(y)$, we notice that in this static environment optimal wedges

8. For a textbook treatment see Salanie (2003).

9. Some of the standard textbook expositions of the mechanism design theory are Fudenberg and Tirole (1991, chap. 7), and Mas-Colell, Whinston, and Green (1995, chap. 23).

10. The full information version of the planner's problem does not require incentive compatibility (2). Thus, its first-order conditions imply that $U_c(c(\theta), y(\theta)/\theta) = U_l(c(\theta), y(\theta)/\theta)/\theta$ for all θ, implying that $\tau'(\theta) = 0$ for all θ. In other words, lump-sum taxes implement the optimal allocation.

directly translate into optimal marginal taxes. In particular, the optimal marginal income tax on type θ, $T'(\theta)$, is given by the wedge in the consumption-labor margin:

$$T'(\theta) = \tau'(\theta).$$

1.2 Insights from Static Environments

One way to explore what this environment suggests about optimal policy is to follow the two-step procedure described above. First, one characterizes the optimal allocations as much as possible. That is, one characterizes the solution to the mechanism design problem (3) and, in particular, examines whether the characterization implies that any individual decisions must be distorted compared to what they normally would be in a full information *ex ante* optimum. Then, one notices that in this static environment optimal marginal distortions, if any, directly translate into optimal marginal taxes. In short, to gain insights into optimal policy, one can characterize constrained efficient allocations and derive results about optimal taxes that implement them.

There are relatively few general insights that can be gained by following this path.[11] We point out the two most sharp and general results. First, optimal marginal tax rates lie between 0 and 1 (Mirrlees 1971). Second, optimal marginal tax rates equal 0 at the top end of the skill distribution and, unless there is a positive measure of agents at the bottom end, optimal marginal tax rates also equal 0 at the very bottom of the skill distribution (Sadka 1976, Seade 1977).

The result about zero marginal tax rate at the top end of the skill distribution (sometimes referred to as "no distortion at the top") is somewhat striking and controversial. However, it is a local result (see Tuomala 1990, chaps. 1 and 6) in the sense that it does not imply that marginal tax rates near the top end of the skill distribution are zero or near zero.

Although the result itself is of limited use, the intuition behind the zero marginal tax rate at the top is instructive. First, note that total tax revenue depends on *average* tax rate, while incentive compatibility is affected by *marginal* tax rates. Now, suppose the marginal tax rate on the top individual in the skill distribution is slightly decreased. Then, she has increased incentive to work but, since the average tax rate is unchanged (as is the rest of the model), the total tax revenue is the same. If this additional incentive effect on the top skill individual is not negligible, then she will increase her income and the total tax revenue will also increase. That is, the top individual is better off without anyone else being worse off. Clearly, this argument can be repeated until the marginal tax rate at the top is zero. There are no agents above the agent with the highest skill and no lower types are better off by claiming to be the highest type. There is no need to distort the highest type's allocations then to provide incentives. Notice also that this argument does not need to work for the next to the top individual

---

11. In particular, Mirrlees (1971) originally analyzes this problem in general form, that is, without assuming specific utility function or the distribution of skills. In this general case, he is able to derive only very weak conditions characterizing optimal tax policies.

since lowering his marginal tax rate will also increase incentives for the top individual to misrepresent herself as a lower type.

Starting already from Mirrlees (1971), it has been realized that based on such general analysis alone it is difficult to develop concrete tax policy guidance. Consequently, from the very beginning, the micro literature attempted to further its insights by using computational methods. The use of numerical calculations is also justified by the very nature of the optimal taxation problem, which requires quantitative answers.

Mirrlees (1971) provides some of the first numerical examples in his attempt to gain further understanding of optimal income tax policy. He uses utilitarian social welfare function, that is, $G(U) = U$, log-linear utility function, and a skill distribution based on the UK wage data. He finds that optimal marginal tax rates are quite low and not monotonically increasing, that is, optimal income tax is not progressive throughout. In particular, Mirrlees concludes that the optimal tax schedule is approximately linear.

Subsequent quantitative work (see, e.g., Stern 1976, Tuomala 1990) questions the implicit assumption about the elasticity of substitution between consumption and work effort implied by the choice of log-linear utility function. The argument is that log-linear utility implies excessive costs of making the tax schedule progressive. Notably, Tuomala (1990, chap. 6) uses a range of realistic values of the elasticity of substitution between consumption and work effort and finds that the optimal tax schedule is substantially nonlinear. He also finds significantly higher optimal marginal tax rates—up to 70% for the utilitarian social objective and up to 90% for maximin social objective, that is, Rawlsian principle. The optimal marginal tax rates in Tuomala (1990) are not monotonically increasing.[12]

1.3 Extension and Connection to Data

Although it provides the foundation for a large body of literature, the general analysis outlined above has few concrete applications as its insights are difficult to relate to policy. An important step forward that brings the static micro approach substantially closer to being policy related is Diamond (1998) and Saez (2001). In static Mirrlees models, Diamond (1998) and Saez (2001) derive easily interpretable formulas for optimal marginal tax rates in terms of elasticities and the shape of income distribution. The elements of the formulas easily connect to empirically observable data. Their work provides an reinterpretation of the first-order conditions for the optimal planning problem and gives insights into forces determining the optimal tax rates.

Diamond (1998) assumes a general increasing and concave social welfare function G and quasi-linear preferences of the form

$$U(c, l) = c + v(1 - l), \tag{5}$$

12. In fact, Tuomala (1990) concludes that in a static Mirrleesian setting "it is difficult (if at all possible) to find a convincing argument for a progressive marginal tax rate structure throughout" (p. 14).

where $v(\cdot)$ is assumed to be strictly concave and twice continuously differentiable. The assumption of quasi-linear preferences implies no income effects. This has an advantage of simplifying the analysis; however, as we discuss later, Saez (2001) shows that the main results of Diamond (1998) can be generalized to preferences with income effects.

Diamond (1998) shows that when preferences satisfy (5), the optimal marginal taxes must satisfy

$$\frac{T'(\theta)}{1 - T'(\theta)} = \left(1 + \frac{1}{\varepsilon(\theta)}\right)\left(\frac{1 - F(\theta)}{\theta f(\theta)}\right)\left(\int_\theta^\infty \left(1 - \frac{G'(U)U(x)}{\lambda}\right)\frac{dF(x)}{1 - F(\theta)}\right),$$

(6)

where $\varepsilon(\theta)$ is the elasticity of labor supply of type θ and λ is the Lagrange multiplier on the government's budget constraint and is given by

$$\lambda = \int_0^\infty G'(U)U(x)dF(x).$$

Equation (6) is a useful representation of the first-order conditions for the planner's problem (3) because it offers intuition for the forces determining optimal marginal taxes. Equation (6) does not represent a closed-form solution for the optimal marginal taxes, $T'(\theta)$. The reason is the integral on right-hand side of equation (6) that depends on the optimal level of utility, U. Consider, for instance, the effects of a lower elasticity of labor supply, $\varepsilon(\theta)$, for some θ. There is a direct effect on the optimal marginal tax rate via an increase in the first term on the right-hand side of equation (6). There is also, however, an indirect effect via the term $G'(U)U$, which is endogenously determined by the optimal allocation.

Nevertheless, equations such as (6) proved to be useful in applications as the intuition they provide often closely matches the direct numerical calculations of the optimal marginal taxes. For examples of that see Diamond (1998), Saez (2001), Weinzierl (2008), Golosov et al. (2010), and Golosov, Troshkin, and Tsyvinski (2010b).

Equation (6) suggests that the optimal marginal tax rates in the static economy are influenced by three key terms that are easily interpretable and can be inferred from empirical data.

The first term, $1 + 1/\varepsilon(\theta)$, is related to the elasticity of labor supply. The more elastic labor supply is, the more distortionary marginal labor taxes are. Thus, higher elasticity of labor supply acts as a force driving the magnitude of the optimal marginal tax rates lower.

The second term on the right-hand side of equation (6) is a tail ratio of the skill distribution, $(1 - F(\theta))/(\theta f(\theta))$. The intuition behind the force provided by this term on the optimal tax rate is the following. A positive marginal tax on a type θ prevents all types above θ from claiming to be θ and receiving the corresponding allocation. If the measure of agents who are more productive than θ is high, that is, $1 - F(\theta)$ is high, an optimal marginal tax on type θ must provide stronger incentives to report type truthfully. This provides a driving force for higher optimal marginal tax on θ.

On the other hand, if the measure of agents of type θ is high, that is, $f(\theta)$ is high, or if they are highly productive, that is, θ is high, then optimal marginal tax on type θ is particularly distortionary. This creates a driving force for lower optimal marginal tax on θ.

Finally, the third term on the right-hand side of equation (6) depends on the curvature of the social welfare function G, which captures the desired degree of redistribution. More concave G tends to raise the third term. Therefore, more redistributive social objective generally acts as a force for higher optimal marginal taxes.

Equations such as (6) can often be used to derive results about the optimal policy. In particular, Diamond (1998) uses equation (6) to prove that optimal marginal taxes are U-shaped if the distribution of skills is single-peaked, with the peak not at the bottom of the distribution, and a Pareto distribution above the peak. That is, given such distribution of skills, for all agents with skills above a certain cutoff the optimal marginal tax is first decreasing up to a certain level of income and monotonically increasing after that. Assuming a Pareto distribution of skills above the modal skill, Diamond (1998) also uses equation (6) to derive the expression for the asymptotic optimal marginal tax. For instance, for any social welfare function G with a property that $\lim_{U \to \infty} G'(U) = 0$, and individual preferences represented by (5), the asymptotic optimal marginal tax rate is given by

$$\lim_{\theta \to \infty} \frac{T'(\theta)}{1 - T'(\theta)} = \frac{1}{a}\left(1 + \frac{1}{\varepsilon(\theta)}\right), \tag{7}$$

where a is the parameter of the Pareto distribution.

Saez (2001) further extends and generalizes this approach. He shows that the results of Diamond (1998) can be extended to preferences with income effects. Saez argues that, while present, the dependence of the results on income effects is generally quite small. He provides a generalization of equation (6) for preferences with income effects. The right-hand-side terms of the generalized equation are still easy to interpret and compute using realistic elasticity parameters and empirical labor earnings distribution obtained from micro data.

Importantly, Saez (2001) numerically computes the optimal tax codes for realistically calibrated versions of the model. He uses the coefficients for income and substitution effects standard in the labor literature. He also uses a simplified representation of the actual U.S. tax code and an empirical distribution of labor earnings—based on the Internal Revenue Service tax returns data—to compute implied distribution function F. He then explores various social welfare functions, G, to study the effect of redistributional objectives.

The quantitative findings of Saez (2001) are consistent with a version of equation (6) and its implications for the shape of the optimal marginal tax and the asymptotic optimal marginal tax rate. In a static model calibrated to empirical cross-sectional distribution of labor income and empirical tax rates, he finds that optimal marginal taxes are U-shaped in the lower part of the income distribution, increase after that,

and the asymptotic tax rates are consistent with equation (7) and are quite high (50–70%).

1.4 Limitations of the Micro Approach

The static approach of the micro literature to exploring the optimal taxation of individuals and more generally the nature of efficient social insurance and redistribution policies comes with several drawbacks. The key drawbacks are the limitations embedded in static environments.

First, because the approach is static in its nature, it is silent about efficient insurance against idiosyncratic shocks over lifetime. The macro approach that we discuss in Section 3 below shows that the evolution of idiosyncratic shocks is one of the chief driving forces behind the optimal income taxation.

Second, just as importantly, a static environment cannot be useful in addressing optimal savings taxation when agents receive dynamic idiosyncratic shocks. Because the static micro approach is silent about optimal savings taxation in such environments, it does not offer a clear way to explore how labor decisions are affected by savings decisions and savings taxation. Studying the consequences of human capital accumulation decisions and, in particular, educational choices are similarly outside the limits of the static micro approach.

Nevertheless, as we discuss in Section 2, the methods of the micro approach can be used to shed light on dynamic optimal taxes and develop new insights into the optimal taxation and into the nature of efficient social insurance and redistribution policies.

2. MACRO APPROACH

Most of the drawbacks of the static micro approach are summarized by the fact that many important classical problems in public economics and macroeconomics are inherently dynamic. The macro approach extends the static framework of Mirrlees (1971) to dynamic environments to attempt to address these questions.

The macro literature typically makes the environment dynamic by assuming that agents live for $T \leq \infty$ periods and, importantly, that their skills evolve stochastically over time. When agents' skills do not change over time, a variation of the micro approach can be used to study intertemporal taxation. For example, in Atkinson and Stiglitz (1976), one can think of consumption of various goods as consumption over time and, therefore, study taxation of capital. It is essential to note that dynamics in the macro approach comes from the stochastic evolution of skills rather than from a repetition of the static Mirrlees model.

Most of the main insights of the macro approach can be developed with $T = 2$, which is what we do here for simplicity and the ease of exposition. We use this extended dynamic setting to illustrate the few general results that have been obtained

in dynamic environments. Then, we point out the challenges to macro approach posed by macroeconomic and public finance questions that are dynamic in nature.

2.1 Dynamic Environment

We consider a dynamic version of the environment in Section 1. Our goal here is to make as few adjustments to the setup in Section 1 as possible to introduce dynamics in a meaningful way. Once we have our dynamic environment, we can extend the analysis of optimal labor taxes developed in Section 1 to characterize the optimal labor and savings distortions in a dynamic economy and examine their implementations.

Consider an economy similar to that of Section 1 that, however, lasts for two periods: $t = 1, 2$. Every agent lives for two periods and has preferences represented by a lifetime utility function

$$\mathbb{E}_0 \sum_{t=1,2} \beta^{t-1} U\left(c_t, l_t\right),$$

where $c_t \in R_+$ is the agent's consumption in period t, $l_t \in R_+$ is the agent's work effort in period t, $\beta \in (0, 1)$ is the agent's subjective discount factor, and $\mathbb{E}_0$ is the expectation operator. The instantaneous utility function $U(c_t, l_t)$ is the same utility function we discuss in the static economy, except now consumption and work effort are time specific.

In each period t, agents draw their skill types, $\theta_t \in \Theta$. In period $t = 1$, skills are drawn from a distribution $F(\theta)$. Conditional on the realization of the shock θ in period $t = 1$, shocks θ' in period $t = 2$ are drawn from a conditional distribution $F(\theta'|\theta)$ with a conditional density $f(\theta'|\theta)$. Let $\theta^1 = \theta_1$, $\theta^2 = (\theta_1, \theta_2)$ be histories of shocks. The skill shocks and the histories of shocks are privately observed by respective agents and so are work efforts, l_t, and their histories. Output $y_t = \theta_t l_t$ and consumption c_t are observed by everyone, including the planner. Let $\Theta^1 = \Theta$ be the set of possible skill shock histories in period $t = 1$, and $\Theta^2 = \Theta \times \Theta$ be the set of possible skill shock histories in period $t = 2$. Denote by $c_t\left(\theta^t\right) : \Theta^t \to \mathbb{R}_+$ an agent's allocation of consumption and by $y_t\left(\theta^t\right) : \Theta^t \to \mathbb{R}_+$ an agent's allocation of output in period t. Denote by $\sigma_t(\theta^t): \Theta^t \to \Theta^t$ an agent's report in period t. It is easy to see how this environment generalizes to $T \leq \infty$.

Resources can be transferred between periods at the rate of $\delta > 0$ on savings. Assume that all savings are publicly observable.[13] Hence, without loss of generality, we assume that the social planner does all the saving in the economy by choosing the amount of aggregate savings.

13. The assumption of publicly observable savings is common to most of the macro literature. For a treatment of efficient insurance with unobservable savings see Allen (1985), Cole and Kocherlakota (2001), Werning (2002b), Shimer and Werning (2008), and in the context of dynamic optimal taxation Golosov and Tsyvinski (2007). See also Abraham and Pavoni (2008) for a two-period examination of the first-order approach with hidden savings as well as borrowing.

For further simplicity, as in much of Section 1, we assume that the social planner is utilitarian, that is, the social welfare function satisfies $G(U) = U$.[14] An *optimal allocation* is then a solution to the following dynamic mechanism design problem (see, e.g., Golosov, Kocherlakota, and Tsyvinski 2003):

$$\max_{\{c_t(\theta^t), y_t(\theta^t)\}_{\theta_t \in \Theta; t=1,2}} \mathbb{E}_0\{U(c_1(\theta^1), y_1(\theta^1)/\theta_1) + \beta U(c_2(\theta^2), y_2(\theta^2)/\theta_2)\} \qquad (8)$$

subject to the feasibility constraint

$$\mathbb{E}_0\{c_1(\theta^1) + \delta c_2(\theta^2)\} \leq \mathbb{E}_0\{y_1(\theta^1) + \delta y_2(\theta^2)\}$$

and the incentive compatibility constraint

$$\mathbb{E}_0\{U(c_1(\theta^1), y_1(\theta^1)/\theta_1) + \beta U(c_2(\theta^2), y_2(\theta^2)/\theta_2)\}$$
$$\geq \mathbb{E}_0\{U(c_1(\sigma_1(\theta^1)), y_1(\sigma_1(\theta^1))/\theta_1) + \beta U(c_2(\sigma_2(\theta^2)), y_2(\sigma_2(\theta^2))/\theta_2)\}$$
$$\text{for all } \sigma_t(\theta^t), t = 1, 2.$$

The expectation $\mathbb{E}_0$ above is taken over all possible realizations of histories. The first constraint in problem (8) is the dynamic feasibility constraint. The second constraint is a dynamic incentive compatibility constraint that states that an agent prefers to truthfully report his history of shocks rather than to choose a different reporting strategy.

Before we go on to discuss insights offered by this dynamic environment, we make two additional considerations. First, we need to consider private insurance markets. Since the macro literature addresses efficient provision of social insurance, one needs to take a stand on how private insurance markets operate. Clearly, whatever policy prescriptions are implied by the insights from the dynamic macro approach, they depend on the availability of private insurance. As it is done in much of the macro literature, we now look at one extreme case of no private insurance and seek to use this case to provide a useful benchmark. We return to the question of private insurance markets below and discuss some of the recent results about optimal dynamic taxation in the presence of private insurance.

Second, we need to consider how optimal Mirrleesian taxes compare to the actual tax codes. The theoretical framework we discuss here considers integrated systems of all taxes and all transfers. At the same time, for example, the U.S. tax system consists of statutory taxes and a variety of welfare programs. Thus, we are to think of labor distortions as being a sum of the distortions from all of those programs. One interpretation is that this calls for an integrated tax and social insurance system. In

---

14. Throughout, we assume that the planner can commit to the dynamic allocations. The environment without commitment is significantly more complicated as the revelation principle may not hold. For the analysis of such environments see, for example, Bisin and Rampini (2006), Acemoglu, Golosov, and Tsyvinski (2008a, 2008b, 2009a), Farhi and Werning (2008), and Sleet and Yeltekin (2009).

other words, a system where various social insurance programs are integrated into one tax code.

Next, we discuss the main general results and policy prescriptions that come from dynamic models of the macro literature. We examine the results about the characterization of optimal allocations first. Then, we consider implementation results in dynamic settings. We compare the results of the macro approach to the results from the static micro literature and discuss connections to empirical data.

2.2 Implicit Tax on Savings

One of the key general insights in dynamic environments of the macro literature is that when agents' productivities change stochastically over time it is optimal to introduce a positive marginal distortion—an implicit tax— that discourages savings. This distortion manifests itself as an inequality—or a wedge—between the intertemporal marginal rate of substitution and the marginal rate of transformation. More formally, a marginal savings distortion $\tau'_S(\theta)$ in our two-period setting is defined by

$$1 - \tau'_S(\theta) = \frac{\delta U_c(c_1(\theta), y_1(\theta)/\theta)}{\beta \mathbb{E}\{U_c(c_2(\theta^2), y_2(\theta^2)/\theta^2)|\theta\}},$$

where U_cs denote partial derivatives of the utility function with respect to consumption and evaluated at periods $t = 1$ and $t = 2$. Then, one of the main results of the macro approach is that when agents' productivities change stochastically over time, then $\tau'_S(\theta) > 0$ is optimal.

The early versions of this result limited to particular settings are Diamond and Mirrlees (1978) and Rogerson (1985). Golosov, Kocherlakota, and Tsyvinski (2003) provide a proof for a general class of dynamic economies with heterogeneous privately observable skills. They show that this result holds for any stochastic process for skills as long as there is some uncertainty about future idiosyncratic shocks.

To see the origins of this result, consider the following. Assume that preferences are additively separable, that is, $U_c(c(\theta), y(\theta)/\theta) = U_c(c(\theta))$ for all θ. Then in a general class of dynamic economies, when skills are heterogeneous, privately observable, and there is uncertainty about future skills, efficiency dictates that the marginal cost of provision of insurance to each agent follows a martingale. With separable preferences, it can be shown that the marginal cost of insurance is equal to $1/U_c(c(\theta))$. This implies that optimal allocations must satisfy a so-called inverse Euler equation. This equation is a necessary condition for optimality that in the two-period environment of this section states that for any θ

$$\frac{1}{U_c\left(c_1^*(\theta)\right)} = \mathbb{E}\left\{ \left. \frac{\delta}{\beta U_c\left(c_2^*\left(\theta^2\right)\right)} \right| \theta \right\},$$

where $\{c_t^*\}_{t=1,2}$ denote an optimal consumption allocation as before.

Since by Jensen's inequality $\mathbb{E}[\frac{1}{x}] > 1/\mathbb{E}[x]$ whenever $\text{Var}(x) > 0$, it follows from the inverse Euler equation that

$$\delta U_c\left(c_1^*(\theta)\right) < \beta \mathbb{E}\left\{ U_c\left(c_2^*(\theta^2)\right) \Big| \theta \right\},$$

which in turn implies that a positive marginal savings distortion, $\tau_S'(\theta) > 0$, is optimal.

If, however, there is no uncertainty about consumption in period $t = 2$, then the inverse Euler equation becomes

$$\frac{1}{U_c\left(c_1^*(\theta)\right)} = \frac{\delta}{\beta U_c\left(c_2^*(\theta^2)\right)},$$

or simply $\delta U_c(c_1^*(\theta)) = \beta U_c(c_2^*(\theta^2))$, which is a standard Euler equation describing the undistorted behavior of a consumer who chooses savings optimally. In other words, in a model with heterogeneous unobservable skills that do *not* stochastically change over time, it is optimal to have a zero capital tax (Werning 2002a, Golosov, Kocherlakota, and Tsyvinski 2003).

To develop intuition for the positive implicit tax on savings, consider the following perturbation of an optimal allocation. For a particular θ_1, decrease period $t = 1$ consumption by ε for θ_1 and increase period $t = 2$ consumption by ε/δ for (θ_1, θ_2) for all θ_2. Given that we started with an optimal allocation, this perturbation is incentive compatible and thus must not increase social welfare. That is, any positive effects of this perturbation must be cancelled by its negative effects. The first two effects of the perturbation are standard. First, the perturbation increases social welfare by increasing period $t = 2$ expected utility by $\beta \frac{\varepsilon}{\delta}\mathbb{E}\{U_c(c_2^*(\theta^2))|\theta_1\}$. Second, the perturbation decreases social welfare and the utility in period $t = 1$ by $\varepsilon U_c(c_1^*(\theta))$. However, there is also a third effect related to the provision of incentives given the information friction. The perturbation reduces incentives to work in period $t = 2$ by reducing covariance between the skills θ_2 and period $t = 2$ utility of consumption. This further reduces social welfare. Since the increase in the social welfare due to the first effect must be equal to the sum of the second and the third effects, we obtain that $\varepsilon U_c(c_1^*(\theta)) < \beta(\varepsilon/\delta)\mathbb{E}\{U_c(c_2^*(\theta^2))|\theta_1\}$. This implies that a positive marginal savings distortion, $\tau_S'(\theta) > 0$, is optimal. In other words, distorting the savings decisions at the optimum improves provision of dynamic incentives.

It is important to note, however, that the optimality of the positive intertemporal wedge—or implicit tax on savings—does not necessarily imply that optimally there needs to be a positive capital tax. Nor does it imply that wedges are necessarily equal to taxes. Rather, the main insight here is that any optimal dynamic tax policy or a social insurance system has to take into account agents' ability to save. Generally, though, taking into account agents' ability to save implies that savings should be discouraged.

This result is in sharp contrast with the Chamley–Judd result (Judd 1985, Chamley 1986) obtained in representative agent macroeconomic Ramsey settings. The Chamley–Judd result states that in the long-run capital should go untaxed.[15]

2.3 Quantitative Insights

In step with theoretical advances, several studies have carried out quantitative analyses of the optimal size of wedges, levels and shapes of taxes that implement the optimum, and welfare gains from improving tax policy. When it comes to computationally solving for a constrained dynamic optimum, one major roadblock is the size of the problem. On the face of it, the number of incentive constraints seems to be the culprit because it increases exponentially as the number of periods goes up or the number of types increases. However, the deeper underlying reason for the large size of these problems is history dependence. That is, the dependence of allocations on all—in the general case—of the previous realizations of shocks. Thus, any restriction that curtails history dependence makes quantitative explorations easier.[16]

One extreme is to assume i.i.d. shocks, that is, $F(\theta'|\theta) = F(\theta')$, as, for example, Albanesi and Sleet (2006) do. A way to exploit the assumption of i.i.d. shocks is to formulate the problem recursively with a one-dimensional state variable that can be interpreted as promised utility from that period on. The ability to formulate the planner's dynamic problem recursively with low-dimensional state variables is a significant computational advantage. Albanesi and Sleet (2006) assume i.i.d. shocks to skills and follow Atkeson and Lucas (1992) to rewrite the problem recursively. For their quantitative examination, Albanesi and Sleet (2006) choose utility function with income effects that is additively separable between consumption and work effort. They compute an implementation of their constrained optimum and examine the levels and shapes of the optimal capital and labor taxes. They find that optimal taxes are generally nonlinear in labor earnings and accumulated wealth and labor earnings taxes are generally lower than what Diamond (1998) and Saez (2001) find using the micro approach.

To help build intuition and further illustrate the case of i.i.d. shocks to skills, in Golosov, Troshkin, and Tsyvinski (2010b), we start by performing numerical simulations for the optimal labor and savings wedges in an illustrative two-period example. The example is based on empirical micro data and realistic parameter values. The analysis there naturally extends the quantitative analysis of the static

---

15. The extension of this analysis to environments with no steady state is provided in Judd (1999).

16. For specific details of the computational approaches taken in the literature, we refer the reader to the discussed papers. Broadly, the approaches can be separated into (i) solving first-order conditions and (ii) direct optimizations. With (i), one simplifies the first-order conditions analytically and numerically solves large systems of (usually differential, but sometimes also integral) equations. With (ii), the planner's problem is treated as a large nonlinear constrained optimization problem and direct optimization algorithms are used (usually interior-point or sequential linear/quadratic programming methods). In both approaches, dynamics is usually handled via value or policy function iteration versions of numerical dynamic programming with continuous states. Importantly, persistence leads one to rely on the first-order approach (to the incentive constraints) to reduce the dimensionality of the state. The validity of the first-order approach is verified *ex post*.

model in Section 1 as well as in Diamond (1998) and Saez (2001). Our optimal labor distortions are U-shaped in both periods.

In Golosov, Troshkin, and Tsyvinski (2010b), we use similar data to the ones used in the literature discussed in Section 1. For simplicity, we assume exponential preferences and a utilitarian planner in the numerical simulations. Note that exponential preferences imply no income effects just as the preferences discussed in Section 1. Therefore, one can compute the implied skills for a dynamic case from the individual static consumption-labor margins as well as one can in the static model. The quantitative results in Golosov, Troshkin, and Tsyvinski show that the marginal labor distortions in period $t = 2$ of the illustrative dynamic two-period example with i.i.d. shocks coincide with those of the static economy. The pattern of optimal marginal labor distortions is similar to the results in Diamond (1998) and Saez (2001) for static Mirrlees economies—they exhibit a U-shaped pattern for lower incomes, increase after that, and tend to a relatively high limit for high income individuals. We also observe a U-shaped pattern of labor distortion in period $t = 1$, although it is less pronounced. An important difference with the static case is that the level of distortions is substantially lower in period $t = 1$ for all income groups and especially for high-income individuals. The intuition for this result is that the dynamic provision of incentives enables the planner to lower distortions in period $t = 1$. Finally, we also find that the savings wedge increases for all income levels and is numerically significant.

Moving to the other side of the spectrum from i.i.d. shocks, another extreme example that restricts history dependence in a different way and facilitates quantitative explorations is the problem of providing disability insurance efficiently.[17] To make our discussion more concrete, consider a two-period example of this dynamic social insurance problem. In period $t = 1$, all agents are able to work. Any able worker can become disabled with some probability in period $t = 2$ (later in life), that is, with positive probability $\theta_2 = 0$ given any θ_1. It is relatively easy for a worker to falsely claim disability. For instance, a worker can pretend to be suffering from back pain, which is difficult to verify. We are interested then in designing an optimal disability insurance system. Such a system would provide adequate transfers to the truly disabled workers, i.e., the ones with $\theta_2 = 0$, while discouraging fake disability applications from those with $\theta_2 > 0$. The decision of a worker to claim disability is necessarily dynamic: a claim in period $t = 2$ is reflected in the worker's choices in period $t = 1$. For example, an able worker facing a given transfer scheme can increase or decrease his savings in period $t = 1$. This savings choice will necessarily increase or decrease his willingness to falsely claim disability benefits in period $t = 2$. In a T-period setting of this problem, Golosov and Tsyvinski (2006) assume permanent disability shocks (i.e., a disabled worker cannot later become able again). They compute the optimal allocation and show that the welfare gains from improving disability insurance system might be large.

---

17. For more on these types of problems see Diamond and Mirrlees (1978) and Golosov and Tsyvinski (2006).

Relative to the two dynamic settings above, environments with some degree of skill shock persistence are markedly less explored quantitatively. This is hardly surprising since persistent shocks pose more challenging computational problems. Dynamic settings with persistent shocks are important examples of environments where history dependence in optimal allocations plays a key role. Empirical studies suggest that there is significant degree of persistence in the idiosyncratic shocks to labor productivity, implying the importance of persistent skill shocks in studying dynamic optimal taxation (see, e.g., Storesletten, Telmer, and Yaron 2004).

The case of a particular form of persistent shocks in a two-period model is considered by Golosov, Tsyvinski, and Werning (2006). They numerically simulate optimal policies when idiosyncratic shocks follow a stochastic process where each agent in period $t = 2$ with equal probability can either stay as productive as he was in period $t = 1$ or receive a shock that makes the agent less productive.

An important step toward quantitatively studying dynamic settings with persistent shocks is made in Kapicka (2010). He suggests a first-order approach to simplify the recursive formulation of the planning problem when shocks are persistent. This leads to a substantial reduction of the state space of the dynamic program and curtails the computational challenges of history dependence. In numerical simulations, Kapicka finds that the optimal marginal distortions differ significantly between the i.i.d. and persistent shock cases.

In Golosov, Troshkin, and Tsyvinski (2010b), we address the case of persistent shocks analytically by combining the elements of micro and macro approaches. The insights we develop there—that are also the basis for the discussion in Section 3— can help interpret our quantitative results. In Golosov, Troshkin, and Tsyvinski (2010b), we quantitatively study multiperiod life-cycle environments with persistent shocks based on empirical micro data and realistic parameter values. To keep the discussion here intuitive, consider a two-period example of such environment. If we consider the two-period example, we find that the pattern of labor distortions in period $t = 1$ in the economy with persistent shocks is similar to the static case in Section 1 and the i.i.d. case above. However, in contrast with the i.i.d. case, different first-period income groups face very different labor distortions in period $t = 2$. The labor distortions in period $t = 2$ of agents who in period $t = 1$ had high incomes are much higher than their labor distortions in period $t = 1$ (and higher than in the i.i.d. case). The labor distortions for agents who in period $t = 1$ had lower incomes do not change significantly from their earlier distortions (and are lower than in the i.i.d. case). Another observation we make in Golosov, Troshkin, and Tsyvinski (2010b) is that the labor distortions no longer follow a U-shaped pattern found in the i.i.d. and static simulations. Finally, we find that the savings wedge increases for all income levels and the overall pattern remains similar to the i.i.d. case with the only difference that the level of the savings distortion is lower. In Golosov, Troshkin, and Tsyvinski (2009), we further quantitatively explore the question of general empirically relevant persistent shock processes at length.

An important contribution of Farhi and Werning (2010) analyzes a different way of characterizing the first-order conditions of the optimal dynamic taxation model.

They provide numerical simulations and also use continuous time setting to derive additional insights. The analyses of Farhi and Werning (2010) and Golosov, Troshkin, and Tsyvinski (2010b) are complementary in an important respect. While Golosov, Troshkin, and Tsyvinski focus on a comprehensive study of cross-sectional properties of optimal wedges and on deriving elasticity based formulas extending Diamond (1998) and Saez (2001), Farhi and Werning (2010) focus on the comprehensive study of the intertemporal properties of allocations and wedges.

The numerical simulations and quantitative insights of the macro literature we discuss above are all looking for an optimal policy and possibly the results of a reform towards it. Another quantitative route to take is to consider partial reforms. Rather than finding the full optimum, a variety of papers using the macro approach consider partial changes in the taxes or insurance systems that can improve upon the current system.

One example of this approach is Farhi and Werning (2009). They consider the welfare gains from partial reforms that introduce optimal savings distortions into the actual tax code but leave the labor allocations unchanged. They compute the efficiency gains from introducing optimal savings distortions by comparing the welfare outcome to an equilibrium where agents' saving decisions are not distorted. The study also investigates how these welfare gains depend on a limited set of features of the economy and finds that general equilibrium effects play an important role.

Another route for a partial tax reform in a dynamic setting is to compute the optimal tax schedule in a model where the tax function is restricted to a specific functional form. By allowing the parameters of the tax function to change optimally, one can allow for a wide range of shapes of tax systems, including progressive taxation, nondiscriminatory lump-sum taxation, and various exemptions. This is the route taken in Conesa and Krueger (2006), Conesa, Kitao, and Krueger (2009), and Golosov, Troshkin, and Tsyvinski (2009). Weinzierl (2008) performs a partial reform study to determine welfare gains and optimal taxes in a calibrated model with age-dependent taxes. He uses individual wage data from the PSID and simulates a dynamic model that generates robust implications. He finds that age dependence lowers marginal taxes on average and especially on high-income young workers. Also, age dependence lowers average taxes on all young workers relative to older workers when private saving and borrowing are restricted. Weinzierl (2008) finds that, despite its simplicity, age dependence generates large welfare gains both in absolute size and relative to fully optimal policy.

Finally, an important quantitative insight is an estimate of the fraction of labor productivity that is private information. A recent study by Ales and Maziero (2007) estimates the fraction of labor productivity that is private information in a life-cycle version of a dynamic Mirrlees economy with publicly and privately observable shocks to individual labor productivity. They find that for the model and data to be consistent, a large fraction of shocks to labor productivities must be private information.[18]

18. See also Farhi and Werning (2007) for the analysis of estate taxation in an intergenerational dynastic model with dynamic private information that shows that estate taxes should be progressive. Hosseini, Jones,

2.4 Implementations

The characterization of optimal allocations and optimal distortions is only one part of the macro approach to dynamic optimal taxation. Ultimately, we are interested in learning what kinds of taxes implement optimal allocations. Unlike in the static settings of the micro literature on optimal taxation, in dynamic Mirrlees taxation models, optimal wedges do not necessarily coincide with marginal taxes implementing optimal allocations (see, e.g., Grochulski and Kocherlakota 2007, Albanesi and Sleet 2006, Golosov and Tsyvinski 2006, Kocherlakota 2005). Thus, the study of the implementations of optimal programs is an important part of the macro approach to taxation. Next, we discuss some recent implementation results in this literature. All of the implementations below have two key features: (i) taxes or transfers have to be conditioned on the amount of savings that agent accumulates, and (ii) there is some degree of history dependence.

First, consider the disability insurance example described earlier. Consider a system of disability transfers that provides a disabled worker with, say, $1,000. An able worker contemplates in period $t = 1$ whether to work or to claim disability in period $t = 2$. If he fakes disability, he will receive $1,000 in period $t = 2$ with probability one. If he does not fake and claims disability only if he is truly disabled, he will receive $1,000 if he is disabled (with some probability less than one) and a higher amount from work if he is able. Given this transfer system, the worker who chooses to falsely claim disability will then have higher savings because he expects to receive $1,000 for sure and not work. A disability insurance scheme that introduces a tax on savings (e.g., by asset testing, i.e., paying benefits only to those with low enough assets) will then discourage fake disability claims and thus move closer to the optimum potentially implementing it.

Golosov and Tsyvinski (2006) show that the optimal disability insurance system can be implemented as a competitive equilibrium with taxes where the optimal allocation is implemented due to the presence of an asset-tested disability insurance system. That is, the system makes a disability benefit payment only if an agent has assets below a specified maximum. Given this type of disability insurance system in place, if an agent considers claiming disability insurance falsely, he will not find doing so beneficial unless he adjusts his savings accordingly. And if the agent increases his savings in the preparation for a false claim of disability insurance, then he will not be able to receive the disability benefits. Golosov and Tsyvinski (2006) quantitatively evaluate the implementation of the optimum with an asset-tested disability insurance system and show that the welfare gains from asset testing are large.

Kocherlakota (2005) studies a dynamic setting with no restrictions on the stochastic evolution of skills over time. He constructs a tax system that implements the optimal allocation in the following way. The taxes are constrained to be linear in an agent's

and Shourideh (2009) in a model of endogenous fertility with private information on productivity show that estate taxes are positive and there are positive taxes on the family size. Finally, Shourideh (2010) takes Mirrleesian approach to study the taxation of capital accumulation and finds that entrepreneurial and nonentrepreneurial capital income should be taxed differently.

accumulated savings but can be arbitrarily nonlinear in his current and past labor incomes. In this implementation, savings taxes in a given period must optimally depend on the individual's labor earnings in that period and the previous ones. However, in any period, the expectation of an agent's savings tax rate in the following period is zero. One possible implementation in these general dynamic environments is one in which capital taxes are regressive.

Several studies consider examples of special cases where implementations are particularly intuitive or practical. One example is Albanesi and Sleet (2006) who show that in a special case of i.i.d. processes for idiosyncratic skill shocks, a nonlinear tax on savings and labor income implements the optimum. They also find that the optimal taxes are generally nonseparable in savings and labor income and relate the shape of marginal savings and labor income tax functions to the properties of individual preferences. Another example is Grochulski and Kocherlakota (2007) who study optimal dynamic policy in environments with habit persistence. They show that in some models with habit formation implementations of the optimal allocation resemble a social security system in which taxes on savings are linear and all optimal taxes and transfers are history dependent only at retirement. An implementation in the context of a model of entrepreneurship is studied in Albanesi (2006). That paper explores optimal taxes under a variety of market structures.

An important recent paper by Werning (2009) characterizes a system of nonlinear taxes on savings that implement any incentive compatible allocation. He restricts the savings tax to be independent of the current state. The tax schedule is differentiable under quite general conditions and its derivative, the marginal tax, coincides with the wedge in the agent's intertemporal Euler equation. Although he allows for nonlinear schedules, a linear tax often suffices. Finally, he shows how the savings tax can be made independent of the history of shocks.

Finally, in Golosov, Troshkin, and Tsyvinski (2010a), we provide a novel implementation of the optimal allocations in general dynamic environments. We refer to this implementation as a consolidated income accounts (CIA) tax system. In a given period in a general dynamic Mirrlees environment, labor income tax depends on that period's labor income and on the balance on the CIA. The savings tax depends only on the amount of that period's savings. The CIA balance is then updated as a function of labor income and its previous balance. We also show that a CIA system takes a particularly simple form if the utility is exponential and the shocks are i.i.d. The tax system consists of a nonlinear tax on capital income,[19] nonlinear labor income tax, and a CIA account. In each period, a taxpayer can deduct the balance of the account from the total labor income tax bill. Thus, while all agents with the same labor income are facing the same marginal tax rate, the total tax bill is smaller for the agents with a higher CIA account. Similarly, updating the CIA balance follows a simple rule. In each period, a change in the CIA balance is determined solely by the individual's labor income in that period.

19. The capital tax implementation is based on Werning (2009).

2.5 Private and Public Insurance

Since the macro literature addresses efficient provision of social insurance, it is important to be explicit about how private insurance markets operate. Policy prescriptions implied by the insights of the dynamic macro approach therefore depend on the availability of private insurance. Above, as it is done in much of the macro literature, we look at one extreme case of no private insurance to provide a useful benchmark. Now, we return to the question of private insurance markets and discuss some of the recent results.

An important aspect of designing optimal dynamic taxation and insurance system is to allow for the possibility of private insurance. In the environments where the only friction is unobservability of types, one can show that the optimal allocation can be decentralized without any need of government intervention. Prescott and Townsend (1984) and Atkeson and Lucas (1992) showed that allocations provided by competitive markets are constrained efficient. The intuition is that the private insurers can offer the same allocations as the planner would. This result does not mean, however, that the wedges present in the optimal allocation disappear in the decentralized competitive equilibrium allocation. Rather, the private insurers offer contracts that have the same wedges (e.g., the same savings wedge) as the social planner would. The only effect of government insurance provision in this environment is complete crowding out of private insurance leaving allocations and welfare unchanged.

The case of observable consumption may have limited empirical relevance in modern economies. It is difficult to imagine that individual firms can preclude individual agents from engaging in credit market transactions or transactions with other firms. In a modern economy, it is very rare that a firm can condition its compensation on how much an agent saves in the bank, how much disability insurance he holds, etc. Golosov and Tsyvinski (2007) study an environment in which consumption is unobservable to the planner as agents can trade unobservably on private markets. An example of this in the context of the disability insurance—that we consider throughout this section—is a setting where workers are able to borrow or lend with a market determined interest rate and such transactions are not observable by the insurance agency. Golosov and Tsyvinski show that private insurance is not efficient and has to be supplemented with public intervention.

Albanesi (2006) considers several market structures that allow multiple assets and private insurance contracts. She explores optimal entrepreneurial capital taxation under these arrangements and proposes implementations of the optimal allocations in a model of entrepreneurship with a variety of market structures.

Ales and Maziero (2009) is a recent study that considers a dynamic Mirrleesian economy in which workers can sign insurance contracts with multiple firms. That is, they extend the dynamic Mirrlees environment to add another friction in the form of nonexclusive contracts on the labor side. Their model endogenously divides the population into agents who are not monitored and have access to nonexclusive contracts and agents who have access to exclusive contracts. Ales and Maziero use the U.S. household level data and find that high school graduates satisfy the optimality

conditions implied by the nonexclusive contracts, while college graduates behave like the group with access to exclusive contracts.

2.6 Challenges of the Macro Approach

The literature on dynamic Mirrlees problems has delivered many important insights into a broad variety of social insurance and taxation issues in dynamic contexts. Nevertheless, many intriguing and challenging questions still lie ahead for the macro approach.

First, it is generally difficult to solve for optimal allocations in dynamic environments, either analytically or computationally. This is especially true in the case of persistent shocks. Second, as a result of optimal allocations in a given period depending on full history of reports, the optimal taxes that are suggested by dynamic environments may depend in a complex way on all of the past choices of individuals. Finally, the key challenge for macro approach is to produce concrete policy recommendations. For example, a recent survey of policy relevance of optimal taxation models by Mankiw, Weinzierl, and Yagan (2009) states, "Most of the recommendations of dynamic optimal tax theory are recent and complex" and that "The theory of optimal taxation has yet to deliver clear guidance on a general system of . . . taxation . . . . Instead, it has supplied more limited recommendations." One reason for that is that the analysis of the dynamic taxation models is often primarily theoretical and uses the language more familiar to a macroeconomist than to a public finance economist. Another reason is that optimal tax systems derived in these models are often difficult to interpret and connect to the empirical data of interest in policy applications. While the macro approach has not yet delivered easily implementable policy insights, Banks and Diamond (2008) argue in their Mirrlees Review chapter on direct taxation for the importance of the Mirrleesian—dynamic and static—models as a guide for policy.

In the next section, we argue that progress can be made by bridging the gap between the macro approach and the more standard to public finance literature micro approach, much of which is set in a static framework. The focus of the next section is on the recent results of an analysis that combines the elements of the micro approach with the dynamics of the macro literature.

3. MERGING THE MICRO AND MACRO APPROACHES

In Golosov, Troshkin, and Tsyvinski (2010b), we suggest a way to merge the elements of micro and macro approaches. This provides a methodology to derive simple formulas that facilitate the interpretation of the forces behind the optimal taxation results in dynamic settings. The formulas are easy to connect to empirically observable data. Obtained by applying the combined analysis, these formulas summarize the first-order conditions for the optimal dynamic labor and savings distortions. As such, the analysis in Golosov, Troshkin, and Tsyvinski extends the micro approach

results of Diamond (1998) and Saez (2001) to dynamic settings of the macro literature discussed in Section 2.

The formulas for the dynamic labor distortions derived in Golosov, Troshkin, and Tsyvinski (2010b) are conceptually similar to those derived in the static models of the micro literature that we discuss in Section 1. As in the static case, the shape of the income distribution, the redistributionary objectives of the government, and labor elasticity play key roles in the determination of optimal labor distortions in dynamic settings. However, the dynamics of the macro approach also adds significant differences to the analysis of optimal distortions. We perform computations for the optimal taxes in empirically realistic calibrated cases and find the results consistent with the insights offered by the formulas.

We first consider the case of i.i.d. shocks. There are two key insights from this part of the analysis for the nature of labor distortions early in the life of an agent. First, the dynamic nature of the incentives represents itself as an additional term in the formula for the optimal distortions. This term effectively alters the welfare weights assigned to agents by the social planner. Second, this reweighing allows the use of dynamic incentives to lower marginal taxes for a fraction of sufficiently skilled agents early in their lives. We also derive a formula representing the savings distortion. The key economic insight of the analysis here is that a high savings distortion should be applied to the high-skilled agents as a way to lower their labor distortion. The intuition is that the effort of the highly skilled agents is highly valuable in production and thus deterring their deviations via a savings tax is particularly important.

In the case of persistent shocks, we are able to show that there are two key insights in addition to the analysis of the static and the i.i.d. cases. The first difference is that the optimal labor distortion formulas now depend on conditional rather than on the unconditional distributions of skills. The second insight is that persistence adds an additional force to the optimal tax problem. When shocks are persistent, an agent misrepresenting his skill early in life has better information than the planner about the true realization of his shocks in the future. This consideration represents itself as a modification of welfare weights in the social welfare function that are assigned to different types of agents. As a result, the planner redistributes away from the types that are more likely to occur after an agent deviated earlier in life.[20]

Finally, we note that in every period of a dynamic environment the planner needs both to redistribute between initial higher and initial lower types and to provide insurance against subsequent shocks. This suggests an implementation via an integrated tax and social insurance system. That is, it is optimal that labor distortions arise from the sum of all tax and social insurance programs rather than from income tax code alone. This also implies that various social insurance programs ought to be integrated. In this regard, in Golosov, Troshkin, and Tsyvinski (2010a), we show that an integrated tax system like a CIA tax system discussed in Section 2 can keep track

---

20. Battaglini and Coate (2008) is one example in which the authors solve for the labor taxes in a dynamic Mirrlees economy. They show that when the utility of consumption is linear, labor taxes of all agents asymptotically converge to zero.

of past labor earning in a summarized fashion and condition transfers and taxes on the summary accounts.

4. OPTIMAL TAXATION AND POLITICAL ECONOMY

One additional issue that is important and closely related to the discussion above is that of the effects of the political economy considerations on optimal taxation. The papers considered above assume that the policymaker is a fictitious benevolent social planner with full commitment. But in reality, the social programs and taxation are determined by politicians. Acemoglu, Golosov, and Tsyvinski (2008b, 2009a) study the optimal Mirrlees taxation problem in a dynamic economy but, in contrast to the approach above, the policy is decided in a classical electoral accountability model of political economy (see also Acemoglu, Golosov, and Tsyvinski 2009b). Politicians are self-interested (fully or partially) and cannot commit to promises. They can misuse the resources and the information they collect to generate rents. An important technical result of the analysis is that a version of revelation principle works despite the commitment problems and the different interests of the government. Using this tool, they show that if the government is as patient as the agents, then the best sustainable mechanism leads in the long run to allocation where the aggregate distortions arising from political economy disappear. In contrast, when the government is less patient than the citizens, there are positive aggregate political economy distortions even asymptotically. Acemoglu, Golosov, and Tsyvinski (2008a) also use this framework to compare centralized mechanisms operated by self-interested rulers to anonymous markets. A related environment is that of the debt policy in dynamic settings with linear taxes and self-interested politicians in Yared (2010).

Farhi and Werning (2008) is a recent study of efficient nonlinear taxation of labor and capital in a dynamic Mirrleesian model that incorporates political economy constraints in which policies are the outcome of democratic elections, and there is no commitment. Their main result is that the marginal tax on capital income is progressive, in the sense that richer agents face higher marginal tax rates. Sleet and Yeltekin (2008) embed a version of the dynamic macro environment considered in Section 2 into a family of game settings that model political credibility considerations. The authors study political game settings with repeated probabilistic voting over mechanisms. That is, voters repeatedly choose among rival political parties and their respective versions of resource allocations. Politically credible allocations are then the allocations that are immune to this revision process via elections. Sleet and Yeltekin (2008) show that optimal politically credible allocations solve a perturbed planning problem with social discount factors greater than the private one and welfare weights that tend to converge to 1. The properties of credible equilibria in dynamic settings with the lack of societal commitment are examined in another recent paper by Sleet and Yeltekin (2009). The authors isolate the forces that promote and retard capital accumulation in these settings, derive the pattern of intertemporal wedges as well as provide an implementation result.

5. CONCLUSION

This paper provides a review of the micro and macro approaches to optimal taxation. We argue that merging these two approaches can provide new insights into the nature of optimal taxation and bring the literature closer to policy implementations.

LITERATURE CITED

Abraham, Árpád, and Nicola Pavoni. (2008) "Optimal Income Taxation and Hidden Borrowing and Lending: The First-Order Approach in Two Periods." Carlo Alberto Notebooks 102, Collegio Carlo Alberto.

Acemoglu, Daron, Mikhail Golosov, and Aleh Tsyvinski. (2008a) "Markets versus Governments." *Journal of Monetary Economics,* 55, 159–89.

Acemoglu, Daron, Mikhail Golosov, and Aleh Tsyvinski. (2008b) "Political Economy of Mechanisms." *Econometrica,* 76, 619–42.

Acemoglu, Daron, Mikhail Golosov, and Aleh Tsyvinski. (2009a) "Dynamic Mirrlees Taxation under Political Economy Constraints." *Review of Economic Studies,* 1–48.

Acemoglu, Daron, Mikhail Golosov, and Aleh Tsyvinski. (2009b) "Political Economy of Ramsey Taxation." NBER Working Paper No. 15302.

Albanesi, Stefania. (2006) "Optimal Taxation of Entrepreneurial Capital with Private Information." NBER Working Paper No. 12419.

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

Ales, Laurence, and Pricila Maziero. (2007) "Accounting for Private Information." Federal Reserve Bank of Minneapolis Working Paper 663.

Ales, Laurence, and Pricila Maziero. (2009) "Non-Exclusive Dynamic Contracts, Competition, and the Limits of Insurance." Working Paper.

Allen, Franklin (1985) "Repeated Principal-Agent Relationships with Lending and Borrowing." *Economic Letters,* 17, 27–31.

Atkeson, Andrew, and Robrert E. Lucas, Jr. (1992) "On Efficient Distribution with Private Information." *Review of Economic Studies,* 59, 427–53.

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

Banks, James, and Peter Diamond. (2008) "The Base for Direct Taxation." *In Dimensions of Tax Design: The Mirrlees Review*, edited by J. Mirrlees, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles and J. Poterba. Oxford, UK: Oxford University Press.

Battaglini, Marco, and Stephen Coate. (2008) "Pareto Efficient Income Taxation with Stochastic Abilities." *Journal of Public Economics,* 92, 844–68.

Bisin, Alberto, and Adriano Rampini. (2006) "Markets as Beneficial Constraints on the Government." *Journal of Public Economics,* 90, 601–29.

Chamley, Christophe. (1986) "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives." *Econometrica,* 54, 607–22.

Cole, Harold, and Narayana R. Kocherlakota. (2001) "Efficient Allocations with Hidden Income and Hidden Storage." *Review of Economic Studies,* 68, 523–42.

Conesa, Juan Carlos, Sagiri Kitao, and Dirk Krueger. (2009) "Taxing Capital? Not a Bad Idea After All!" *American Economic Review,* 99, 25–48.

Conesa, Juan Carlos, and D. Krueger. (2006) "On the Optimal Progressivity of the Income Tax Code." *Journal of Monetary Economics,* 53, 1425–50.

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

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

Farhi, Emmanuel, and Iván Werning. (2007) "Inequality and Social Discounting." *Journal of Political Economy,* 115, 365–402.

Farhi, Emmanuel, and Iván Werning. (2008) "The Political Economy of Non-Linear Capital Taxation." Mimeo, MIT.

Farhi, Emmanuel, and Iván Werning. (2009) "Capital Taxation: Quantitative Explorations of the Inverse Euler Equation." Working Paper.

Farhi, Emmanuel, and Iván Werning. (2010) "Insurance and Taxation over the Life Cycle." Working Paper.

Fudenberg, Drew, and Jean Tirole. (1991) *Game Theory*. Cambridge, MA: MIT Press.

Golosov, Mikhail, Narayana R. Kocherlakota, and Aleh Tsyvinski. (2003) "Optimal Indirect and Capital Taxation." *Review of Economic Studies,* 70, 569–87.

Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski. (2009) "A Quantitative Exploration in the Theory of Dynamic Optimal Taxation." Mimeo, University of Minnesota.

Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski. (2010a) "Consolidated Income Accounts." Working Paper.

Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski. (2010b) "Optimal Dynamic Taxes." Working Paper.

Golosov, Mikhail, Maxim Troshkin, Aleh Tsyvinski, and Maxim Weinzierl. (2010) "Preference Heterogeneity and Optimal Capital Taxation." NBER Working Paper 16619.

Golosov, Mikhail, and Aleh Tsyvinski. (2006) "Designing Optimal Disability Insurance: A Case for Asset Testing." *Journal of Political Economy,* 114, 257–79.

Golosov, Mikhail, and Aleh Tsyvinski. (2007) "Optimal Taxation with Endogenous Insurance Markets." *Quarterly Journal of Economics,* 122, 487–534.

Golosov, Mikhail, Aleh Tsyvinski, and Iván Werning. (2006) "New Dynamic Public Finance: A User's Guide." *NBER Macroeconomics Annual,* 21, 317–63.

Grochulski, Borys, and Narayana R. Kocherlakota. (2007) "Nonseparable Preferences and Optimal Social Security Systems." NBER Working Paper No. 13362.

Hosseini, Roozbeh, Larry E. Jones, and Ali Shourideh. (2009) "Risk Sharing, Inequality and Fertility." NBER Working Paper.

Hurwicz, Leonid. (1960) "Optimality and Informational Efficiency in Resource Allocation Processes." In *Mathematical Methods in the Social Sciences,* edited by K.J. Arrow, S. Karlin, and P. Suppes. Stanford, CA: Stanford University Press.

Hurwicz, Leonid. (1972) "On Informationally Decentralized Systems." In *Decision and Organization*, edited by C.B. McGuire and R. Radner. Amsterdam: North-Holland.

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

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

Kapicka, Marek. (2010) "Efficient Allocations in Dynamic Private Information Economies with Persistent Shocks: A First Order Approach." Mimeo, University of California Santa Barbara.

Kocherlakota, Narayana R. (2005) "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation." *Econometrica,* 73, 1587–621.

Kocherlakota, Narayana R. (2010) *The New Dynamic Public Finance*. Princeton, NJ: Princeton University Press.

Mankiw, N. Gregory, Matthew Weinzierl, and Danny Yagan. (2009) "Optimal Taxation in Theory and Practice." NBER Working Paper No. 15071.

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green. (1995) *Microeconomic Theory*. New York: Oxford University Press.

Meade, James. (1978) *The Structure and Reform of Direct Taxation*. London: Institute for Fiscal Studies.

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

Mirrlees, James A. (1976) "Optimal Tax Theory: A Synthesis." *Journal of Public Economics,* 6, 327–58.

Mirrlees, James A. (1986) "The Theory of Optimal Taxation." *Handbook of Mathematical Economics,* 3, 1197–249.

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

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

Sadka, Efraim. (1976) "On Income Distribution, Incentive Effects and Optimal Income Taxation." *Review of Economic Studies,* 43, 261–7.

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

Salanie, Bernard. (2003) *The Economics of Taxation*. Cambridge, MA: MIT press.

Seade, Jesus K. (1977) "On the Shape of Optimal Tax Schedules." *Journal of Public Economics,* 7, 203–35.

Shimer, Robert, and Iván Werning. (2008) "Liquidity and Insurance for the Unemployed." *American Economic Review,* 98, 1922–42.

Shourideh, Ali. (2010) "Optimal Taxation of Capital Income: A Mirrleesian Approach to Capital Accumulation." Mimeo, University of Minnesota.

Sleet, Christopher, and Sevin Yeltekin. (2008) "Politically Credible Social Insurance." *Journal of Monetary Economics,* 55, 129–51.

Sleet, Christopher, and Sevin Yeltekin. (2009) "Allocation and Taxation in Uncommitted Societies." Tepper School of Business Paper 460.

Stern, N. (1976) "On the Specification of Models of Optimum Income Taxation." *Journal of Public Economics,* 6, 123–62.

Stiglitz, Joseph E. (1987) "Pareto Efficient and Optimal Taxation and the New New Welfare Economics." *Handbook of Public Economics,* 2, 991–1042.

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

Tuomala, Matti. (1990) *Optimal Income Tax and Redistribution*. New York: Oxford University Press.

Weinzierl, Matthew. (2008) "The Surprising Power of Age-Dependent Taxes." Mimeo, Harvard University.

Werning, Iván. (2002a) "Optimal Dynamic Taxation and Social Insurance." Ph.D. Dissertation, University of Chicago.

Werning, Iván. (2002b) "Optimal Unemployment Insurance with Unobservable Savings." Mimeo, MIT.

Werning, Iván. (2009) "Nonlinear Capital Taxation." Working Paper.

Yared, Pierre. (2010) "Politicians, Taxes, and Debt." *Review of Economic Studies,* 77, 806–40.