

# Policy Implications of Dynamic Public Finance

Mikhail Golosov<sup>1</sup> and Aleh Tsyvinski<sup>2</sup>

<sup>1</sup>Department of Economics, Princeton University, Princeton, New Jersey 08544;  
email: golosov@princeton.edu

<sup>2</sup>Department of Economics, Yale University, New Haven, Connecticut 05520-8268;  
email: a.tsyvinski@yale.edu

Annu. Rev. Econ. 2015. 7:147–71

First published online as a Review in Advance on  
March 18, 2015

The *Annual Review of Economics* is online at  
economics.annualreviews.org

This article's doi:  
10.1146/annurev-economics-080614-115538

Copyright © 2015 by Annual Reviews.  
All rights reserved

JEL codes: H21, I38, E62

## Keywords

social insurance, redistribution, optimal taxation, new dynamic  
public finance

## Abstract

The dynamic public finance literature underwent significant changes over the past decade. This research agenda has now reached a stage at which it is able to analyze the design of social insurance programs and optimal taxation in rich environments that can be closely matched to microeconomic data. We provide an overview of the recent advances in this literature, discuss the key trade-offs, and explain how the prescriptions for the optimal policy depend on the specific parameters that can be estimated in the data. We also describe the relationship between the dynamic mechanism design approach to optimal taxation and the approach that considers sophisticated tax functions chosen within parametrically restricted classes.

## 1. INTRODUCTION

The theory of taxation and social insurance underwent a significant transformation over the past decade. Advances in theoretical methods and computational techniques dramatically increased the realism of the models used for the analysis. It is now possible to study optimal policy in environments with rich heterogeneity and realistic uncertainty that closely match microeconomic data. The goal of this article is to provide an overview of recent advances in this literature and discuss its policy implications.

We focus first on the area that is most researched and in which the link with the empirical literature is especially tight: the theory of redistribution and insurance against idiosyncratic labor income shocks. Section 3 starts the discussion with a static model that builds on the work of Mirrlees (1971). Although simple, this model highlights important economic trade-offs in designing social insurance programs that are also present in dynamic environments. This analysis also allows us to illustrate the insights of Diamond (1998) and Saez (2001) into how empirical estimates of the labor supply elasticity and the hazard rates of the income distribution can be used to obtain sharp qualitative and quantitative predictions about optimal taxes and transfers.

We then extend our discussion to a canonical life-cycle model with idiosyncratic shocks. This environment has long been a workhorse of the empirical labor literature (see Storesletten et al. 2004, Heathcote et al. 2010, and Blundell et al. 2012 for some examples of recent work). In Section 4, we explain how the trade-offs emphasized in the static environment interact with additional dynamic considerations and describe the formulas that link the optimal labor and capital distortions with structural parameters of the model that can be estimated empirically. Critically, reliable estimates of some important parameters have only recently become available. For example, the theory emphasizes that higher moments of the distribution of shocks, such as kurtosis, are important parameters that determine the form of the optimal insurance programs. Reliable estimation of these moments became possible only recently, when high-quality administrative data became available for economic research. We conclude Section 4 by showing simulations for the optimal insurance programs using these estimates.

The optimal redistribution and insurance discussed in Section 4, constrained only by informational frictions, can be implemented using a complex system of taxes and transfers that uses history dependence and sophisticated joint taxation of several sources of income. This system provides a useful benchmark that we discuss in Section 5. In Section 6, we discuss alternative approaches based on the analysis of simpler tax systems. We first describe how research on optimal policy using parametrically defined tax functions achieved significant progress in analyzing realistic, empirically rich models with significant heterogeneity and a variety of frictions. We then show how the analysis of dynamic tax reforms can be used to evaluate various elements of the optimal tax systems and decompose the welfare and revenue gains coming from age dependence, savings effects, and the joint taxation of capital and income.

In Section 7, we describe several extensions and applications of the baseline framework. First, we discuss the models in which shocks affect return to capital rather than labor income. This specification is particularly relevant to the analysis of entrepreneurial decisions that account for a significant portion of the upper tail of the wealth distribution. The main result of the analysis is that there is a need to differentially tax and carefully consider various forms of investment—whether they are primarily financial or invested in the productive business capital. Second, we discuss an active literature on endogenizing the skills through human capital; significant recent progress has been achieved in analyzing both these models and their policy implications for optimal taxation and a system of education subsidies. Third, we discuss recent work on estate taxation, which emphasizes the need to carefully consider the social welfare criteria and heterogeneity in altruism that drives bequests. Finally, we discuss the implications of dynamic optimal taxation to the design of pension systems.

## 2. EFFECTIVE MARGINAL TAX RATES IN THEORY AND IN THE DATA

Most of tax theory focuses on the characterization of marginal labor and capital distortions. Their analogs in the data are the effective marginal tax rates introduced by tax and transfer systems. The effective marginal tax rates measured in the data consist of two elements. The first element comprises various taxes on income, levied on either individuals or employers, as well as taxes on consumption. The second element comes from welfare transfers and various social insurance programs. The phaseout rules of such programs, which make welfare benefits available to individuals only if their income does not exceed a certain threshold, have an effect that is economically equivalent to a positive marginal tax.<sup>1</sup>

In the United States, effective tax rates vary by the level and source of income, age, family status, type of residence, etc. Different states have different eligibility rules for welfare programs, and as a result, there is a substantial heterogeneity in effective tax rates (Maag et al. 2012). We illustrate general patterns using the state of Colorado as an example. **Figure 1** shows income eligibility for transfer programs, and **Figure 2** shows the effective tax rate for a single parent with two children. The effective marginal tax rates are the highest for moderately low and relatively high annual earnings. The former are driven by the phaseout of welfare programs, the latter by progressivity in the income tax schedule. Individuals with very low earnings, below the poverty line, often face negative effective taxes owing to such welfare programs as the EITC. Apart from the rates for the very poor, the effective marginal taxes are approximately U-shaped. These patterns of distortions exist in many other states (Maag et al. 2012) and also are present in federal tax programs (Congr. Budg. Off. 2005), although there is a very substantial heterogeneity in both the shape of the effective tax rate schedules and the size of the effective tax rates, depending on the state of residence, family status, and the number of children.

## 3. REDISTRIBUTION AND TAXATION IN A STATIC MODEL

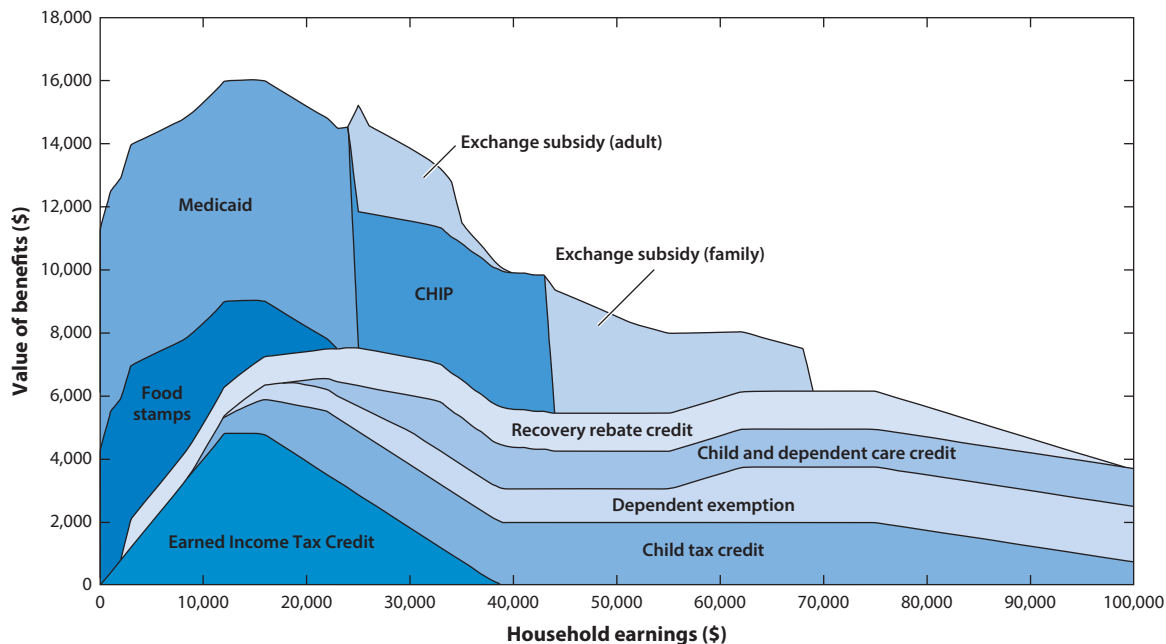
We start our discussion of optimal taxation by considering a static model first developed by Mirrlees (1971). We assume that individuals have a standard utility function  $U(c, l)$  over consumption  $c$  and effort  $l$ , and that they are heterogeneous with respect to their skills  $\theta$ . Given the same amount of effort, the higher-skilled individuals can produce more labor income,  $y = \theta l$ . Welfare is evaluated using a weighted sum of utilities of all households, with  $\alpha(\theta)$  denoting the Pareto weight.

There are two ways to think about optimal taxation in this model. One approach is to postulate an income tax schedule  $T(y)$  and find the tax function that is budget feasible and maximizes social welfare.  $T(y)$  may be positive or negative, and it captures the net effect of all tax and transfer programs. An alternative approach is to consider first the best allocation that the government can achieve if it has limited information about individuals' abilities and then to back out the implied system of taxes and transfers that can achieve this optimum. The two approaches are equivalent when the government has no information about  $\theta$ , the assumption that we maintain through most of this review.<sup>2</sup>

We focus on the qualitative and quantitative properties of the optimal  $T(y)$  and its generalizations for dynamic economies. We discuss which parameters determine the shape of the optimal

<sup>1</sup>Some of the transfers, such as the Earned Income Tax Credit (EITC) in the United States, provide a subsidy to low-income earners. Such programs may simultaneously have negative effective tax rates (for the income levels at which they are phased in) and positive effective tax rates (for the income levels at which they are phased out).

<sup>2</sup>Several papers study the implications for taxation if the government can use additional characteristics that are correlated with an individual's earning ability (see, e.g., Mankiw & Weinzierl 2010 for a discussion of height-dependent taxation and Alesina et al. 2011 for a discussion of gender-dependent taxation). One way to think about such taxes is as a joint function  $T(y, \text{characteristics})$ , where characteristics are the non-income-related information used by the government.



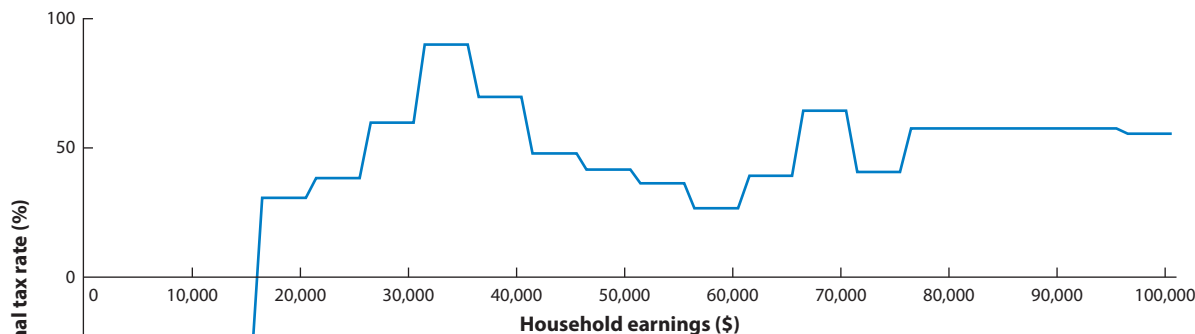
**Figure 1**

Universally available tax and transfer benefits for a single parent with two children in Colorado in 2008. Figure reproduced from Maag et al. (2012) with permission from the National Tax Association. Abbreviation: CHIP, Children’s Health Insurance Program.

tax schedule and use empirical estimates of those parameters to quantify the size of the optimal effective marginal rates. The properties of optimal taxes can generally be traced to two main trade-offs that can be illustrated by the following argument.<sup>3</sup> Let  $H(y)$  be the distribution of labor earnings in an economy in which all individuals maximize their utility facing a tax schedule  $T(y)$ . Let  $h(y)$  be the probability density function of  $H(y)$ . Consider a perturbation of this schedule that increases the marginal tax rate on all incomes in an interval  $[\bar{y}, \bar{y} + \Delta]$  by a small amount, keeping the marginal taxes at all other income levels fixed. The additional revenues from this perturbation are equally distributed to everyone through a uniform shift in the tax schedule. This perturbation is shown in Figure 3.

This perturbation has three effects. First, all the individuals with income in the interval  $[\bar{y}, \bar{y} + \Delta]$  face an increase in the marginal tax rate on an extra dollar they earn. The reduction in tax revenues collected from this group is determined by the compensated (price) elasticity of labor supply and the total income that this group earned before the perturbation, which is approximately equal to  $\bar{y}h(\bar{y})$  for small  $\Delta$ . The second effect comes from an increase in average taxes for all income levels above  $\bar{y} + \Delta$ . The amount of additional revenues collected from this group is determined by the size of this group,  $1 - H(\bar{y})$ , and the income elasticity of labor supply. The net revenue gain from these two effects may be positive or negative. Because the net revenues are then returned back to households uniformly, there is the third effect from the redistribution of resources between those households that earn more than  $\bar{y}$  and those that earn less. This redistribution effect is evaluated

<sup>3</sup>Piketty (1997) and Saez (2001) developed this line of argument.



**Figure 2**

Effective marginal tax rate for a single parent with two children in Colorado in 2008. Figure reproduced from Maag et al. (2012) with permission from the National Tax Association.

using the Pareto-adjusted marginal utility of consumption,  $\alpha U_c$ . In the optimal tax system, the sum of the three effects should be zero.

This argument points to the elasticity of labor supply, the income effect, and the hazard rate  $(1 - H(y))/yh(y)$  as the key parameters that determine the optimal marginal taxes in this model. All else being equal, the effective marginal tax rate is high when the compensated elasticity is small, and the income effect and the hazard rate are large. Redistributive objectives of the government, captured by the Pareto weights, obviously play a role as well. Some statements can be made, however, for a broad class of redistributive objectives. Most common welfare criteria assign weakly lower Pareto weights to richer individuals. Because richer individuals also have lower marginal utility of consumption, such criteria imply that, holding the elasticities and the hazard ratio fixed, they should face higher marginal tax rates.<sup>4</sup>

We can now discuss the policy implications of this model. First, as long as Pareto weights  $\alpha$  are nonincreasing in the skill or income, negative effective marginal tax rates are not optimal. A negative marginal tax on income  $y$  provides a transfer to agents with incomes above  $y$  at the

<sup>4</sup>Recent research has explored the normative assumptions often used in optimal tax theory. Mankiw & Weinzierl (2010) point out a puzzling implication of standard utilitarian optimal tax theory—namely, that it recommends much greater use of tagging, the taxation of personal characteristics statistically correlated with income such as height, than is found in existing policy. Weinzierl (2014) shows that the tagging puzzle, among others, can be resolved by using an objective function for tax policy that reflects a mixed normative criterion. In particular, he argues that the classical principle of equal sacrifice likely plays a role in prevailing judgments of taxation and helps explain features of existing policy at odds with conventional optimal results.

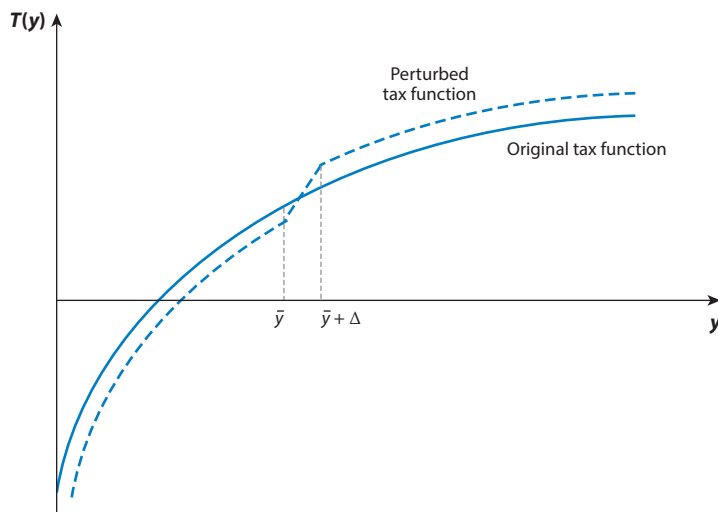


Figure 3

Perturbation of an income tax scheme in a static model.

expense of agents with incomes below  $y$ . Because the agents with incomes below  $y$  are poorer, their marginal utility of consumption is lower, and hence such redistribution is suboptimal. Therefore, programs similar to the EITC are not desirable in this environment.<sup>5</sup>

The second broad implication for the optimal marginal tax rate can be derived using the properties of the hazard rate  $(1 - H(y))/y h(y)$  observed in the data. Using US tax return data, Saez (2001) documents that this hazard rate is U-shaped: It is high at low income levels, decreases and reaches its minimum at approximately \$80,000 annual earnings (in 1992 dollars), and then increases again and stabilizes around a value of 0.5. For a broad class of welfare criteria, this implies that the optimal marginal taxes should also be U-shaped (see Diamond 1998 for a formal analysis).

The previous argument overlooks the following subtle distinction. We applied our perturbation to the income distribution  $H(y)$  generated by the optimal tax schedule  $T(y)$ , whereas in the data we observe the income distribution generated by the existing, potentially suboptimal tax code. The two are linked through the distribution of  $\theta$ , which can be inferred from the data using one of the two methods. One possibility is to postulate a functional form for preferences, for example,

$$U(c, l) = \frac{1}{1 - \sigma} c^{1 - \sigma} - \frac{1}{1 + 1/\varepsilon} l^{1 + 1/\varepsilon}, \quad (1)$$

and use the available empirical estimates of  $\sigma$  and  $\varepsilon$  together with the observed tax code in the data to impute the distribution of  $\theta$ . Another possibility is to observe that the labor income of type  $\theta$  and their skill are linked by a formula

<sup>5</sup>The negative marginal effective tax rates may be optimal if the government uses high Pareto weights for the agents in the middle of the income distribution at the expense of the agents in the tails. Such situations naturally arise in voting equilibria. For example, Brett & Weymark (2014) consider a model in which taxes are chosen using majoritarian voting and show that the marginal tax rates are typically negative for those with low income.

$$\frac{\dot{y}(\theta)}{y(\theta)} = \frac{1 + \zeta^u(\theta)}{\theta} + \dot{y}(\theta) \frac{T''(y(\theta))}{1 - T'(y(\theta))} \zeta^c(\theta),$$

where  $\zeta^c$  and  $\zeta^u$  are compensated and uncompensated elasticities of labor supply, respectively, and  $T$  is any (optimal or not) tax schedule. Given empirical estimates of  $\zeta^c$  and  $\zeta^u$  and the effective tax rates observed in the data, one can use this formula to back out the distribution of  $\theta$ .

The two approaches each have their strengths and weaknesses. The first approach is more indirect and requires a structural estimation of preference parameters. The second one is more intuitive but generally requires estimation of relevant elasticities at different points of the income distribution, as they typically vary with income and are endogenous to the tax code; this approach is also harder to generalize to dynamic and stochastic environments. For this reason, throughout this review we use the first approach. In the static setting, they produce similar results.

The empirical observations about the patterns of skill distribution can be used to obtain tight quantitative predictions about optimal taxation of the very rich. Standard preferences imply that the marginal utility of consumption of the very rich approaches zero. As long as the Pareto weights do not increase in skill, this implies that it is efficient to set taxes to maximize the tax revenues collected from these agents. If skills are drawn from a distribution with a Pareto tail and preferences are given by Equation 1, the revenue-maximizing tax rate is given by

$$\frac{1}{\frac{1}{1 + \varepsilon\sigma} + a \frac{\varepsilon}{1 + \varepsilon}}, \quad (2)$$

where  $a$  is the Pareto coefficient for skill distribution.<sup>6</sup>

To get a sense of the magnitude of the optimal top marginal tax rate, assume that  $\sigma = 1$ . In this case, the Pareto tail of the skill distribution coincides with the Pareto tail of the empirical income distribution (see Golosov et al. 2013c for details), which Saez (2001) estimates to be approximately 2. Then one can use Equation 2 and empirical estimates of  $\varepsilon$  to compute the top marginal tax rate. The elasticity parameter  $\varepsilon$  has long been recognized as crucial for estimating labor supply responses. There is a substantial controversy about the value of this parameter depending on whether a researcher uses micro or macro data sets. The literature using the micro data typically finds small values of  $\varepsilon$ , with the labor supply elasticity being 0.3 or less.<sup>7</sup> These estimates imply a top marginal tax rate of 80% or more. Conversely, the macro literature often finds significantly higher elasticities, with  $\varepsilon$  frequently between 1 and 2 and sometimes as high as 4.<sup>8</sup> For these estimates, the top marginal tax rate is between 55% and 65%.

Summarizing our discussion, conventional welfare criteria imply that negative labor distortions are not optimal. For a large class of welfare criteria, the observed income distribution and

<sup>6</sup>Saez (2001) obtains this expression in terms of compensated and uncompensated elasticities of labor supply and the Pareto coefficient of the income distribution. The formula we use is identical to his when the elasticities are derived in terms of structural parameters of preferences (Equation 1). Saez also provides calculations for the optimal top marginal tax rates for a range of parameters.

<sup>7</sup>The three commonly used elasticities are Frisch, compensated (Hicksian), and uncompensated (Marshallian). In our preferences,  $\varepsilon$  measures the Frisch elasticity. When marginal taxes are constant and nonlabor income is a negligible fraction of total earnings, the compensated and uncompensated elasticities are  $\zeta^c = \varepsilon/(1 + \varepsilon\sigma)$  and  $\zeta^u = \varepsilon(1 - \sigma)/(1 + \varepsilon\sigma)$ , respectively.

<sup>8</sup>Readers are referred to Keane & Rogerson (2012) for an overview of the two literatures and a discussion of the way to reconcile their estimates. Note that the ideal parameter for the static model is the elasticity of the lifetime labor supply. The discussion here also abstracts from joint family labor supply decisions, which is an important margin at which incentives to work operate within families (Blundell et al. 2012), and from the distinction between intensive and extensive margins (see Saez 2002, Laroque 2005, and Werquin 2014 for an analysis of the optimal taxation with adjustments along the extensive margin).

commonly estimated preference parameters imply that the optimal labor distortions are U-shaped, with top marginal taxes for the rich often exceeding 60%.

#### 4. DYNAMIC ECONOMY

The static model discussed in the previous section illustrates some important trade-offs in optimal taxation and provides a useful benchmark, but it also has its limitations. By implicitly attributing all cross-sectional income heterogeneity to permanent skill differences, it overstates the underlying inequality in the economy. Both the deterministic life-cycle skill changes and transitory shocks make the cross-sectional distribution appear more dispersed than the true underlying heterogeneity. The static model also abstracts from capital taxation, and it does not allow one to consider the design of retirement or social insurance systems that provides insurance against specific idiosyncratic shocks such as disability or a job loss.

In this section, we discuss redistribution and the design of social insurance in a canonical life-cycle model. We assume that individuals have finite lives and are subject to idiosyncratic shocks. Their initial skills are drawn from a distribution  $F_0(\theta)$ , and then  $\theta_t$  follows a Markov process. The drift, persistence, volatility, and higher moments of this shock process may depend on age. In this section, we assume that this process is exogenous; in Section 7.2, we discuss extensions of this framework to human capital accumulation. To simplify the discussion, we focus on isoelastic preferences (Equation 1) and assume that the government can freely borrow and lend with a riskless interest rate  $R$ .

Throughout this section, we assume that skills  $\theta_t$  are unobservable to the government and characterize properties of the incentive-compatible allocations that maximize Pareto-weighted lifetime utilities of agents. Under some technical invertibility assumptions (see Kocherlakota 2005), this is equivalent to choosing the optimal taxes of the form  $T_t(y_t, k_t; y^{t-1})$ , where  $T_t$  is the tax schedule in period  $t$  on capital  $k_t$  and labor income  $y_t$ . The tax may also be a function of the history of past incomes, summarized by a vector  $y^{t-1} = (y_0, \dots, y_{t-1})$ .

It is useful to focus on the socially optimal incentive-compatible allocations and the implied tax system that decentralizes them for several reasons. They provide a natural upper bound on what can be achieved with social insurance, at least as long as the policies do not use non-income-related information to infer individual skills. Once the properties of the fully optimal tax  $T_t$  are known, they can be used as guidance to design simpler insurance systems. Finally, the analytical tools to solve such models—the recursive contract methods developed by Green (1987), Spear & Srivastava (1987), Thomas & Worrall (1990), Atkeson & Lucas (1992), Fernandes & Phelan (2000), and Kapicka (2013)—are readily available, and they allow one to obtain clean analytical insights into the main economic forces that determine the optimal taxes and allocations.

To characterize properties of the optimal allocations, we focus on distortions, or wedges, in the consumption-labor and Euler equations. Formally, the labor distortion in period  $t$ ,  $\tau_t^y$ , for an individual with a history of skill shocks  $\theta^t$  is defined as

$$\tau_t^y(\theta^t) = 1 - \frac{-U_l(\theta^t)}{\theta_t U_c(\theta^t)},$$

and the capital or savings distortion  $\tau_t^s$  is defined as

$$\tau_t^s(\theta^t) = 1 - \frac{1}{\beta(1+R)} \frac{U_c(\theta^t)}{\mathbb{E}_t U_c(\theta^{t+1})},$$

where  $\beta$  is the discount rate. The discussion of the results in this section is based on the work of Farhi & Werning (2013b) and Golosov et al. (2013c). For simplicity, we drop explicit references to past histories  $\theta^t$  in our notation [e.g.,  $\tau_t^y$  denotes  $\tau_t^y(\theta^t)$  for a given  $\theta^t$ ].



In dynamic economies, a planner pursues two goals while choosing incentive-compatible allocations in a given period. First, the planner needs to provide insurance against new shocks that an individual experiences in that period. This problem is essentially identical to the static model with utilitarian Pareto weights discussed in the previous section. All the arguments from the static model continue to apply with the caveat that the hazard rate of the distribution of period- $t$  shocks rather than the cross-sectional distribution determines the size of the optimal labor distortion. Second, the planner needs to ensure that period- $t$  allocations provide incentives to reveal information in previous periods, which is needed both for provision of insurance against idiosyncratic shocks in earlier periods and for redistribution. For commonly used shock processes, this effect is proportional to the persistence of the stochastic processes and the previous period's labor distortion.

One implication of this discussion is that negative labor distortions are typically suboptimal in dynamic settings. Such distortions are not desirable for the provision of insurance against period- $t$  shocks and, as long as Pareto weights on the low-income individuals are not too low, for redistribution.<sup>9</sup>

Farhi & Werning (2013b) characterize the law of motion of the labor distortions when shocks are log-normally distributed with persistence  $\rho$ . They show that distortion dynamics satisfies

$$\mathbb{E}_t \left[ \frac{\tau_{t+1}^y}{1 - \tau_{t+1}^y} \frac{1}{U_{c,t+1}} \right] = \left( 1 + \frac{1}{\varepsilon} \right) \text{cov}_t \left( \ln \theta_{t+1}, \frac{1}{U_{c,t+1}} \right) + \rho \frac{\tau_t^y}{1 - \tau_t^y} \frac{1}{U_{c,t}}.$$

The first term on the right-hand side of this equation captures the intratemporal insurance motive. Because skills and consumption are positively correlated, it shows a force for higher expected labor distortions in future periods. As more shocks are being realized over time, the need to provide insurance increases, requiring higher labor distortions. The second term on the right-hand side captures the intertemporal incentive motive. Because the empirical literature generally finds that the uninsurable component of the idiosyncratic shocks is highly persistent (Storesletten et al. 2004, Guvenen et al. 2014a), this law of motion implies that on average labor distortions should be higher later in life.

The two objectives that the planner faces when choosing period- $t$  allocations have different impacts on labor distortions for high and low shocks. Golosov et al. (2013c) show that the optimal labor distortions for unexpectedly high shocks are mainly determined by the need to provide intratemporal insurance. Therefore, many arguments from the static model carry over to dynamic environments largely unchanged for such shocks. In particular, if the distribution of shocks is fat tailed, the top labor income distortion is still given by the expression in Equation 2, with the only difference being that  $a$  is the tail parameter of the shock process rather than the tail of the cross-sectional distribution. Even when the tails of the stochastic process are thin (e.g., if shocks are log-normal), marginal labor distortions are approximately constant for a large range of high shocks.<sup>10</sup> Conversely, the need to provide incentives for information revelation in previous periods determines

<sup>9</sup>It is possible to construct examples of stochastic processes for which negative labor distortions sometimes are optimal. The discussion here is based on log-normal shocks (Farhi & Werning 2013b) and a mixture of log normals chosen to match higher-moment shocks observed in the data (Golosov et al. 2013c).

<sup>10</sup>Technically, when shocks are drawn from a log-normal distribution or a mixture of log-normal distortions, the top marginal distortion converges to zero, but the rate of convergence is very slow, of the order of  $\ln \theta$ . When plotted against income, such distortions appear approximately flat. A higher kurtosis of the stochastic process implies higher labor distortions for large positive shocks.

the optimal labor distortions for unexpectedly low shocks. For low realizations of  $\theta_t$ , they approximately satisfy

$$\frac{\tau_t^y}{1 - \tau_t^y} \approx \rho \frac{\tau_{t-1}^y}{1 - \tau_{t-1}^y} \left( \frac{c_t}{c_{t-1}} \right)^{-\sigma}.$$

These expressions allow us to identify the key parameters that determine the size of the optimal labor distortion. The distortions for high shocks depend mainly on the elasticity of labor supply, the income effect, and the tail properties of the hazard rate, all of which can be estimated in the data. The distortions do not depend on Pareto weights, age, or past shock history. Alternatively, the optimal labor distortions for low shocks are determined by the persistence of shocks, the redistributory objectives, and the history of past shocks, summarized by the term  $\tau_{t-1}^y / (1 - \tau_{t-1}^y)$ .

The discussion so far has focused on labor distortions; we now turn to capital distortions. Golosov et al. (2003) demonstrate that as long as preferences are separable between consumption and labor, the optimal consumption allocations satisfy the following expression, the inverse Euler equation:

$$\frac{1}{U_{c,t}} = \frac{1}{\beta(1+R)} \mathbb{E}_t \frac{1}{U_{c,t+1}}.$$

This equation implies that a positive savings distortion is optimal as long as there is some unrealized uncertainty in the next period. Farhi & Werning (2012) further quantify the size of the savings distortion for realistic shock processes. First, observe that if the utility of consumption is logarithmic and consumption is log-normally distributed with variance  $\sigma_\varepsilon^2$ , the optimal savings distortion is given by

$$\tau_t^s = 1 - \exp(-\sigma_\varepsilon^2), \quad (3)$$

so that  $\tau_t^s \approx \sigma_\varepsilon^2$  when  $\sigma_\varepsilon$  is small. The empirical estimates of the permanent component of consumption volatility in the data are fairly low. For example, Blundell et al. (2008) use the Panel Study of Income Dynamics data set to estimate it at approximately 0.01, Deaton & Paxson (1994) find it to be 0.0069 using a different methodology, and Heathcote et al. (2014a) estimate it at 0.0056. As these are the estimates of the consumption volatility under the current, likely suboptimal tax system, better insurance should reduce this number further. Therefore, the optimal savings distortions should not be very high.<sup>11</sup> Farhi & Werning (2012) show that they are further reduced in general equilibrium and conclude that savings distortions play a modest role in the provision of insurance.

The quantitative properties of the optimal labor and capital distortions depend crucially on the stochastic process for the idiosyncratic shocks, in particular on the higher moments of that process. Until recently, the empirical labor literature mainly used household surveys. It is difficult to reliably estimate higher moments with the small samples and top coding prevalent in such surveys, and for this reason, the literature often estimates only the persistence and volatility of idiosyncratic shocks. Newly available for economic research, high-quality administrative data made the estimation of higher moments possible. Guvenen et al. (2014a,b) use a sample of 10% of the working-age males in the United States and find that the stochastic process for labor earnings has kurtosis and skewness significantly larger than that implied by the normal distribution. Golosov et al. (2013c) use the US tax code and the empirical moments for earnings reported by Guvenen et al. (2014a) to estimate the stochastic process for skills  $\theta$ . **Figure 4** shows the quantitative properties of the optimal labor and capital distortions, which are reported for isoelastic preferences (Equation 1) with  $\sigma = 1$  and  $\varepsilon = 0.5$ .

<sup>11</sup>However, this is a distortion in the gross interest rate  $1 + R$ .

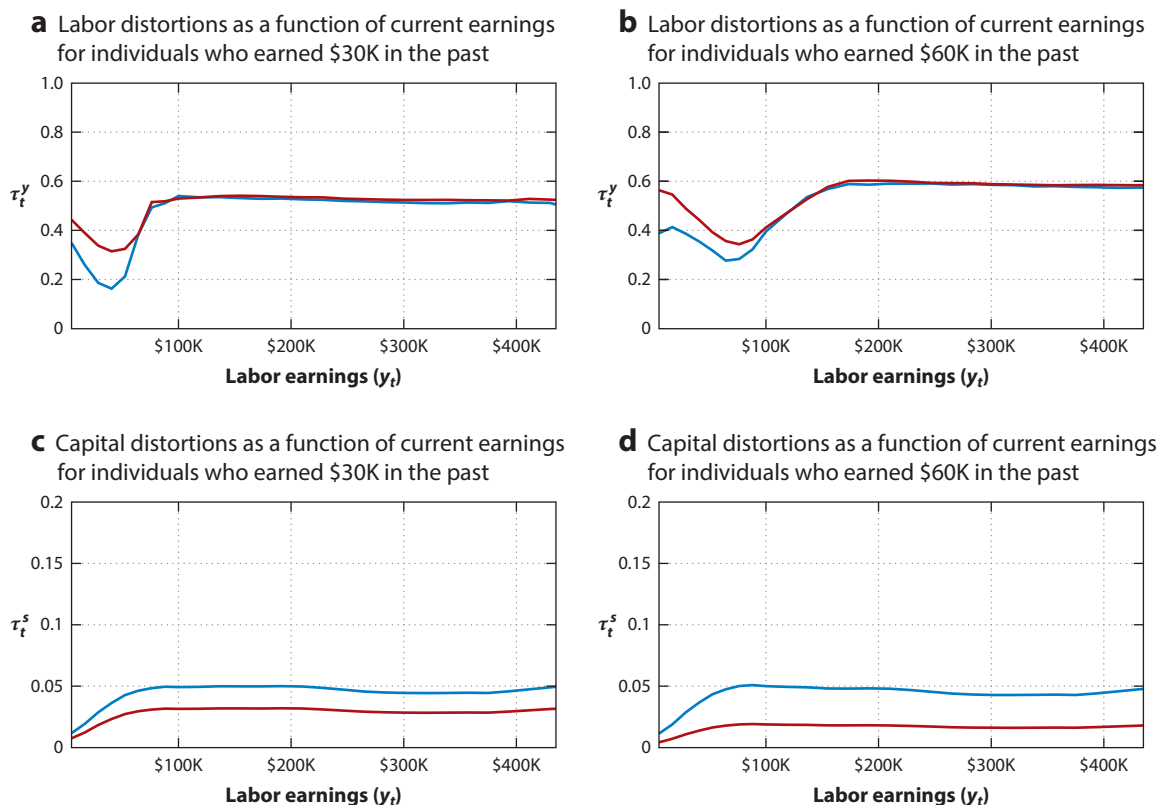


Figure 4

(a,b) Labor and (c,d) savings distortions in a life-cycle model. Red lines correspond to higher  $t$  (i.e., to distortions later in life). Figure reproduced from Golosov et al. (2013c).

Figure 4a,b shows the optimal labor distortion as a function of income in period  $t$  of an individual who earned \$30,000 and \$60,000 in all previous periods, respectively. The optimal labor distortions are U-shaped, with the smallest distortions around the previous labor income. Individuals who had lower income in the past have more regressive labor distortions for earnings in period  $t$  and more progressive labor distortions for high earnings. Recall from Figure 2 that high effective marginal tax rates in the data for low earnings are associated with the phaseout of social insurance programs. Therefore, this pattern of labor distortions is consistent with a slower phaseout of social insurance programs for individuals with a history of higher earnings. Labor distortions in the left tail increase with age and with the redistributory objectives of the government (not shown on the figure) and are approximately constant in the right tail. The U-shaped pattern of labor distortions is optimal because of the high kurtosis of idiosyncratic shocks, implied by the empirical findings of Guvenen et al. (2014a); with log-normally distributed shocks, these distortions would have been mildly regressive and approximately flat.<sup>12</sup>

<sup>12</sup>Recall from our discussion of the static economy that labor distortions are higher at income levels at which the hazard ratio  $(1 - H(y))/y b(y)$  is high and vice versa. When the shocks, and hence income, follow a highly leptokurtic stochastic process, the hazard ratio is large in the tails and small around the mean, which explains the U-shaped pattern.

Savings distortions are shown in **Figure 4c,d**. Savings distortions are progressive and decreasing with age. Savings distortions are smaller at low incomes as those individuals receive more insurance. Savings distortions decrease with age because the same shock implies a smaller loss of the present value of earning later in life and hence affects consumption less. As seen in Equation 3, a lower volatility of consumption is associated with lower labor distortions.<sup>13</sup>

## 5. DECENTRALIZATION

Once the optimal allocations are known, it is possible to find a tax system  $T_t(y_t, k_t; y^{t-1})$  that decentralizes them in a competitive equilibrium with taxes. The optimal allocations are generally unique, whereas there are many different tax functions  $T_t(y_t, k_t; y^{t-1})$  that decentralize them. Several papers investigated whether it is possible to decentralize the optimal allocations using relatively simple tax systems.

One of the first discussions of decentralization in dynamic environments is by Golosov & Tsyvinski (2006), who study optimal insurance against disability shocks. They show that the optimal disability insurance can be implemented with asset-tested disability benefits, which make insurance payments only if the value of an individual's assets is below a certain threshold.<sup>14</sup> The decentralization in Golosov & Tsyvinski (2006) is simple, but their analysis is restricted to only one type of shock (permanent disability). Albanesi & Sleet (2006) consider a distribution of shocks with arbitrary support under an assumption that shocks are independently and identically distributed. They show that it is not necessary to keep track of a past history of earnings in such settings and implement the optimum with a simple joint tax on labor and capital income. Kocherlakota (2005) discusses decentralizations with more general shocks.

A common feature of these decentralizations is that the marginal tax rate on capital income is decreasing in current-period labor income. The intuition for the efficiency of such tax is as follows. It is desirable to provide the largest transfers to the poorest individuals. A policy that makes individuals with higher wealth ineligible for some transfers appears as a high effective capital tax rate. The asset-tested disability benefits of Golosov & Tsyvinski (2006) are an extreme example of such phaseouts. Further discussion of decentralizations of dynamic economies with idiosyncratic shocks can be found in Werning (2007) and Fukushima (2010).<sup>15</sup>

An important issue is how much welfare society would lose with simpler tax systems compared to the full optimum. Mirrlees was the first to investigate this question in his original study of the optimal nonlinear taxation. He computed the optimal nonlinear tax in an economy with logarithmic preferences and log-normal shocks and found that such tax can be closely approximated by a proportional labor income tax together with a uniform lump-sum subsidy. In their study of a dynamic economy with log-normally distributed idiosyncratic shocks, Farhi & Werning (2013b) reach a similar conclusion. They find that the optimal linear capital and labor taxes result in a welfare loss of 0.3% of lifetime consumption. This number can be further cut in half if taxes depend on age. Golosov et al. (2013c) find welfare losses from linear taxes to be larger in a model with initial heterogeneity, leptokurtic idiosyncratic shocks, and a concave social welfare function.

<sup>13</sup>Farhi & Werning (2013b) were the first to show that the optimal savings distortions decrease with age in these settings.

<sup>14</sup>Asset tests are a part of eligibility criteria for some welfare programs in the United States.

<sup>15</sup>Another strand of the literature studies implementation and the role of endogenous private insurance in these environments (see, e.g., Golosov & Tsyvinski 2007, Farhi et al. 2009, Ales & Maziero 2009).

## 6. EVALUATING THE ELEMENTS OF THE OPTIMAL TAXES

The mechanism design approach to optimal taxation described above starts with an informational friction, solves for the constrained optimal allocation to characterize the wedges, and then finds the tax functions that implement the optimum. The tax functions are a priori unrestricted, and the informational frictions determine the properties of the optimal taxes. An alternative approach is to start with parametric restrictions on the tax function rather than with the informational friction. These restrictions still capture some important elements of a tax code, such as progressivity or age dependency.

There are several advantages to this method. First, optimizing over a set of parameters of the restricted tax functions is often significantly easier than finding the full informationally constrained optimum. Therefore, the quantitative solutions can be found in very rich models—with multiple dimensions of heterogeneity, sophisticated processes for ongoing stochastic shocks, and incomplete markets. Second, the tax functions typically used in this approach are already based on the features of the existing tax code. Moreover, it is easy to add realistic features that may not necessarily be currently present in the tax code but may be desirable. Third, as the analysis already starts with the tax functions, there is no distinction between taxes and wedges. The main drawback of this approach is its sensitivity to the exact specification of the tax function. It is possible that slightly varying the form of the tax functions may lead to rather different conclusions on the nature of the optimum. The mechanism design approach and the restricted tax functions approach are complementary. One promising direction to further unify the two approaches is to incorporate the key prescriptions of the mechanism design approach into the restricted tax function. Below we describe several prominent papers that exemplify the optimal taxation with restricted tax functions approach.

Conesa et al. (2009) study a model in which household productivity has three components: a deterministic age-dependent component that provides an explicit life-cycle structure, a type-dependent fixed effect that captures heterogeneity in the innate ability to generate income, and a persistent idiosyncratic shock. The markets are incomplete. The tax system over which the government optimizes is restricted as follows. The capital income tax is type independent and linear. The income tax function is assumed to be of the form proposed by Gouveia & Strauss (1994). There are two important parameters in this functional form. The first parameter determines the level of the average tax. The second parameter determines the progressivity of the tax code and thus allows one to consider progressive, flat, or regressive tax functions. Therefore, the government chooses three elements of the tax function: the level of the linear capital tax, the average income tax, and the degree of the income tax progressivity. Conesa et al. (2009) find that the optimal tax system is given by a substantial tax on capital income of 36% and a labor income tax that is flat at 23% with a deduction of \$7,200. Moreover, the optimal tax system yields welfare that is significantly higher (by 1.33% in consumption equivalents) than welfare under the current system. The authors provide a comprehensive quantitative analysis of the main determinants of the optimal tax. To begin, consider the capital tax. The main reason for the positivity and magnitude of the capital tax stems from the theoretical analysis of the life-cycle models of endogenous labor supply. Garriga (2001) and Erosa & Gervais (2002) show that in life-cycle models, when the elasticity of labor supply differs with age, and when the labor income tax is not dependent on age, a capital income tax mimics the optimal age dependency of the labor income tax. The key reason for the progressivity of the labor income tax is the government's motive to redistribute across households with different innate abilities. Introducing uninsurable labor income shocks during the lifetime of the agents further strengthens the insurance and redistribution motive.<sup>16</sup>

<sup>16</sup>An earlier paper (Conesa & Krueger 2006) also uses a similar restriction on the income tax function but does not distinguish between capital and labor income. It finds that the optimal tax is a flat tax of 17% with a deduction of \$9,400.

Several papers further extend this work. Kitao (2010) shows that incorporating labor-dependent capital taxes can approximate optimal age-dependent taxation of both labor and capital income and improve on the prescriptions of the high proportional capital tax. This result is related to the prescriptions of the mechanism design approach for the optimality of joint conditioning of capital and labor income taxes and on the age dependency of the optimal tax system. Gervais (2012) provides further theoretical justifications of why progressivity in the tax system may mimic the optimal age dependency of the optimal tax. Peterman (2013) considers an extension in which the elasticity of the labor supply is constant and the government is allowed to tax accidental bequests at a separate rate from ordinary capital income. He finds that these considerations imply a lower optimal tax on capital and lower welfare gains of the optimal restricted tax system.

Heathcote et al. (2014b) study a perpetual youth model in which households face idiosyncratic labor market shocks of two types, privately insurable and uninsurable shocks. There is also ex ante heterogeneity of two types: learning ability and disutility of work effort. The government policy is restricted to a nonlinear income tax system that provides social insurance and finances public goods. There are two parameters of the function that can be chosen: the progressivity of the system and the level of output devoted to public goods. The key result that the authors derive is the closed-form solution for the optimal welfare that allows them to clearly understand how various parameters affect the degree of optimal progressivity. They find that a utilitarian government would choose less progressivity than in the current US tax system, and the gains of switching to the optimum are on the order of 0.5% of aggregate consumption. The main role in the lower progressivity is played by the endogenous labor supply, endogenous skill investment, and the externality related to public goods.

Fukushima (2010) and Heathcote & Tsujima (2014) come closest to the aim of bridging the gap between the mechanism design approach and the approach using restricted instruments. Fukushima (2010) studies two lessons from the dynamic mechanism approach to taxation: (a) nonseparability in current labor and asset income with negative cross-partial derivatives and (b) the history dependence of optimal taxes. He replaces Conesa et al.'s (2009) optimal flat tax with an optimal nonlinear tax that is allowed to be arbitrarily age and history dependent and finds a welfare gain on the order of a 10% increase in consumption for every household. This gain mostly comes from higher per capita consumption and shorter per capita hours, as well as a shift of labor supply toward productive households, which increases aggregate productivity. This result indicates that the gains of considering more sophisticated tax functions may be large. Heathcote & Tsujima (2014) study an environment in which groups of individuals can insure the shocks among themselves in addition to available private insurance, and the planner can tax the income of the families. The mechanism design problem then becomes a static problem as the stochastic shocks are insured within the families. The authors solve for the constrained optimum and explore whether parametric tax functions can come close to achieving those allocations.

The discussion here illustrates the costs and benefits of the restricted tax functions approach. On the one hand, the limited number of parameters to optimize over allows researchers to consider very rich models of both heterogeneity and household choices and derive closed-form expressions, as well as quantitative decompositions of the sources of the welfare gains and the determinants of the optimal taxes. On the other hand, the set of tax functions considered is limited, which potentially leaves a large part of the welfare gains unexplored.

A different approach to analyzing the elements of the optimal taxes is to consider tax reforms. This approach lies between the fully optimal tax schedules and the models with the restricted tax functions. The primary goal of this approach is to find the effects of changing the tax system by incorporating various elements of the fully optimal system.

Weinzierl (2011) considers the idea that age-dependent taxation, a relatively simple partial reform, could capture a large share of the gains that a fully optimal, history-dependent policy would yield. If the shape of the income-earning ability distribution varies with age, for instance, if it is more compressed when workers are young, tax schedules that depend on age can achieve desired redistribution more efficiently than age-independent taxes can. In addition, if young workers face constraints on borrowing against future earnings, age-dependent taxes can relax those borrowing constraints by shifting resources from older to younger workers. Quantitatively, Weinzierl's calibrated simulations suggest that age dependence could yield welfare gains equivalent to 1% of aggregate output, capturing more than 60% of the gains from the fully optimal, history-dependent policy. Referencing this research and earlier, related work by Kremer (2002) in their chapter for *The Mirrlees Review*, Banks & Diamond (2010) cite age dependence as one of the most promising near-term reforms of tax policy.

Golosov et al. (2014) propose a general method to analyze tax reforms in dynamic settings. They study a dynamic model in which individuals' characteristics evolve deterministically over their lifetimes. Instead of solving for a constrained optimal problem and then backing out the implied optimal taxes  $T_t(y_t, k_t; y^{t-1})$  that decentralize the optimum, they develop a method to optimize with respect to the tax function  $T_t$  directly. This method builds on the perturbation ideas that Piketty (1997) and Saez (2001) apply to a static economy. The advantage of the optimization with respect to  $T_t$  is that it is possible to impose restrictions a priori on the type of taxes that are available to the government and find the optimum within that class. Another advantage of this approach is that it allows the authors to evaluate gains from local tax reforms, such as introducing some amount of history dependence or increasing progressivity in some part of the tax code. Those gains are expressed in terms of parameters that can be estimated in the data, such as the elasticities of labor and savings and multivariate hazard rates.

## 7. OTHER APPLICATIONS

Our discussion so far has focused on optimal taxation in models that abstract from occupational choice, human capital, capital income risk, bequests, and many other important margins along which people make savings and labor decisions. In this section, we discuss recent literature that considers the policy implications of such decisions.

### 7.1. Capital Income Risk and Taxation of Wealth

Most analyses discussed above are based on an environment in which the primary source of risk is to labor income. At the same time, a large literature on entrepreneurship emphasizes another source of risk, capital income risk, that is more pertinent to the issues of taxation of wealth. For example, Quadrini (2000) and Gentry & Hubbard (2004) show that of the top 5% of the wealthiest Americans, approximately 70% are business owners. Moreover, a significant proportion of the business owners' wealth is concentrated in their enterprises and hence is quite risky.

Here, we discuss two papers (Albanesi 2005 and Shourideh 2012) that focus on how entrepreneurial and capital income risks affect the prescriptions of the optimal taxation of wealth. Albanesi (2005) derives important results on the possibility of a negative intertemporal capital wedge for entrepreneurs and on the implementation of the optimal allocations under different market structures in an environment in which investment is observable. Shourideh (2012) derives results on the progressivity of taxation of various forms of investment in an environment in which investment is unobservable. Both papers highlight the importance of carefully considering the source of investment in deriving prescriptions for optimal taxation in environments with capital income risk.



Albanesi (2005) considers a model in which an entrepreneur exerts effort that determines the stochastic returns on investment. The effort is private information, but investment is observable. In this environment, there is a usual intertemporal investment wedge that is positive, as in Golosov et al. (2003). This wedge captures the incentive effects of increasing holdings of a risk-free asset with the return equal to the expected return to entrepreneurial capital. However, in addition to this aggregate wedge, there is also an individual entrepreneur's intertemporal wedge, which is the difference between the individual marginal benefit of increasing capital by one unit and the individual marginal cost. The key insight is that idiosyncratic returns to capital, which depend on entrepreneurial effort, differ from aggregate capital returns. Whereas the aggregate wedge is always positive, the individual wedge may be negative. The intuition behind this result is as follows. There are two effects of increasing entrepreneurial capital. The first effect is the adverse effect of additional capital providing more insurance and hence worsening incentives and decreasing effort. The second effect stems from a positive dependence of expected returns on entrepreneurial effort. When the second effect dominates, the resulting individual wedge is negative. This characterization of the optimal allocations shows an important issue that the capital taxes on entrepreneurs should carefully distinguish between the purpose of investment. If an entrepreneur accumulates financial investment such as a risk-free bond, then the positive aggregate wedge implies that it should be taxed. If the entrepreneur increases capital allocated to business, then the possibility of the negative individual wedge implies that this investment may be subsidized. The paper further explores the implementation of the optimal allocation under different market structures and shows that the differential asset taxation is optimal.

Shourideh (2012) considers a model in which there are two frictions: The returns on investment are private, and the investment itself is private as it can be diverted from productive capital to other assets. The main conclusion of the paper is that the taxes on various forms of investment, either an entrepreneurial investment or a financial investment, should be progressive. Consider a model in which production technology is constant returns to scale, and the return on investment consists of two components: heterogeneous entrepreneurial skill and unobservable idiosyncratic shocks. When there are no frictions—if the agent types, shocks, consumption, and investment are observable—the planner optimally allocates all investment to the most productive agent. Moreover, the utilitarian planner redistributes across entrepreneurial types and also provides full insurance against the investment return shocks. However, this allocation is not incentive compatible. For example, the agent with the highest entrepreneurial skill has no incentive to invest as his or her consumption does not depend on the amount of investment. Consider, in contrast, an economy with two informational frictions: Consumption and investment of the agent, as well as the investment returns, are unobservable. The income—the product of investment, entrepreneurial skill, and investment shocks—is observable. For simplicity, assume that the entrepreneurial types are also observable; most of the results carry over to the case of unobservable types. Suppose that in addition to investing in this risky project (inside savings), an entrepreneur can also invest in a risk-free asset (outside investment). The main question is how the informational frictions determine the distortions that the planner optimally imposes on these two types of savings. Consider first the outside savings. The trade-off from the point of view of the planner is as follows. As in the frictionless case, the planner wants to allocate investment to the most productive agent. However, because of the unobservability of investment, there is a cost that increases with the amount invested—given that full insurance is not possible, higher investment implies higher riskiness of consumption. In other words, more productive agents face higher risk. Suppose that there are no taxes on the outside savings. The more productive agents want to self-insure and hence invest more in the risk-free bond. This outside investment worsens incentives to invest—the planner then should impose progressive taxes on the outside investment. The intuition behind



taxing the inside savings is more involved. On the one hand, the more productive agents face a higher degree of consumption risk, and this implies a force for regressive taxes. On the other hand, the more productive agents invest more and therefore have higher income—a redistributive planner then wants to impose a progressive tax. Shourideh (2012) shows that in numerical simulations, the first force dominates, and the taxes on the inside savings are also progressive.

Several other papers study the optimal taxation of entrepreneurs. Scheuer (2013) analyzes a tractable multidimensional private information problem in which agents differ in their costs of setting up a firm and in their skill. He derives the results on the optimal nonlinear taxation of profits and labor income in this private information economy with endogenous firm formation. Panousi & Reis (2012) study the optimal taxation of capital in a model with uninsurable idiosyncratic investment risk. They combine elements of both Mirrlees and Ramsey frameworks and consider an affine tax system. Choi (2012) studies capital structure and the choice of debt versus equity in an optimal taxation framework.

Despite the important theoretical progress described here, the literature on optimal Mirrleesian taxation of entrepreneurs is still in relatively early stages. One promising direction is to move from stylized predictions in theoretical models to more quantitative studies that bring to the forefront a variety of empirical facts on entrepreneurship. This is similar in spirit to the evolution of the models described in Section 4 that moved from theory to quantitative implications and policy predictions. Such evolution will also bring the optimal taxation literature described here closer to the more classic public finance issues on this topic, such as Gentry & Hubbard (2000) and Cullen & Gordon (2007).

## 7.2. Human Capital

The most commonly analyzed shocks in the dynamic mechanism design models of optimal taxation are those to the agents' skills. An important question involves how to characterize implications for policy in the environments in which these skills are endogenously determined by human capital investment.

It is useful to separate the literature into three dimensions. The first two are models of either observable or unobservable human capital accumulation. The third includes models focusing on optimal education policies.<sup>17</sup>

We start the discussion with models of observable human capital accumulation. Bohacek & Kapicka (2008) study a model of human capital accumulation in which the individual skill is unobservable and is constant over time. Because the skill does not change over time, the informational friction reduces the model to an essentially static setting, and the dynamics comes from the accumulation of observable human capital. If there are no informational frictions and individual ability is observable, it is optimal for the allocations to satisfy two conditions. The first is that the marginal rate of substitution between consumption and leisure is equated to the wage rate. The second is that the Euler equation for human capital investment is undistorted. The unobservability of agents' skills in the absence of human capital introduces a (static) distortion between the marginal rate of substitution of consumption and leisure and the wage. The question is how this distortion interacts with human capital accumulation. Because there is a distortion in the labor choice, labor is lower than in the unconstrained optimum. If human capital accumulation is undistorted and the Euler equation is satisfied, then schooling is also distorted downward

---

<sup>17</sup>We also note several other papers that build on earlier studies of taxation with endogenous human capital (Bovenberg & Jacobs 2005, 2011; da Costa & Maestri 2007; Anderberg 2009).

compared to the undistorted optimum. The trade-off that the planner faces is whether to subsidize human capital at the expense of introducing an additional distortion in the labor choice. Bohacek & Kapicka (2008) show that human capital should be subsidized as long as it is optimal to increase over time the agent's labor supply relative to schooling. We return to the issues of implementation of the optimum with education subsidies below.

Kapicka & Neira (2012) study a two-period model in which the ability is constant over time and the learning effort is unobservable and yields a stochastic (risky) human capital accumulation. The model is conceptually similar to that of Shourideh (2012) discussed in Section 7.1. The difference, however, is that the unobservability of investment in Shourideh (2012) necessarily implies the unobservability of consumption, which essentially adds another friction. In Kapicka & Neira (2012), consumption is observable, and a model is closer to a classic moral hazard model applied to human capital accumulation. They show that there exists a savings wedge similar to that described in Section 4 and that this wedge in turn implies an increasing expected labor wedge. Moreover, the gains of switching to the optimal system from the current US system are large and of the order of 8% of consumption.

The most comprehensive analysis of human capital in the mechanism design approach to optimal taxation is the work of Stantcheva (2014). She studies a dynamic life-cycle model of labor supply and risky human capital formation. The wage is a function of both stochastic, persistent, and exogenous ability and endogenous human capital. Human capital is acquired throughout the life cycle through monetary expenses and training time and is observable.<sup>18</sup> The informational friction is the private observability of the initial ability of agents and of the lifetime evolution of ability. The main result is the characterization of the implicit subsidy to human capital and a decomposition of the key forces determining it. The first force is to counterbalance the distortions to human capital from implicit income and savings taxes. The second is to indirectly stimulate labor supply by increasing the wage (i.e., the returns to labor). The third is to redistribute, taking into account the differential effect of human capital on the pretax income of high- and low-ability people. When the wage elasticity with respect to ability is decreasing in human capital, human capital has a positive redistributive effect on after-tax income and a positive insurance value. It is then optimal to subsidize human capital expenses beyond simply insuring a neutral tax system with respect to human capital expenses.

The second strand of papers deals with the environments in which human capital investment is unobservable. These models are more difficult to analyze as they essentially add another friction, making consumption unobservable. In an early paper, Kapicka (2006) studies an environment in which an individual's productivity does not change over time, human capital is unobservable, and the government is restricted to the use of taxes that depend only on current income. This restriction is in the spirit of the models we discussed in Section 6. However, beyond the restriction of the absence of history independence, the taxes are otherwise unconstrained. He shows that the unobservability of human capital tends to decrease the labor wedge. Kapicka (2015) provides a more comprehensive model in which skills do not change over time, and human capital accumulation is unobservable. The key insight is that unobservable human capital effectively makes preferences over labor, nonseparable across age. Then the analysis can be reduced to analyzing classic Diamond (1998) and Saez (2001) elasticity-based formulas. Kapicka (2015) further shows that the optimal marginal income taxes decrease with age and that the behavior of cross-Frisch elasticities is essential in explaining the decline (see also Grochulski & Piskorski 2010 for a model of unobservable human capital accumulation).

---

<sup>18</sup>Stantcheva (2014) also extends the results to unobservable human capital accumulation.

Finally, we consider the issue of optimal education policies. The implementation of the optimal allocations in Bohacek & Kapicka (2008) and Stantcheva (2014) has direct policy relevance. The positive human capital wedge in Bohacek & Kapicka (2008) can be implemented with positive marginal schooling subsidies. The simulations of the welfare gains of implementing the optimal schooling subsidies compared to the system with a flat income yield a significant income gain, on the order of 8–9% of consumption. The implementation in Stantcheva (2014) is more involved as the model is stochastic. She shows two ways of implementing the constrained efficient allocations: income-contingent loans and a deferred deductibility system. In the deferred deductibility scheme, only part of the current investment in human capital can be deducted from current taxable income. The remainder is deducted from future taxable income to account for the risk and nonlinearity of the tax schedule. The deduction schemes for higher education expenses are rather common but are usually contemporaneous to the expense. The income-contingent loans proposed to implement the optimum are, however, rather different from the existing programs as the optimal loan repayment schedules are contingent on the past history of earnings and human capital investments. Most importantly, the income-contingent loans observed in practice focus almost exclusively on the downside, so that repayments can be deferred or forgiven in times of economic hardship. The optimal scheme is also focused on the upside, with repayments potentially increasing after a good history of earnings.

A sequence of papers by Findeisen & Sachs focuses specifically on education policies. Findeisen & Sachs (2012) study optimal education and tax policies in a heterogeneous agents model with idiosyncratic uncertainty. They show that optimal allocations can be decentralized by a combination of income taxes and student loans that come with income-contingent repayment. Their quantitative exploration reveals that the optimal loan repayment is increasing in income. Furthermore, they find significant welfare gains from the income contingency of loan repayment. Addressing the issue of time consistency, Findeisen & Sachs (2014b) analyze optimal education subsidies for the case in which the government lacks commitment. Once human capital investment is sunk, the government is tempted to make the income tax code more progressive than announced. This creates an additional rationale to distort the education margin. Because the government anticipates its temptation to increase income taxes, it makes education subsidies more progressive. The reason is that this implies a lower degree of inequality in human capital and therefore mitigates the incentives to increase taxes. From a more applied and quantitative point of view, Findeisen & Sachs (2014a) study the design of college subsidies and income taxes within a life-cycle model in which individuals differ among multiple dimensions of heterogeneity. In a carefully conducted quantitative exploration, they find that a universal increase in college subsidies or grants is to a large degree self-financing via higher tax revenue in the future. Targeting such a grant increase to students from low-income households even triggers an increase in future tax revenue that exceeds the costs of the grant increase. Concerning tax design, they find that the excess burden via the college channel is low as compared to the excess burden via the labor supply channel. Consequently, quantitative results for optimal income taxes are barely affected by the endogeneity of the college decision.

Finally, Krueger & Ludwig (2013) bridge the gap between the literature described here and the more standard quantitative analysis of the education subsidies. They use a restricted tax system approach. The government chooses the degree of progressivity (through a level of deduction in an otherwise linear income tax) and the level of the labor income tax, educational subsidies, and the amount of government debt. They quantitatively characterize the optimal mix of progressive income taxes and education subsidies in a large-scale overlapping generations model with endogenous human capital formation, borrowing constraints, income risk, intergenerational transmission of wealth and ability, and incomplete financial markets. Moreover, they also study

the transition to the steady state. The main result is that the welfare-maximizing fiscal policy is characterized by a substantially progressive labor income tax code and a sizable subsidy for college education. Both the degree of tax progressivity and the education subsidy are larger than they are under the current US policy.

### 7.3. Taxation of Bequests

An important policy question is how estate taxation should be designed. Farhi & Werning (2010) and Piketty & Saez (2013) are two recent contributions to this issue.

It is easiest to illustrate the intuition for Farhi & Werning's (2010) results through the two-period example they develop. Consider an economy with two generations: parents and children. The friction is the unobservability of the ability of the parents. The children do not work and only consume bequests. The parents' utility places a positive weight on children's utility. The key issue that determines the prescription for taxes is the nature of the social welfare function. Consider first the social welfare objective that maximizes the expected utility of the parents. In this case, the Atkinson & Stiglitz (1976) uniform commodity taxation result applies, and there is no need to distort the intertemporal choices of the parents. The estate tax is zero, and there is a nonlinear labor income tax on parents' income. Now, consider a social welfare criterion that places a positive weight on the utility of the second generation, the children. Farhi & Werning (2010) show two stark results: that estate taxation should be progressive and that the marginal estate taxes should be negative so that all parents face a marginal subsidy on bequests. The intuition is best understood through an externality interpretation. One can think of the consumption of the children that the parents enjoy as a good that imposes a positive externality on the generation of the children. The externality is, however, not a function of consumption but rather a function of the utility of consumption. The externality is thus stronger for children with low consumption. Hence, the implicit subsidy that corrects this nonlinear externality should be progressive. The authors then show that the optimal allocation can be implemented with a simple nonlinear tax on bequests, levied separately from the income tax.<sup>19</sup> Farhi & Werning (2013a) further develop the theory of estate taxation and focus on the differences in the degree of altruism. They show that optimal estate taxes depend crucially on redistributive objectives. These range from zero taxes in an environment in which there is no direct weight placed on children under a proper normalization of utility, to subsidies when there is a positive weight on children's welfare, to taxes when there is a Rawlsian preference for the equality of opportunity for children.

Piketty & Saez (2013) consider a dynamic stochastic model in which there are two types of heterogeneity: in tastes for bequests and in labor productivities. They consider a restricted tax system that is linear or has two brackets. Their key contribution is to express the optimal tax formula in terms of estimable sufficient statistics, including behavioral elasticities, distributional parameters, and social preferences for redistribution. They show that the optimal tax rate is positive and quantitatively large if the elasticity is low, bequests are quantitatively large and highly concentrated, and society cares mostly about those receiving little inheritance. In contrast, the optimal tax rate can be negative when society cares mostly about inheritors. For the realistic parameters, the optimal inheritance tax rate is on the order of 50–60%, or even higher for top bequests. In summary, these papers deliver important policy-relevant predictions for estate taxation, concluding that the sign of the tax depends on the choice of the welfare function and on the heterogeneity of altruism.

---

<sup>19</sup>One implication of the Shourideh (2012) model is that the bequest taxes should be negative.

## 7.4. Design of Optimal Pension Systems

Designing optimal pension systems is another key area of optimal taxation. In this section, we discuss several examples of recent work on optimal pension systems, using both the restricted tax functions approach and the dynamic mechanism design approach.

Huggett & Parra (2010) analyze how well a stylized version of the US social insurance system provides social insurance. They first solve for the fully optimal incentive-constrained allocation in a model in which there are only permanent shocks. They find that the gains of switching from the current US pension system to the optimal are very large and on the order of 4% of consumption. Then they consider two reforms in which they change a restricted tax function in a model with idiosyncratic stochastic shocks. First, Huggett & Parra (2010) consider an optimal reform of the social security benefit function, which is chosen in the class of piecewise-linear functions of average past earnings without changing the social security tax rate or the income tax system. This reform leads to almost no welfare gain in the permanent-shock model. In the model with stochastic shocks, the best benefit function leads to gains worth a 1.15% increase in consumption. The second reform is to eliminate the social insurance system and replace it with an optimal tax on the present value of earnings. An optimal present-value tax chosen within a class of increasing step functions achieves a welfare gain of 3.95% of consumption in the permanent-shock model—nearly all of the maximum possible welfare gain. However, in the full model, this optimal reform leads to no welfare gain.<sup>20</sup>

We now turn to the analysis of pension systems within the mechanism design approach. The work on optimal disability insurance (Golosov & Tsyvinski 2006) can also be thought of as understanding some elements of the optimal pension system that address permanent shocks.<sup>21</sup> The main insights discussed above are that this insurance should feature asset testing and the transfers should be conditioned on the wealth of the agents. The most comprehensive study of the issue is by Shourideh & Troshkin (2013). They study a life-cycle environment with both intensive and extensive labor margins. The key tension in their environment is between the redistribution and provision of incentives: Although it may be more efficient to have highly productive individuals work more and retire older, earlier retirement may be needed to give them incentives to fully realize their productivity while they work. The authors show that efficient retirement ages must increase in productivity and implement the optimal allocations with pension benefits that not only depend on the age of retirement, but are designed to be actuarially unfair. In a calibrated model, they find that it is optimal for the most productive workers to retire later than they do in the data and to retire older than their less productive peers, in contrast to the pattern observed in the data. They find significant welfare gains from switching to the optimal system, ranging from 1% to 5% of consumption. The key message of their paper is that distorting individual retirement decisions provides a powerful policy tool to undo incentives to retire earlier that are imbedded in any labor income tax.

Hosseini (2015) studies the role of social security in providing insurance when there is adverse selection in the annuity market in a life-cycle model in which individuals have private information about their mortality. The key additional friction is that contracts are nonexclusive and insurers cannot observe individuals' trades. These two informational frictions, unobservability of types and trades, lead the equilibrium price of annuities to be higher than the overall actuarially fair value of their payment. Hosseini develops a sophisticated quantitative analysis that uses data on subjective survival probabilities in the Health and Retirement Study, calibrates the model, and matches the key facts on the

<sup>20</sup>Golosov et al. (2013b) extend Huggett & Parra's (2010) model to one with both extensive and intensive margins.

<sup>21</sup>We also briefly mention a related literature that deals with designing sophisticated welfare systems, such as Pavoni & Violante (2007).

average replacement ratio of benefits and the average fraction of retirement wealth that is annuitized outside social security. The quantitative results show that social security can improve welfare ex ante, but the welfare gains are small. One reason for this is that the gains to low-mortality individuals are small owing to high annuity prices in the presence of social security. Social security leads to a crowding effect, which reduces the annuity demand by high-mortality types. This leads to more severe adverse selection in the private annuity market and a high premium. The crowding-out effect is related to the work on endogenous insurance in the dynamic optimal taxation models (Golosov & Tsyvinski 2007) and in the models with endogenously incomplete markets (Krueger & Perri 2011).

## 8. CONCLUSION

The dynamic public finance literature achieved significant progress in a relatively short period of time. What started as an abstract optimal dynamic contracting framework now has become an active research agenda that delivers theoretical, quantitative, and empirical results that are increasingly relevant to policy. An important development is the convergence between two approaches on optimal dynamic taxation: the dynamic mechanism design approach and the approach using sophisticated parametrically restricted tax functions. The dynamic mechanism design approach is moving in a direction of using increasingly rich and empirically motivated environments, delivering quantitative analysis that is increasingly policy relevant. The approach using the parametrically restricted tax functions is moving in a direction that incorporates increasingly more sophisticated taxes that are often motivated by the theoretical and quantitative findings of the dynamic mechanism design approach. One area of research that we hope will bring these approaches even closer together is the analysis of sophisticated tax reforms, attempted in Golosov et al. (2014).

We now briefly outline several other important strands of research that we omitted owing to space limitations. An active research agenda involves optimal unemployment insurance with a variety of frictions (Werning 2002; Shimer & Werning 2007, 2008; Chetty 2008; Pavoni 2009; Golosov et al. 2013a). Several recent papers study the design of sophisticated welfare policies in the context of labor markets (Pavoni & Violante 2007), disability insurance (Low & Pistaferri 2010), and the interaction of unemployment and disability insurance (Kitao 2013). Rothschild & Scheuer (2013, 2014) study the effect of taxation on occupational choice, Blundell & Shephard (2012) and Kleven et al. (2009) analyze taxation with more realistic family structure, and Golosov et al. (2013d) examine optimal taxation with multiple dimensions of unobservable heterogeneity.

## DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## ACKNOWLEDGMENTS

The authors thank the NSF for financial support. They also thank Maxim Troshkin and Gianluca Violante for helpful comments and Karen Lou, Kelley Lou, and Pedro Olea for assistance in preparing this manuscript.

## LITERATURE CITED

- Albanesi S. 2005. *Optimal taxation of entrepreneurial capital with private information*. Work. Pap., Columbia Univ., New York
- Albanesi S, Sleet C. 2006. Dynamic optimal taxation with private information. *Rev. Econ. Stud.* 73:1–30



- Ales L, Maziero P. 2009. *Accounting for private information*. Work. Pap., Finance Dep., Wharton Sch., Univ. Pennsylvania, Philadelphia
- Alesina A, Ichino A, Karabarbounis L. 2011. Gender-based taxation and the division of family chores. *Am. Econ. J. Econ. Policy* 3(2):1–40
- Anderberg D. 2009. Optimal policy and the risk properties of human capital reconsidered. *J. Public Econ.* 93:1017–26
- Atkeson A, Lucas RE. 1992. On efficient distribution with private information. *Rev. Econ. Stud.* 59:427–53
- Atkinson AB, Stiglitz JE. 1976. The design of tax structure: direct versus indirect taxation. *J. Public Econ.* 6:55–75
- Banks J, Diamond P. 2010. The base for direct taxation. In *Dimensions of Tax Design: The Mirrlees Review*, ed. J Mirrlees, S Adam, T Besley, R Blundell, S Bond, et al., pp. 548–648. New York: Oxford Univ. Press
- Blundell R, Pistaferri L, Preston I. 2008. Consumption inequality and partial insurance. *Am. Econ. Rev.* 98:1887–921
- Blundell R, Pistaferri L, Saporta-Eksten I. 2012. *Consumption inequality and family labor supply*. NBER Work. Pap. 18445
- Blundell R, Shephard A. 2012. *Optimal taxation in an empirical life-cycle model of labor supply*. Work. Pap., Dep. Econ., Univ. Pennsylvania, Philadelphia
- Bohacek R, Kapicka M. 2008. Optimal human capital policies. *J. Monet. Econ.* 55:1–16
- Bovenberg L, Jacobs B. 2005. Redistribution and education subsidies are Siamese twins. *J. Public Econ.* 89:2005–35
- Bovenberg L, Jacobs B. 2011. Optimal taxation of human capital and the earnings function. *J. Public Econ. Theory* 13:957–71
- Brett C, Weymark J. 2014. *Citizen candidates and voting over incentive-compatible nonlinear income tax schedules*. Work. Pap., Vanderbilt Univ., Nashville
- Chetty R. 2008. Moral hazard versus liquidity and optimal unemployment insurance. *J. Polit. Econ.* 116:173–234
- Choi KJ. 2012. *Mirrlees meets Modigliani-Miller: double taxation and capital structure*. Work. Pap., Univ. Calgary
- Conesa JC, Kitao S, Krueger D. 2009. Taxing capital? Not a bad idea after all! *Am. Econ. Rev.* 99:25–48
- Conesa JC, Krueger D. 2006. On the optimal progressivity of the income tax code. *J. Monet. Econ.* 53:1425–50
- Congr. Budg. Off. 2005. *Effective marginal tax rates on labor income*. CBO Pap., Congr. Budg. Off., Washington, DC
- Cullen JB, Gordon RH. 2007. Taxes and entrepreneurial risk-taking: theory and evidence for the US. *J. Public Econ.* 91:1479–505
- da Costa CE, Maestri LJ. 2007. The risk properties of human capital and the design of government policies. *Eur. Econ. Rev.* 51:695–713
- Deaton A, Paxson C. 1994. Intertemporal choice and inequality. *J. Polit. Econ.* 102:437–67
- Diamond P. 1998. Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates. *Am. Econ. Rev.* 88:83–95
- Erosa A, Gervais M. 2002. Optimal taxation in life-cycle economies. *J. Econ. Theory* 105:338–69
- Farhi E, Golosov M, Tsyvinski A. 2009. A theory of liquidity and regulation of financial intermediation. *Rev. Econ. Stud.* 76:973–92
- Farhi E, Werning I. 2010. Progressive estate taxation. *Q. J. Econ.* 125:635–73
- Farhi E, Werning I. 2012. Quantitative explorations of the inverse Euler equation. *J. Polit. Econ.* 120:398–445
- Farhi E, Werning I. 2013a. Estate taxation with altruism heterogeneity. *Am. Econ. Rev.* 103:489–95
- Farhi E, Werning I. 2013b. Insurance and taxation over the life cycle. *Rev. Econ. Stud.* 80:596–635
- Fernandes A, Phelan C. 2000. A recursive formulation for repeated agency with history dependence. *J. Econ. Theory* 91:223–47
- Findeisen S, Sachs D. 2012. *Education and optimal dynamic taxation: the role of income-contingent student loans*. Work. Pap., Univ. Cologne
- Findeisen S, Sachs D. 2014a. *Designing efficient education and tax policies*. Work. Pap., Univ. Cologne
- Findeisen S, Sachs D. 2014b. *Education policies and taxation without commitment*. Work. Pap., CRM Univ. Cologne

- Fukushima K. 2010. *Quantifying the welfare gains from flexible dynamic income tax systems*. Work. Pap., Univ. Wisconsin–Madison
- Garriga C. 2001. *Optimal fiscal policy in overlapping generations models*. Work. Pap. E01/66, Univ. Barcelona
- Gentry W, Hubbard G. 2000. Tax policy and entry into entrepreneurship. *Am. Econ. Rev.* 90:283–87
- Gentry W, Hubbard G. 2004. Entrepreneurship and household saving. *Adv. Econ. Anal. Policy* 4:1053
- Gervais M. 2012. On the optimality of age-dependent taxes and the progressive US tax system. *J. Econ. Dyn. Control* 36:682–91
- Golosov M, Kocherlakota N, Tsyvinski A. 2003. Optimal indirect and capital taxation. *Rev. Econ. Stud.* 70:569–87
- Golosov M, Maziero P, Menzio G. 2013a. Taxation and redistribution of residual income inequality. *J. Polit. Econ.* 121:1116–204
- Golosov M, Shourideh A, Troshkin M, Tsyvinski A. 2013b. Optimal pension systems with simple instruments. *Am. Econ. Rev.* 103:502–7
- Golosov M, Troshkin M, Tsyvinski A. 2013c. *Redistribution and social insurance*. Work. Pap., Princeton Univ., Princeton, NJ
- Golosov M, Troshkin M, Tsyvinski A, Weinzierl M. 2013d. Preference heterogeneity and optimal capital income taxation. *J. Public Econ.* 97:160–75
- Golosov M, Tsyvinski A. 2006. Optimal disability insurance: a case for asset testing. *J. Polit. Econ.* 114:257–79
- Golosov M, Tsyvinski A. 2007. Optimal taxation with endogenous insurance markets. *Q. J. Econ.* 122:487–534
- Golosov M, Tsyvinski A, Werquin N. 2014. *A variational approach to the analysis of tax systems*. Work. Pap., Princeton Univ., Princeton, NJ
- Gouveia M, Strauss RP. 1994. Effective federal individual income tax functions: an exploratory empirical analysis. *Natl. Tax J.* 47:317–39
- Green E. 1987. Lending and the smoothing of uninsurable income. In *Contractual Arrangements for Intertemporal Trade*, ed. EC Prescott, N Wallace, pp. 3–25. Minneapolis: Univ. Minnesota Press
- Grochulski B, Piskorski T. 2010. Risky human capital and deferred capital income taxation. *J. Econ. Theory* 145:908–43
- Guvenen F, Karahan F, Ozkan S, Song J. 2014a. *What do data on millions of US workers say about lifecycle earnings risk?* Work. Pap., Univ. Minnesota
- Guvenen F, Ozkan S, Song J. 2014b. The nature of countercyclical income risk. *J. Polit. Econ.* 122:621–60
- Heathcote J, Storesletten K, Violante G. 2010. The macroeconomic implications of rising wage inequality in the United States. *J. Polit. Econ.* 118:681–722
- Heathcote J, Storesletten K, Violante G. 2014a. Consumption and labor supply with partial insurance: an analytical framework. *Am. Econ. Rev.* 104:2075–126
- Heathcote J, Storesletten K, Violante G. 2014b. *Redistributive taxation in a partial-insurance economy*. Work. Pap., New York Univ.
- Heathcote J, Tsujima H. 2014. *Optimal income taxation: Mirrlees meets Ramsey*. Work. Pap., Fed. Reserve Bank, Minneapolis
- Hosseini R. 2015. Adverse selection in the annuity market and the role for social security. *J. Polit. Econ.* In press
- Huggett M, Parra J. 2010. How well does the US social insurance system provide social insurance? *J. Polit. Econ.* 118:76–112
- Kapicka M. 2006. Optimal income taxation with human capital accumulation and limited record keeping. *Rev. Econ. Dyn.* 9:612–39
- Kapicka M. 2013. Efficient allocations in dynamic private information economies with persistent shocks: a first-order approach. *Rev. Econ. Stud.* 80:1027–54
- Kapicka M. 2015. Optimal Mirrleesian taxation with unobservable human capital formation. *Am. Econ. J. Macroecon.* In press
- Kapicka M, Neira J. 2012. *Optimal taxation in a life-cycle economy with endogenous human capital formation*. Work. Pap., Univ. Calif., Santa Barbara
- Keane M, Rogerson R. 2012. Micro and macro labor supply elasticities: a reassessment of conventional wisdom. *J. Econ. Lit.* 50:464–76
- Kitao S. 2010. Labor-dependent capital income taxation. *J. Monet. Econ.* 57:959–74



- Kitao S. 2013. *A life-cycle model of unemployment and disability insurance*. Work. Pap., Hunter Coll., New York
- Kleven H, Kreiner C, Saez E. 2009. The optimal income taxation of couples. *Econometrica* 77:537–60
- Kocherlakota N. 2005. Zero expected wealth taxes: a Mirrlees approach to dynamic optimal taxation. *Econometrica* 73:1587–622
- Kremer M. 2002. *Should taxes be independent of age?* Work. Pap., Harvard Univ, Cambridge, MA
- Krueger D, Ludwig A. 2013. Optimal progressive labor income taxation and education subsidies when education decisions and intergenerational transfers are endogenous. *Am. Econ. Rev.* 103:496–501
- Krueger D, Perri F. 2011. Public versus private risk sharing. *J. Econ. Theory* 146:920–56
- Laroque G. 2005. Income maintenance and labor force participation. *Econometrica* 73:341–76
- Low H, Pistaferri L. 2010. *Disability risk, disability insurance and life cycle behavior*. Work. Pap., Stanford Univ., Stanford, CA
- Maag E, Steuerle CE, Chakravarti R, Quakenbush C. 2012. How marginal tax rates affect families at different levels of poverty. *Natl. Tax J.* 65:759–82
- Mankiw NG, Weinzierl MC. 2010. The optimal taxation of height: a case study of utilitarian income redistribution. *Am. Econ. J. Econ. Policy* 2(1):155–76
- Mirrlees J. 1971. Exploration in the theory of optimal income taxation. *Rev. Econ. Stud.* 38:175–208
- Panousi V, Reis C. 2012. *Optimal capital taxation with idiosyncratic investment risk*. Work. Pap., Fed. Reserve Board, Washington, DC
- Pavoni N. 2009. Optimal unemployment insurance with human capital depreciation and duration dependence. *Int. Econ. Rev.* 50:323–62
- Pavoni N, Violante GL. 2007. Optimal welfare-to-work programs. *Rev. Econ. Stud.* 74:283–318
- Peterman WB. 2013. Determining the motives for a positive optimal tax on capital. *J. Econ. Dyn. Control* 37:265–95
- Piketty T. 1997. La redistribution fiscale face au chômage. *Rev. Fr. Econ.* 12:157–201
- Piketty T, Saez E. 2013. A theory of optimal inheritance taxation. *Econometrica* 81:1851–86
- Quadri V. 2000. Entrepreneurship, saving, and social mobility. *Rev. Econ. Dyn.* 3:1–40
- Rothschild C, Scheuer F. 2013. Redistributive taxation in the Roy model. *Q. J. Econ.* 128:623–68
- Rothschild C, Scheuer F. 2014. *A theory of income taxation under multidimensional skill heterogeneity*. NBER Work. Pap. 19822
- Saez E. 2001. Using elasticities to derive optimal income tax rates. *Rev. Econ. Stud.* 68:205–29
- Saez E. 2002. Optimal income transfer programs: intensive versus extensive labor supply responses. *Q. J. Econ.* 117:1039–73
- Scheuer F. 2013. Adverse selection in credit markets and regressive profit taxation. *J. Econ. Theory* 148:1333–60
- Shimer R, Werning I. 2007. Reservation wages and unemployment insurance. *Q. J. Econ.* 122:1145–85
- Shimer R, Werning I. 2008. Liquidity and insurance for the unemployed. *Am. Econ. Rev.* 98:1922–42
- Shourideh A. 2012. *Optimal taxation of wealthy individuals*. Work. Pap., Dep. Finance, Wharton Sch., Univ. Pennsylvania, Philadelphia
- Shourideh A, Troshkin M. 2013. *Providing efficient incentives to work: retirement ages and the pension system*. Work. Pap., Finance Dep., Wharton Sch., Univ. Pennsylvania, Philadelphia
- Spear SE, Srivastava S. 1987. On repeated moral hazard with discounting. *Rev. Econ. Stud.* 54:599–617
- Stantcheva S. 2014. *Optimal taxation and human capital policies over the life cycle*. Work. Pap., Dep. Econ., Mass. Inst. Technol., Cambridge, MA
- Storesletten K, Telmer C, Yaron A. 2004. Consumption and risk sharing over the life cycle. *J. Monet. Econ.* 51:609–33
- Thomas J, Worrall T. 1990. Income fluctuation and asymmetric information: an example of a repeated principal-agent problem. *J. Econ. Theory* 51:367–90
- Weinzierl M. 2011. The surprising power of age-dependent taxes. *Rev. Econ. Stud.* 78:1490–518
- Weinzierl M. 2014. The promise of positive optimal taxation: normative diversity and a role for equal sacrifice. *J. Public Econ.* 118:128–42
- Werning I. 2002. *Optimal unemployment insurance with unobservable savings*. Work. Pap., Dep. Econ., Mass. Inst. Technol., Cambridge, MA
- Werning I. 2007. *Nonlinear capital taxation*. Work. Pap., Dep. Econ., Mass. Inst. Technol., Cambridge, MA
- Werquin N. 2014. *Income taxation with frictional labor supply*. Work. Pap., Dep. Econ., Yale Univ., New Haven, CT



# Contents

Knowledge-Based Hierarchies: Using Organizations to Understand the  
Economy  
*Luis Garicano and Esteban Rossi-Hansberg* . . . . . 1

Beyond Ricardo: Assignment Models in International Trade  
*Arnaud Costinot and Jonathan Vogel* . . . . . 31

The Roots of Gender Inequality in Developing Countries  
*Seema Jayachandran* . . . . . 63

Reconciling Micro and Macro Labor Supply Elasticities: A Structural  
Perspective  
*Michael Keane and Richard Rogerson* . . . . . 89

International Trade, Multinational Activity, and Corporate Finance  
*C. Fritz Foley and Kalina Manova* . . . . . 119

Policy Implications of Dynamic Public Finance  
*Mikhail Golosov and Aleh Tsyvinski* . . . . . 147

Media and Politics  
*David Strömberg* . . . . . 173

Forecasting in Nonstationary Environments: What Works and What Doesn't  
in Reduced-Form and Structural Models  
*Raffaella Giacomini and Barbara Rossi* . . . . . 207

Political Decentralization  
*Dilip Mookherjee* . . . . . 231

Household Debt: Facts, Puzzles, Theories, and Policies  
*Jonathan Zinman* . . . . . 251

Making Progress on Foreign Aid  
*Nancy Qian* . . . . . 277

Credit, Financial Stability, and the Macroeconomy <i>Alan M. Taylor</i> . . . . .	309
Job Creation, Job Destruction, and Productivity Growth: The Role of Young Businesses <i>John Haltiwanger</i> . . . . .	341
The Evolution of Social Norms <i>H. Peyton Young</i> . . . . .	359
Crime and Economic Incentives <i>Mirko Draca and Stephen Machin</i> . . . . .	389
Entrepreneurship and Financial Frictions: A Macroeconomic Perspective <i>Francisco J. Buera, Joseph P. Kaboski, and Yongseok Shin</i> . . . . .	409
The US Electricity Industry After 20 Years of Restructuring <i>Severin Borenstein and James Bushnell</i> . . . . .	437
Methods of Identification in Social Networks <i>Bryan S. Graham</i> . . . . .	465
Affirmative Action in Undergraduate Education <i>Peter Arcidiacono, Michael Lovenheim, and Maria Zhu</i> . . . . .	487
Is College a Worthwhile Investment? <i>Lisa Barrow and Ofer Malamud</i> . . . . .	519
The Schumpeterian Growth Paradigm <i>Philippe Aghion, Ufuk Akcigit, and Peter Howitt</i> . . . . .	557
Climate and Conflict <i>Marshall Burke, Solomon M. Hsiang, and Edward Miguel</i> . . . . .	577
The Gains from Market Integration <i>Dave Donaldson</i> . . . . .	619
Valid Post-Selection and Post-Regularization Inference: An Elementary, General Approach <i>Victor Chernozhukov, Christian Hansen, and Martin Spindler</i> . . . . .	649
<b>Indexes</b>	
Cumulative Index of Contributing Authors, Volumes 3–7	689
Cumulative Index of Article Titles, Volumes 3–7	692

## Errata

An online log of corrections to *Annual Review of Economics* articles may be found at <http://www.annualreviews.org/errata/economics>