Using the Discrete Model to Derive Optimal Income Tax Rates

Spencer Bastani*

First version: Oct 2, 2013

This version: July 24, 2014

Forthcoming in *FinanzArchiv*

Abstract

In this paper I demonstrate how the discrete model of optimum income taxation can be used to derive the structure of optimal income tax rates. I compare simulations of the discrete and continuous models of optimum income taxation under identical circumstances based on US wage data. The two models produce similar results once the number of types used to represent the skill distribution is sufficiently large. Regardless if the discrete or continuous model is used, in order to accurately capture the shape of the optimal schedule of marginal tax rates, a large number of taxpayers should be employed.

Keywords: optimum income taxation; simulations; computational methods

JEL Classification: H21; H24; C63

*Department of Economics, Uppsala University, Box 513, SE-751 20 Uppsala, Sweden. E-mail address: spencer.bastani@nek.uu.se. I am grateful to two anonymous referees, Sören Blomquist, Martin Hellwig, Luca Micheletto, Dominik Sachs, Laurent Simula, and Matti Tuomala for comments on an earlier draft of this paper. Financial Support from Riksbankens Jubileumsfond and The Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

1 Introduction

It was recognized already by Mirrlees (1971) that optimal taxation is a quantitative exercise. Analytical expressions alone offer little guidance about the shape of the optimal income tax with the exception of, that under reasonable assumptions, optimal marginal tax rates lie between zero and one and the maximum income should be taxed at a zero marginal rate if the skill distribution is bounded.[1] To derive further insight into the structure of optimal income taxes or to calculate the welfare effects of public policy it is necessary to turn to numerical simulations.

The primary purpose of this paper is to demonstrate the usefulness of the discrete optimal income tax model in deriving the shape of the optimal income tax schedule using numerical simulations. To the best of my knowledge this is also the first paper to compare simulations of the discrete and continuous formulations of the optimal income tax problem.

The discrete model of optimal income taxation was introduced by Stiglitz (1982), Stern (1982) and Guesnerie and Seade (1982). While Stiglitz and Stern focused on a model with two types of agents, Guesnerie and Seade introduced the so called chain-property and considered an arbitrary, but finite number of types. The discrete model has been further developed in the quasi-linear case by Weymark (1986) and Simula (2010).

The advantage of the discrete model was to clarify the mimicking issues behind taxation that were not completely transparent in the continuous model, where they were expressed in terms of differential equations. The model has paved the way for numerous important contributions in theoretical optimal income taxation. Nonetheless, the discrete model has not received much attention in scholarly debates about the structure of optimal income taxation.[2] A contributing factor in this regard has certainly been that the relationship between the continuous and discrete models of optimal income taxation has not been made very clear in the literature and that the

[1] There is also a result stating that the marginal tax rate should be zero at the lowest income level if there is no bunching there. However, this result is of little practical relevance and has been rejected by simulation studies of optimal income taxation.

[2] In two recent survey articles Diamond and Saez (2011) and Piketty and Saez (2012) describe the discrete model as being best suited for making theoretical points within small stylized models and not for characterizing the shape of the optimal schedule or for policy recommendations.

discrete model is most often discussed with reference to the two type model first analyzed by Stiglitz (1982). However, in a recent paper Hellwig (2007a) proposes a unifying framework for optimal income taxation which embeds the finite and continuum type models as special cases. Hellwig demonstrates that the continuous and finite approaches are in fact mathematically equivalent.[3]

Another, perhaps more fundamental reason for the discrete model being perceived as lacking policy relevance, is the lack of simulation studies using the discrete model.[4] The most influential simulation results of optimal income tax schedules presented in the literature, such as those of Mirrlees (1971), Tuomala (1984, 1990), Diamond (1998) and Saez (2001) have been based on the optimal income tax model with a continuum of types. These studies have been very useful, and the continuous model has made it possible to link theoretical optimal tax results to empirically observable elasticities, however it does not seem necessary to focus on the continuous model when performing numerical simulations. In fact, numerical simulations could be made easier by adopting a discrete model of optimum income taxation since it allows researchers to compute optimal income tax schedules by solving a simple constrained nonlinear optimization problem.[5]

The simulation approach to optimal income taxation using the discrete model amounts to employing a large set of discrete taxpayer types, solve for the optimal allocation using constrained optimization techniques, and then construct the optimal income tax schedule. This numerical procedure is very transparent and relies on principles which are well understood. Moreover, it is perhaps more accessible to researchers as compared with traditional simulation approaches using the continuous model.[6] An additional benefit of the discrete model is that

[3] In a different paper, Hellwig (2007b) uses the maximum theorem to obtain a continuity result in the transition from discrete to continuous distributions under an additional assumption (weakly decreasing consumption-specific risk aversion).

[4] A reason for this might have been computational. Given recent computational developments it is feasible to solve discrete type optimal income tax problems by setting up a constrained nonlinear optimization problem with a large number of taxpayer types.

[5] Computing the optimal income tax in the continuous framework of Saez (2001) for instance requires solving a fixed-point problem, a procedure which can be cumbersome to implement numerically.

[6] Simulations of the discrete model have of course been presented earlier in the literature (see for instance Weinzierl 2011 and Bastani et al. 2013) but have usually been restricted to models with a small number of types and have not been used to characterize the overall shape of the optimal income tax schedule.

no smoothness or continuity assumptions are imposed on the allocation, and the second order condition is incorporated explicitly as a set of constraints into the problem.[7]

The discrete model poses a well behaved convex programming problem under mild assumptions on individual preferences and the redistributive taste of the government. In this case one only has to take into account the downwards self-selection constraints linking two adjacent individuals in the skill distribution. Convexity, of course, significantly simplifies computation. It is important to note, however, that computations of the discrete optimal tax model are not restricted to cases where agent monotonicity holds. It is possible to solve models when there is heterogeneity both in terms of skills and preferences so that the population is indexed by a two-dimensional set.[8]

For comparative purposes I present simulations based on the continuous formulation of the optimal income tax problem, adopting a numerical procedure of the type used in Saez (2001). Note that simulations of the continuous model have to be made using an approximate solution to a functional equation on a discrete grid. Thus simulations of optimal income tax schedules employ discrete skill distributions regardless if the type set in the underlying economic model is discrete or continuous.

I find that numerical simulations based on the discrete and continuous model produce similar results, given that the number of types used to represent the skill distribution is sufficiently large. By varying the number of types used in the simulations, I find evidence that in order to accurately capture the shape of the optimal schedule of marginal tax rates, a large number of taxpayers should be employed. This holds true both for the discrete and continuous model. Moreover, there seems to be a tendency of the two models to converge to the same optimal

[7] In contrast, when using standard approach to simulate the continuous model, the second order condition needs to be checked ex-post. See Brito and Oakland (1977) and Ebert (1992) for the importance of the second order condition in the continuous model and Guesnerie and Seade (1982) or Hellwig (2007a) for its role in the discrete model.

[8] With multi-dimensional heterogeneity there is no obvious way to order individuals along a unidimensional scale, thus single-crossing cannot be established and a full set of incentive constraints has to be included into the problem. The method can still be used if all these constraints are included into the optimization, however convergence to an optimal solution cannot be guaranteed.

income tax schedule as the number of types increase.

The paper is organized as follows. In section 2 I briefly describe the discrete optimal tax model, section 3 discusses the calibration strategy and section 4 presents the results.

2 The Discrete Optimal Income Tax Model

Individuals have preferences over consumption C and labor supply ℓ represented by a twice-differentiable, strictly concave utility function $u(C, \ell)$ where I assume $u'_C > 0$, $u'_\ell < 0$. Moreover I assume that leisure (the negative of labor supply) is a normal good. Agents are indexed by their exogenous productivity (or skill level) w, have a continuous labor supply choice (there are no fixed costs of work), and generate income equal to $Y = w \cdot \ell$. The government levies a nonlinear tax $T(Y)$ which depends only on income, hence neither ℓ or w are observable to the government. Let $U(C, Y, w) \equiv u(C, Y/w)$. The individual solves

$$\max_{\ell} \; u(w \cdot \ell - T(w \cdot \ell), \ell) \quad \Leftrightarrow \quad \max_{Y} \; U(Y - T(Y), Y, w) \tag{1}$$

At points where the tax function $T(Y)$ is differentiable, we can express the (private) optimality condition for the above problem

$$1 - T'(Y) = -\frac{U_Y(C, Y, w)}{U_C(C, Y, w)} \equiv \Omega(C, Y, w) \tag{2}$$

where $\Omega(C, Y, w)$ is the marginal rate of substitution for an agent of ability type w at the point (C, Y) in the consumption-income space representing the slope of the indifference curve at this point. This notation allows us, as is customary in the literature, to express the marginal tax rate as $T'(Y) = 1 - \Omega(C, Y, w)$.

Consider a population consisting of N agents indexed by the set $\mathcal{S} = \{1, 2, \ldots, N\}$. Assign to each agent $i \in \mathcal{S}$ the productivity parameter w_i and assume without loss of generality that

$w_j > w_i$ whenever $j > i$. Let π_i be the proportion (mass) of individuals of type i ($\sum_{i=1}^{N} \pi_i = 1$) and denote by α_i the weight the government attaches to the welfare of individual i. The problem of the government is the maximization of total welfare defined as

$$\max_{\{C_i, Y_i\}_{i=1}^{N}} \sum \alpha_i \pi_i U(C_i, Y_i, w_i) \tag{3}$$

subject to the public budget constraint

$$\sum_{i=1}^{N} \pi_i (Y_i - C_i) \geq R \tag{4}$$

and the incentive compatability constraints

$$\forall i, j \in \mathcal{S} : \quad U(C_i, Y_i, w_i) \geq U(C_j, Y_j, w_i) \tag{5}$$

where R is an exogenous revenue requirement.

The formulation of the social objective embeds the Utilitarian social welfare function (obtained by setting $\alpha_i = 1, i = 1, \ldots, N$) and the Max-min social welfare function (obtained by setting $\alpha_1 = 1$ and $\alpha_j = 0, j = 2, \ldots, N$) as special cases.

As shown by Röell (1985), given the concavity of the utility function and the normality of leisure, the single-crossing (Spence-Mirrlees) condition $\Omega_w^{'}(C, Y, w) < 0$ is satisfied. This implies that individuals with higher productivity have flatter indifference curves in the (C, Y) space. Moreover, the concavity of the utility function implies that redistribution is desirable even under a Utilitarian social welfare function. As shown by Hellwig (2007a), these properties allow the optimal income tax problem above to be replaced by a weakly relaxed problem involving only downwards self-selection constraints linking two adjacent individuals plus a monotonicity condition on consumption.[9]

[9] Guesnerie and Seade (1982) refer to this as a 'simple monotonic chain to the left'.

Thus, the above problem is replaced with the weakly relaxed problem obtained by replacing the set of incentive constraints (5) with the set of downwards adjacent constraints:

$$\forall i \in \{2, \ldots, N\} : \quad U(C_i, Y_i, w_i) \geq U(C_{i-1}, Y_{i-1}, w_i) \tag{6}$$

plus a weak monotonicity constraint on consumption[10]:

$$\forall i \in \{2, \ldots, N\} : c_i \geq c_{i-1} \tag{7}$$

I attach the Lagrange multipliers μ to the public budget constraint (4), λ^i to the ith self-selection constraint in (6) and σ_i to the ith montonicity constraint in (7). At points where the tax schedule is differentiable, and the monotonicity constraint is not binding ($\sigma_i = 0$), the first order conditions to the government's problem can be combined to produce an expression for the marginal tax rate. The optimal marginal tax rate for an individual of productivity $i \in \{1, \ldots, N-1\}$ is

$$T'(Y^i) = \frac{1}{\pi^i} \frac{\frac{d\widetilde{U}_{i+1}}{dC_i}}{\mu} \lambda^i \left(\Omega(C_i, Y_i, w_i) - \widetilde{\Omega}(C_i, Y_i, w_{i+1}) \right) \tag{8}$$

where $\frac{d\widetilde{U}_{i+1}}{dC_i}$ and $\widetilde{\Omega}(C_i, Y_i, w_{i+1})$ denotes, respectively, the marginal utility of consumption and the marginal rate of substitution of an agent of productivity type $i+1$ choosing the income point intended for productivity type i. As is well known, for the most productive agent $i = N$ we get $T'(Y_i) = 0$.

Note that this implicit expression for the marginal tax rate is not used to compute the

[10]Hellwig shows that irrespective of the type set being continuous or discrete, under single-crossing and a redistributive assumption, upwards incentive constraints can be replaced with the monotonicity condition. Since these assumptions plus incentive compatability implies monotonicity, downwards incentive compatability and monotonicity is weaker than incentive compatability. Nonetheless, the solutions to the weakly relaxed optimal income tax problem and the optimal income tax problem are the same. When the type set is discrete, as in the current context, downwards incentive compatability is equivalent to downwards *adjacent* incentive compatability.

optimal profile of marginal tax rates for the discrete model. Instead the allocation is computed *directly* by finding the optimal income points $\{C_i, Y_i\}_{i=1}^N$ as a solution to the social optimization problem above.

It is useful to spend some time discussing the determinants of optimal marginal tax rates as this will become relevant when interpreting the numerical results. Formula (8) highlights in an intuitive manner the considerations involved in the determination of the optimal marginal tax rate. A positive distortion is imposed on agent i in order to deter the more productive agent $i+1$ from reducing his/her labor in an effort to reproduce the labor income of agent i. Redistribution is achieved by imposing positive marginal tax rates. A positive marginal tax rate is possible only insofar as type $i+1$ has a greater willingness to supply income than type i at the income point (C_i, Y_i). This explains the presence of the difference $\Omega(C_i, Y_i, w_i) - \widetilde{\Omega}(C_i, Y_i, w_{i+1})$. Due to the single-crossing condition we have $\Omega(C_i, Y_i, w_i) > \widetilde{\Omega}(C_i, Y_i, w_{i+1})$ and marginal tax rates are guaranteed to be positive throughout the income distribution. Furthermore, the optimal distortion is decreasing in the probability mass of agents of type i (π_i) and increasing in the value of redistribution to agent i as determined jointly by the social welfare weight α_i and the marginal utility of consumption.

3 Calibration

To perform simulations the skill distribution must be specified. Two approaches have been taken in the literature. Saez (2001) calibrated the skill distribution so that his model of taxable income supply matched the empirical earnings distribution under the actual tax system in the United States. The procedure is difficult to implement unless there exists a one-to-one mapping between earnings and skills (ability) which can only be retrieved for a small class of utility functions. In addition the procedure is sensitive to how the actual tax system is modeled. Beginning with Mirrlees (1971) another common approach is to specify the skill distribution directly using wages as a proxy for skills.[11] In this paper I use the same data as Mankiw et al. (2009) and follow their calibration procedure closely. These authors use the March wave of the Current Population Survey (CPS) to parameterize the U.S. wage distribution. I employ the wages for 2007 using a lognormal parameterization with parameters $(\mu, \sigma) = (2.757, 0.5611)$ up to a wage level of \$42.50, and from this point on fit a Pareto tail with parameter $a = 2$, following Saez (2001). The distribution is then scaled so that it is continuous and integrates to one. In accordance with most of the earlier literature I assume that there is a mass of disabled workers, earning a wage of \$0.01. The fraction of disabled agents in the population is assumed to be 5%. The wages should be interpreted as hourly wages.[12]

The wage rates are constructed so that the distance between any two consecutive wage rates is fixed. Increasing the number of types used to represent a given wage distribution thus decreases the distance between any two wage rates by the same amount.

An alternative is to divide the wage distribution into a set of equally sized bins, where each bin represents a given fraction of the total population. The median wage for the individuals in each bin can then be calculated and used to represent the wages of the individuals belonging to each bin.[13] This procedure implies that the distance between the points used to approximate the wage distribution in regions where many individuals are located is smaller compared to in regions where few people are located (such as in the top of the income distribution). I have performed simulations with the discrete model for various ways to represent the wage distribution, and found that bunching is more likely to occur using this alternative way of representing the wage distribution.

[11] Kanbur and Tuomala (1994) modified the approach of Mirrlees (1971) by choosing the values of the skill distribution parameters so that the income distribution inferred from the skill distribution matches the empirical distribution of earnings under the optimal nonlinear income tax system.

[12] It is worth mentioning that instead of constructing wages by sampling from some pre-determined parameterized statistical distribution, it is possible to use any nonparametric density estimator on the actual wages in the data. I have made kernel density estimations based on actual CPS wage data obtaining very similar wage distributions as those obtained using the combination of the Lognormal and Pareto distribution.

[13] If the number of bins is 100, this is equivalent to using percentiles to represent the wage distribution.

In the numerical simulations I employ the utility function

$$u(c,h) = \log(c) - \xi \frac{h^{1+k}}{1+k}, \tag{9}$$

which is a standard utility function commonly used in the optimal tax and quantitative macroeconomics literature. This is also one of the utility functions used by Saez (2001).[14]

Following Mankiw et al. (2009) I set $\xi = 2.55$ and the parameter k is set equal to 2 implying a Frisch elasticity of labor supply equal to 0.5. The parameter ξ is a scaling parameter and does not affect the shape of the optimal tax schedule.[15] In the simulations presented below I will focus on a Utilitarian social welfare function.

4 Results

The discrete model is set up in the modeling language AMPL and solved using the state-of-the-art nonlinear optimization package KNITRO developed by Ziena Inc. A first goal is to make a direct comparison of the discrete approach with the continuous approach under identical circumstances. For this purpose I have compared the tax schedules I obtain with the discrete model, with the tax schedules I obtain from the algorithm that Mankiw et al. (2009) used to produce their simulations of the continuous model.[16]

In figure 1 I present the simulations for the discrete model and in figure 2 I present the simulations for the continuous model, where in both figures I vary the number of types in the simulations. Results are presented for 125, 250, 500, 1000, and 2000 types. I find that both the continuous and discrete models are sensitive to the number of types used in the simulations, but both models display clear signs of convergence as the number of types increase.[17]

[14] Using a non-separable utility function would pose no additional difficulty in the current setting, but I focus on the separable specification to maintain comparability with previous studies.

[15] Note that choosing ξ is equivalent (up to a constant) to changing the scale of the wage distribution.

[16] The continuous model is solved in MATLAB. I thank Danny Yagan for making the code available and for the excellent documentation which made it possible to replicate their results.

[17] Mankiw et al. (2009) use the continuous formulation of the optimal income tax problem employing 144 discrete points to represent the wage distribution.

Note that both the discrete and continuous model relies on a discrete skill distribution to approximate the underlying 'true' continuous (Lognormal + Pareto) skill distribution. One possible reason for the sensitivity of the marginal tax rates to the number of types in the continuous model is that the solution algorithm to the continuous model assumes that the marginal tax rate is constant between two adjacent wage levels. Another reason, applying both to the continuous and discrete model, relates to the approximation of the underlying skill distribution. It is a well-known fact that as the number of data points in such a discrete approximation increases, a better fit with the underlying continuous skill distribution is obtained. In general, it is difficult to separate this approximation issue from other potential reasons for why the shape of simulated optimal income tax schedules change as the number of types is being varied. An in-depth analysis of this issue is left as a topic for future research.

In figure 3 I plot the 125,250, and 500-type cases from figures 1 and 2, side-by-side, making it easier to compare the convergence properties of the discrete and continuous model. As evident from figure 3, the discrete model converges from below, and the continuous model from above.

Finally, in figure 4 I graph the schedules for both models using 2000 types. From this figure it can be seen that, when the number of types is very high, the schedules are almost identical. Hence, one can conclude that as the number of types increase, the discrete and continuous model produce very similar results.

The conclusion I draw from the results presented is that the discrete model should be considered a viable alternative to the continuous model when using numerical simulations to derive optimal income tax rates.

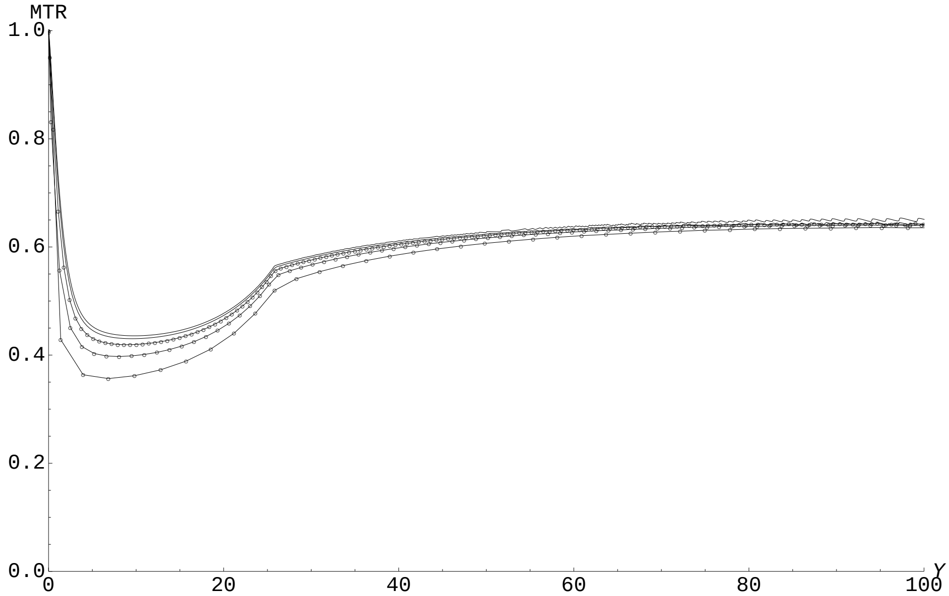

Figure 1: Optimal profiles of marginal tax rates for discrete model using 125, 250, 500, 1000, and 2000 types (from bottom to top).

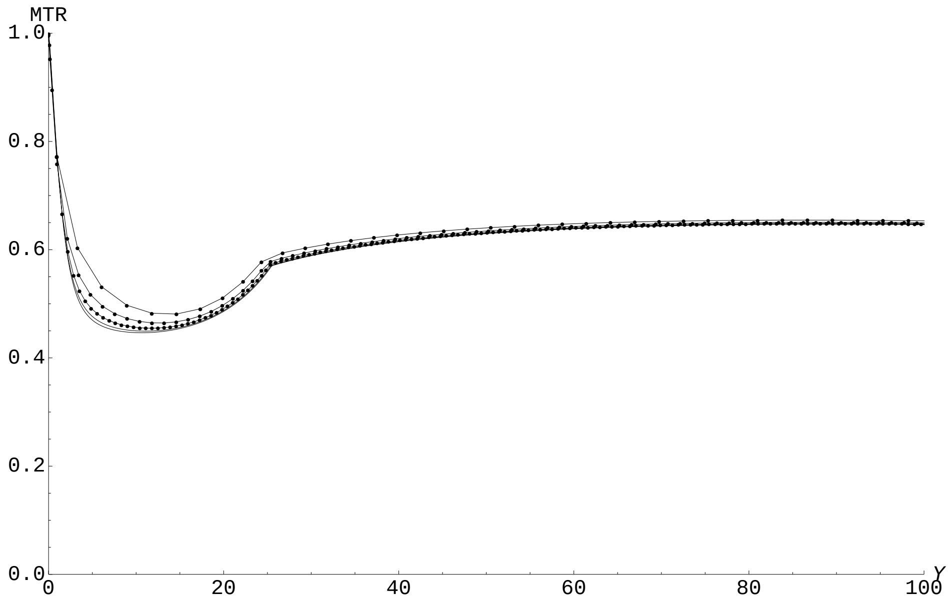

Figure 2: Optimal profiles of marginal tax rates for continuous model using 125, 250, 500, 1000, and 2000 types (from top to bottom).

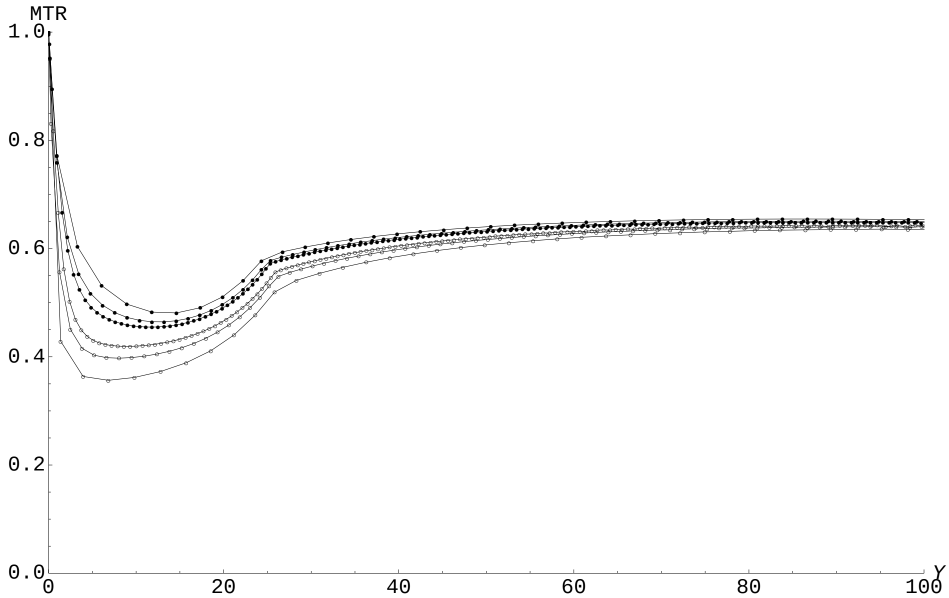

Figure 3: Optimal profiles of marginal tax rates for continuous model (top three graphs) and discrete model (bottom three graphs) using 125, 250, and 500 types. The figure displays a subset of the schedules presented in figures 1 and 2.

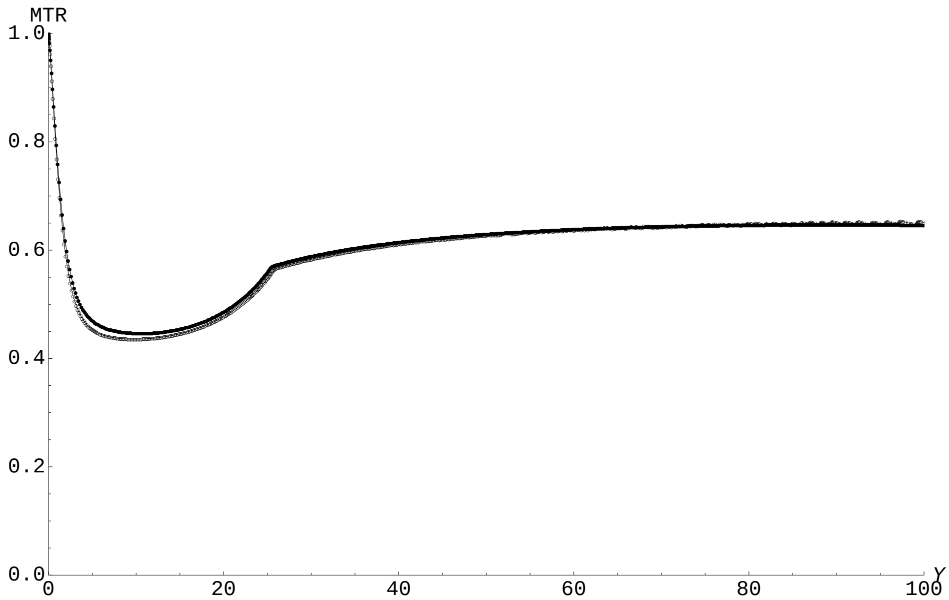

Figure 4: Optimal profiles of marginal tax rates for continuous model (top graph) and discrete model (bottom graph) using 2000 types.

References

Bastani, S., Blomquist, S., and Micheletto, L. (2013). The welfare gains of age-related optimal income taxation. *International Economic Review*, Vol 54-4.

Brito, D. L. and Oakland, W. H. (1977). Some properties of the optimal income-tax. *International Economic Review*, 18(2):pp. 407–423.

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

Diamond, P. A. and Saez, E. (2011). The case for a progressive tax: From basic research to policy recommendations. *Journal of Economic Perspectives*, 25(4):165–90.

Ebert, U. (1992). A reexamination of the optimal nonlinear income tax. *Journal of Public Economics*, 49(1):47 – 73.

Guesnerie, R. and Seade, J. (1982). Nonlinear pricing in a finite economy. *Journal of Public Economics*, 17:157–179.

Hellwig, M. F. (2007a). A contribution to the theory of optimal utilitarian income taxation. *Journal of Public Economics*, 91:1449–1477.

Hellwig, M. F. (2007b). The undesirability of randomized income taxation under decreasing risk aversion. *Journal of Public Economics*, 91:791 – 816.

Kanbur, R. and Tuomala, M. (1994). Inherent inequality and the optimal graduation of marginal tax rates. *Scandinavian Journal of Economics*, 96(2):275–82.

Mankiw, N. G., Weinzierl, M., and Yagan, D. (2009). Optimal taxation in theory and practice. *Journal of Economic Perspectives*, 23(4):147–74.

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

Piketty, T. and Saez, E. (2012). Optimal labor income taxation. NBER Working Papers 18521, National Bureau of Economic Research, Inc.

Röell, A. A. (1985). A note on the marginal tax rate in a finite economy. *Journal of Public Economics*, 28(2):267 – 272.

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies*, 68:205–229.

Simula, L. (2010). Optimal nonlinear income tax and nonlinear pricing: optimality conditions and comparative static properties. *Social Choice and Welfare*, 35(2):199–220.

Stern, N. H. (1982). Optimum taxation with errors in administration. *Journal of Public Economics*, 17:181–211.

Stiglitz, J. E. (1982). Self-selection and Pareto-efficient taxation. *Journal of Public Economics*, 17:213–240.

Tuomala, M. (1984). On the optimal income taxation: Some further numerical results. *Journal of Public Economics*, 23(3):351–366.

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

Weinzierl, M. C. (2011). The surprising power of age-dependent taxes. *Review of Economic Studies*, 78(4):1490–518.

Weymark, J. A. (1986). A reduced-form optimal nonlinear income tax problem. *Journal of Public Economics*, 30(2):199–217.