

Optimal Income Taxation with Multidimensional Taxpayer Types*

Preliminary and Incomplete

Kenneth L. Judd[†] Che-Lin Su[‡]

November 22, 2006

Abstract

Beginning with Mirrlees, the optimal taxation literature has generally focused on economies where individuals are differentiated only by their productivity. In this paper we examine models where individuals are differentiated by two or more characteristics. We examine cases where individuals differ in productivity, elasticity of labor supply, and their “basic needs”. We find that the extra dimensionality produces substantively different results. In particular, we find cases of negative marginal tax rates for some high productivity taxpayers. In our examples, income becomes a fuzzy signal of who should receive a subsidy under the planner’s objective, and the planner chooses less redistribution than it would in more homogeneous societies. We also show examine optimal taxation in an OLG model, and find that there is much less redistribution if the planner, as most governments do, does not discriminate on the basis of age. Multidimensional optimal tax problems are difficult nonlinear optimization problems because the linear independence constraint qualification does not hold at all feasible points and often fails to hold at the solution. To robustly solve these nonlinear programs, we use SNOPT with an elastic-mode, which has been shown to be effectively for degenerate nonlinear programs.

*This research is still a work in progress.

[†]Hoover Institution, Stanford, CA 94305 (judd@hoover.stanford.edu) and NBER.

[‡]CMS-EMS, Kellogg School of Management, Northwestern University (c-su@kellogg.northwestern.edu).

1 Introduction

The Mirrlees (1971) optimal tax analysis and much of the literature that followed assumed that people differ only in their productivity, and shared common preferences over consumption and leisure. The world is not so simple. People differ in more ways than their productivity. Any realistic model would account for multidimensional heterogeneity. For example, some high ability people have low income because they prefer leisure, or the life of a scholar and teacher. In contrast, some low ability people have higher-than-expected income because circumstances, such as having to care for many children, motivate them to work hard. Despite the unrealistic aspects of this one-dimensional analysis, its conclusions have been applied to real tax problems. For example, Saez (2001) says “optimal income tax schedules have few general properties: we know that optimal tax rates must lie between 0 and 1, and that they equal zero at the top and bottom.”

The presence of multidimensional heterogeneity is critically important for optimal taxation. In one-dimensional models, there is often a precise connection between what the government can observe, income, and how much the government wants to help (or tax) the person. Incentive constraints alter this connection, but the solution often involves full revelation of each individual’s type. However, this clean connection between income and “merit” is less precise in the presence of multidimensional heterogeneity. If an individual has low income because he has low productivity, then we might want to help him even when we would not want to help a high-productivity person choosing the same income because of his preference for leisure.

The basic question we ask here is “Does the presence of multidimensional heterogeneity reduce the optimal level of redistribution?” The intuition is clear: a one-dimensional signal like income is a noisy signal of merit, and as the signal to noise ratio falls, we will rely less on it to implement any policy. Our initial results support this conjecture.

There have been some attempts to look at multidimensional tax problems with a continuum of types. For example Mirrlees considers a general formulation, and Wilson (1993, 1995) looks at similar problems in the context of nonlinear pricing. However, both assume a first-order approach. This approach is justified in one-dimensional cases where the single-crossing property holds and implies that at the solution each type is tempted only by the bundles offered to one of the his two neighboring types. This approach leads to a system of partial differential equations. Tuomala (1999) solves one such example numerically, as does Wilson in the nonlinear pricing context. Unfortunately, no one has found useful assumptions that justify the first-order approach in the multidimensional case. The first-order approach assumes that the only alternatives that are tempting to a taxpayer are the choices made by others who are very similar in their characteristics. The single-crossing property in the one-dimensional case creates a kind of monotonicity that can be exploited to rule out the need to make global comparisons. However, there is no comparable notion of monotonicity in higher dimensions since there is no simple complete ordering of points in multidimensional Euclidean spaces.

The absence of an organizing principle like single-crossing does not alter the general theory; it only makes it harder. The general problem is still a constrained optimization problem - maximize social objective subject to incentive and resource constraints. However, the number of incentive constraints is enormous, and, unlike the single-crossing property in the one-dimensional case, there are no plausible assumptions that allow us to reduce the size of this problem.

There have been several studies which have extended the Mirrlees analysis in multidimensional directions. The references lists several such papers. Our initial perusal of that literature¹ indicates that there has been only limited success. Some papers look at cases with a few (such as four) types of people, some consider using other instruments, such as commodity taxation, to sort out types, and some prove theorems of the form “IF the solution has property A, THEN it also has property B” leaving us with little idea about how plausible Property A is.

Since multidimensional models are clearly more realistic than one-dimensional models, we numerically examine optimal taxation with multidimensional heterogeneity. First, we examine a two-dimensional case with both heterogeneous ability and heterogeneous elasticity of labor supply. This is a particularly useful example since it demonstrates how easy it is to get results different from the one-dimensional case and how any search for simplifying principles like single crossing is probably futile. In particular, we find that the optimal tax rate can be negative for the highest income earners! This contradicts one of the most basic results in the optimal tax literature, and the contradiction is due to the failure of binding incentive constraints to fall into a simple pattern. Second, we look at another two-dimensional model where people differ in ability and “basic needs”. In this model income is a bad signal of a person’s marginal utility of consumption (which, at the margin, is what the planner cares about) because high income could indicate high wages or an individual with moderate ability but large expenditures, such as medical expenses, that consume resources more than they contribute to utility. Third, we compute the solution to the optimal tax policy for a case of three-dimensional heterogeneity combining heterogeneous ability, elasticity of labor supply and “basic needs”.

Fourth, we consider the case where people differ in age. One can pursue a standard Mirrlees approach in dynamic models where an individual announces his type and then accepts a sequence of income-consumption bundles from the government. This is the approach pursued in Golosov, et al. (2003) and Kocherlakota (2005). The tax policies in these models may be strongly history dependent. That is, a person’s tax payment today depend on past income. Actual income tax policies do not have this form. In contrast, US tax liabilities for a year depend on the income earned in that year². We make no attempt to explain why US

¹This draft does not contain any detailed description of the literature. Our apologies to those whose work we ignore in this draft.

²There are some deviations from this. In the past, income averaging created some intertemporal connections. There are a few special features based on age. However, these exceptions are of minor importance and are surely much smaller than the interdependencies implied by dynamic mechanism models. Even the capital gains tax rules allows a taxpayer to take the records of his purchases to the grave with him.

tax law does not use memory³; instead, we examine the optimal, incentive compatible, but memoryless tax policy. Here again the government will face a difficult problem in deciding taxation policies. If a person has low income, is it because he is a middle-aged individual with low ability, or is it a young person with high-ability at the beginning of a steep life-earnings profile? The government may want to help the former, but not the latter. The solution to the memoryless optimal tax problem is a solution to another incentive problem that takes into account the lack of memory on the part of the planner.

Our computations not only demonstrate the feasibility of our approach, but they also point towards interesting economic conclusions. When we compare the results for the one-, two-, and three-dimensional models, we find that the optimal level of redistribution is significantly less as we add heterogeneity. In the OLG model, we also find that the inability to discriminate on age substantially reduces redistribution. The intuition is clear: in a complex world, income is a less reliable signal of whether we want to tax or subsidize a particular type of individual; hence, we use the signal less and implement less redistribution.

One response to the reduced ability to redistribute income is to add instruments to tax policy, such as taxing those commodities demanded by those we want to tax or to gather more facts about a taxpayer, such as his wage. Of course, the marginal value to a planner of making the tax system more complex will have to be balanced against the implementation costs. While it is reasonable to conjecture that we do not want to make the tax code as complex as the world, the exact balance between complexity costs and benefits is not immediately clear. In any case, the approach we take can be used to address this issue.

In all of these cases, we take a numerical approach. Our results show that numerical solutions, implemented on desktop computers⁴, of complex incentive mechanism design problems are possible when one uses the appropriate algorithms. This is not as difficult as commonly perceived, but there are some numerical difficulties. The lack of convexity in the incentive constraints implies the possible presence of multiple local optima; we use standard multi-start and other diagnostic techniques to avoid spurious local maxima. The more serious problem is the frequent result that the solution does not satisfy LICQ (linear independence constraint qualification). This makes it difficult for most optimization algorithms to find solutions. We use methods that can deal with moderate failure of LICQ, but many of our problems lie at the frontier of the state of the art in numerical optimization.

In this version of this paper, we just examine a small number of examples that highlight the ideas and present some interesting and suggestive initial results. Far more computation

³Whatever their economic value, tax rules with strong history dependence are currently politically infeasible. One problem is the fear that the accumulation of personal histories in government hands would lead to abuse, and convert the IRS into an agency more like the NKVD or Stasi. Most Americans have a very negative view of such institutions. I suspect that this would not change even if those more familiar with the NKVD, Stasi, and similar organizations, could reassure us that our worries are baseless.

⁴Relative to the computing power available in modern high-performance computing environments, using a desktop computer is like running the marathon after you have shot yourself in both feet. Far more will be possible once we move to modern computing environments.

will be needed to check the robustness of the insights.

We take the discrete-type approach since it makes no extra assumptions about the solution. The continuous-type approach used in Wilson and Tuomala allow them to use powerful PDE methods but only after they have made strong assumptions about the solution. Since we want to avoid unjustified assumptions about the solution, we stay with models with finite types. This turns out to be justified since we find results that violate standard presumptions. In particular, we find cases where the marginal tax rate on the top income is negative. This appears to violate previous results. In particular, Corollary 6.1 in Guesnerie and Seade (1982) finds that marginal tax rates are always nonnegative in their multidimensional model, but they use “Assumption B”, which is essentially a statement that the single-crossing property holds for some ordering of the distinct utility functions. Guesnerie and Seade admit that this assumption will not hold when there are many types with a multidimensional structure. Indeed, we find rather small examples that violate Assumption B and produce negative marginal tax rates.

The examples in this paper show that heterogeneity has substantial impact on optimal tax policy. We also show that computational approaches to optimal tax problems are possible when one uses high-quality optimization software. Further work will examine the robustness of these examples, but the efficiency of our algorithms will allow us to look at a wide variety of specifications for tastes and productivities.

2 Taxation with One-Dimensional Taxpayer Types

We begin with examples of the classical Mirrlees problem. We will then compare them to the optimal tax policies in more heterogeneous models.

We assume $N > 1$ taxpayers. There are two goods: consumption and labour services. Let c_i (l_i) denote taxpayer i ’s consumption (labor supply). Taxpayer i ’s productivity is represented by his wage rate, w_i . We index the taxpayers so that taxpayer i is less productive than taxpayer $i + 1$; therefore,

$$0 < w_1 < \dots < w_N. \quad (1)$$

A type i taxpayer i has pre-tax income equal to

$$y_i := w_i l_i, \quad i = 1, \dots, N \quad (2)$$

Individuals have common utility function over consumption and labor supply: $u : R \times R_+ \rightarrow R$. We assume that u is continuously differentiable, strictly increasing, and strictly concave function with $u_c(0, l) = \infty$, and $\lim_{c \rightarrow \infty} u_c(c, l) = 0$. We next define the implied utility function $U^i : R \times R_+ \rightarrow R$ over income and consumption

$$U^i(c_i, y_i) := u(c_i, y_i/w_i), \quad i = 1, \dots, N. \quad (3)$$

For many preferences (such as quasilinear utility) over income and consumption, higher ability individuals have flatter indifference curves, and indifference curves in income-consumption space of different individuals intersect only once; this defines the *single-crossing property*.

An *allocation* is a vector $a := (y, c)$ where $y := (y_1, \dots, y_N)$ is an *income vector* and $c := (c_1, \dots, c_N)$ is a *consumption vector*. The *social welfare function* $W : R^N \times R_+^N \rightarrow R$ is of the weighted utilitarian form,

$$W(a) := \sum_i \lambda_i U^i(c_i, y_i), \quad (4)$$

We will typically assume that the weights λ_i are positive and nonincreasing in ability. The case of where λ_i equals the population frequency of type i is the utilitarian social welfare function. We take the utilitarian approach. We will assume Output is proportional to total labor supply, which is the only input. Therefore, technology imposes the constraint

$$\sum_i c_i \leq \sum_i y_i \quad (5)$$

We also assume $c_i \geq 0$.

We assume that the government knows the distribution of wages and the common utility function, that it can measure the pretax income of each taxpayer but cannot observe a taxpayer's labor supply nor his wage rate. This corresponds to assuming that each taxpayer's tax payment is a function solely of his labor income. We also assume that all taxpayers face the same tax rules. Therefore, each taxpayer can choose any (y_i, c_i) bundle suggested by the government. The government must choose a schedule such that type i taxpayers will choose the (y_i, c_i) bundle; therefore, the allocation must satisfy the incentive-compatibility or self-selection constraint:

$$U^i(c_i, y_i) \geq U^i(c_j, y_j), \text{ for all } i, j; \quad (6)$$

which states that each person weakly prefers the commodity bundle meant for his type to that chosen by any other type. If the tax policy satisfies (10), then it is common knowledge that an individual with wage w_i will choose (y_i, c_i) from the set $\{(y_1, c_1), \dots, (y_N, c_N)\}$.

The optimal nonlinear income tax problem is equivalent to the following nonlinear optimization problem where the government chooses a set of commodity bundles:

$$\begin{aligned} \max_{y_i, c_i} \quad & \sum_i \lambda_i U^i(c_i, y_i) \\ U^i(c_i, y_i) \geq & U^i(c_j, y_j), \text{ for all } i, j \\ \sum_i c_i \leq & \sum_i y_i \\ c_i \geq & 0, \text{ for all } i \end{aligned} \quad (7)$$

The zero tax commodity bundle for type i is the solution to

$$\max_{l_i} u^i(w_i l_i, l_i)$$

which we denote (c^*, l^*, y^*) .

2.1 Mirrlees cases

We first consider examples of the the following form

$$\begin{aligned} u(c, l) &= \log c - l^{1/\eta+1}/(1/\eta + 1) \\ N &= 5 \\ w_i &\in \{1, 2, 3, 4, 5\} \\ \lambda_i &= 1 \end{aligned}$$

where different values of η correspond to different examples. The zero tax solution is $l_i^* = 1$ and $c_i^* = w_i$. We compute the solutions for $\eta = 1, 1/2, 1/3, 1/5, 1/8$, and report in Tables 1-5 the following:

$$\begin{aligned} y_i, \quad i &= 1, \dots, N, \\ \frac{y_i - c_i}{y_i}, \quad i &= 1, \dots, N, \quad (\text{average tax rate}) \\ 1 - \frac{u_l}{w u_c}, \quad i &= 1, \dots, N, \quad (\text{marginal tax rate}) \\ l_i/l_i^*, \quad i &= 1, \dots, N, \\ c_i/c_i^*, \quad i &= 1, \dots, N, \end{aligned}$$

The pattern of the binding incentive-compatibility constraints is the simple monotonic chain to the left property as expected in nonlinear optimal tax problems in one dimension. Note that the results are as expected. Marginal and average tax rates on the types that pay taxes are moderately high, and increase as the elasticity of labor supply falls. The subsidy rates to the poor fall as we move from the high-elasticity world to the low-elasticity world because the high marginal rates the poor face depress their labor supply much more in the high-elasticity world; remember, all people in each of these economies have the same elasticity.

Table 1. $\eta = 1$					
i	y_i	$\frac{y_i - c_i}{y_i}$	MTR_i	l_i/l_i^*	c_i/c_i^*
1	0.40	-2.87	0.36	0.40	1.56
2	1.31	-0.45	0.38	0.65	0.95
3	2.56	0.03	0.29	0.85	0.83
4	4.01	0.16	0.16	1.00	0.84
5	5.54	0.19	0.00	1.10	0.90

Table 2. $\eta = 1/2$					
i	y_i	$\frac{y_i - c_i}{y_i}$	MTR_i	l_i/l_i^*	c_i/c_i^*
1	0.60	-2.09	0.31	0.60	1.87
2	1.54	-0.39	0.35	0.77	1.08
3	2.69	0.02	0.29	0.89	0.87
4	3.99	0.17	0.17	0.99	0.82
5	5.41	0.21	0.00	1.08	0.85

Table 3. $\eta = 1/3$					
i	y_i	$\frac{y_i - c_i}{y_i}$	MTR_i	l_i/l_i^*	c_i/c_i^*
1	0.70	-1.91	0.28	0.70	2.06
2	1.66	-0.38	0.33	0.83	1.15
3	2.77	0.02	0.29	0.92	0.90
4	3.99	0.17	0.18	0.99	0.82
5	5.33	0.23	0.00	1.06	0.82

Table 4. $\eta = 1/5$					
i	y_i	$\frac{y_i - c_i}{y_i}$	MTR_i	l_i/l_i^*	c_i/c_i^*
1	0.80	-1.84	0.22	0.80	2.29
2	1.78	-0.39	0.29	0.89	1.24
3	2.85	0.02	0.27	0.95	0.93
4	4.01	0.19	0.18	1.00	0.81
5	5.25	0.26	0.00	1.05	0.77

Table 5. $\eta = 1/8$					
i	y_i	$\frac{y_i - c_i}{y_i}$	MTR_i	l_i/l_i^*	c_i/c_i^*
1	0.87	-1.84	0.17	0.87	2.48
2	1.86	-0.41	0.24	0.93	1.31
3	2.91	0.02	0.23	0.97	0.95
4	4.02	0.20	0.16	1.00	0.80
5	5.19	0.28	0.00	1.03	0.73

2.2 Multidimensional Heterogeneity

We next consider models with multidimensional heterogeneity. One kind of multidimensional heterogeneity will be where people differ in both productivity and elasticity of labor supply, η . We will examine that and other types of heterogeneity. More generally, we consider utility functions of the following form

$$u(c, l) = \frac{(c - a)^{1-1/\gamma}}{1 - 1/\gamma} - \psi \frac{l^{1/\eta+1}}{1/\eta + 1}, \quad (8)$$

where a, γ, ψ , and η are possible taxpayer heterogeneities, in addition to wage w . Each term has a natural economic interpretation. The parameter a represents basic “needs”, a minimal level of consumption. A high a implies a higher marginal utility of consumption at any c . The parameter γ represents the elasticity of demand for consumption, whereas the parameter ψ represents the level of distaste for work. The parameter η represents labor supply responsiveness to the wage. This general specification implies a five-dimensional specification of taxpayer types.

For simplicity, we assume there are N types in each heterogeneities. We also define a product index sets $\mathcal{T} := \{1, \dots, N\}$ ⁵. Since wage heterogeneity will always be assumed, let t_1 always represent the wage type. For each $t = (t_1, \dots, t_5) \in \mathcal{T}$, we denote the utility function for type t as⁵ $u^t(c, l)$. The optimal nonlinear income taxation with the utility function (8) is formulated as follows:

$$\begin{aligned} \max_{(y_t, c_t)_{t \in \mathcal{T}}} \quad & \sum_{t \in \mathcal{T}} \lambda_t u^t(c_t, y_t/w_{t_1}) \\ \text{s.t.} \quad & u^t(c_t, y_t/w_{t_1}) - u^t(c_{t'}, y_{t'}/w_{t_1}) \geq 0 \quad \forall t, t' \in \mathcal{T} \\ & \sum_{t \in \mathcal{T}} c_t \leq \sum_{t \in \mathcal{T}} y_t \\ & \sum_{t \in \mathcal{T}} c_t \geq 0, \end{aligned} \quad (9)$$

The problem size is $2|\mathcal{T}|$ variables with $|\mathcal{T}| * (|\mathcal{T}| - 1)$ nonlinear constraints. Notice that the number of constraints is the square of the number of the number of variables. This is a feature of all incentive problems; only in some with simplifying principals like single-crossing are we able to significantly reduce the number of constraints.

2.3 Computational Difficulties with NLPs from Incentive Problems

The nonlinear program (9) is a difficult problem to solve numerically. The objective is concave, but there is a large number of constraints. Furthermore, there is no reason to

⁵To minimize notational complexity, we allow the wage type to show up in the notation u^t even though the wage type is not an argument of u .

believe that the constraints are concave. This creates two problems. First, we cannot ignore the possibility of multiple local optima. We will deal with this in well-known standard ways, and so will not discuss the details of that

The second problem is more challenging: the failure of useful constraint qualifications. Recall the structure of constrained optimization problems. Consider the inequality constrained problem

$$\min f(x) \text{ subject to } c(x) \geq 0,$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$ are assumed smooth. There are many numerical algorithms for solving such problems, but they generally assume that the solution, x^* , satisfies some constraint qualification. Define the set of binding constraints

$$A^* = \{i = 1, 2, \dots, m | c_i(x^*) = 0\}.$$

The linear independence constraint qualification (LICQ) states that the gradients, $\nabla c_i(x^*)$, of the binding constraints $i \in A^*$ are linearly independent. The Mangasarian-Fromovitz constraint qualification (MFCQ) assumes that there is a direction such that $\nabla c_i(x^*)^T d > 0$ for all $i \in A^*$.

In one-dimensional models with single-crossing problems, the LICQ will generally hold. However, multidimensional problems can easily run afoul of the LICQ for one simple reason: if the number of binding constraints exceed the number of constraints, such as in the case of multidimensional pooling, it is impossible for LICQ to hold. Since multidimensional problems are not likely to satisfy a simple pattern of binding incentive constraints. In fact, we found many cases where the LICQ could not hold.

LICQ is a sufficient condition for local convergence of many optimization algorithms, not a necessary condition. However, the failure of LICQ will at least slow down convergence. Our computations converged, but in many cases the number of major iterations needed was unusually large, sometimes in the order of thousands, even for a two-dimensional problem with 25 total types. The failure of the LICQ is probably the sole source of difficulties in solving these nonlinear programs. Other issues, such as scaling, e.g., the range of input parameters w or η , can also cause problems. However, we found LICQ will often fail at solutions of these problems.

Nonlinear programs with constraint qualification failures have been the object of much research in numerical optimization in the past decade. More generally, much progress has been made on a new class of problems called *mathematical programs with equilibrium constraints* (MPECs); see Luo, Pang and Ralph (1996), Outrata, Kocvara and Zowe (1998). One well-known property about MPECs is that the standard CQs fail at every feasible point. We are optimistic that we will be able to apply those methods to allow us to solve larger and more complex problems.

3 Numerical Examples

We now examine economies with multiple dimensions of heterogeneity.

3.1 Wage-Labor Supply Elasticity Heterogeneity

We first consider the case where taxpayers differ in terms of their wage, w , and elasticity of labor supply, η . We assume that $a = 0$, $\gamma = 1$ (log utility), and $\psi = 1$. We consider an optimal nonlinear income tax problem with two-dimensional types of taxpayers:

$$\begin{aligned}
& \max_{y_{ij}, c_{ij}} \quad \sum_{i=1}^N \sum_{j=1}^N \lambda_{ij} u^j(c_{ij}, y_{ij}/w_i) \\
& \text{s.t.} \quad u^j(c_{ij}, y_{ij}/w_i) - u^j(c_{i'j'}, y_{i'j'}/w_i) \geq 0, \text{ for all } (i, j), (i', j') \\
& \quad \sum_{i=1}^N \sum_{j=1}^N c_{ij} \leq \sum_{i=1}^N \sum_{j=1}^N y_{ij} \\
& \quad \sum_{i=1}^N \sum_{j=1}^N c_{ij} \geq 0,
\end{aligned} \tag{10}$$

where $u^j(c, l) = \log c - l^{1/\eta_j+1}/(1/\eta_j + 1)$, and (c_{ij}, y_{ij}) is the allocation for the (i, j) -type taxpayer. The zero tax solution for type (i, j) is $(l_{ij}^*, c_{ij}^*, y_{ij}^*) = (1, w_i, w_i)$. We choose the following parameters: $N = 5$, $w_i \in \{1, 2, 3, 4, 5\}$, $\eta_j \in \{1, 1/2, 1/3, 1/5, 1/8\}$, and $\lambda_{ij} = 1$. We use the zero tax solution (c^*, y^*) as a starting point for the NLP solver, SNOPT.

We report numerical results in Tables 6 and 7 below. There are several points to emphasize.

First, all taxpayers with wage type $w_i = 4$ are pooled. Table 7 shows that there are many more binding constraints than there are variables. This tells us that our worries about failures of LICQ are well-founded.

Second, this example violates the general result in one-dimensional optimal tax theory that marginal tax rates lie between zero and one. Consider taxpayers with wage rate $w_i = 5$. In particular, the taxpayers with low labor supply elasticity η tend to work less and make less income. However, they pay more tax than those taxpayers with high labor supply elasticity η , who have higher income. We also find negative marginal tax rate for the high productivity types with $w = 5$. These results are quite different from the results for one-dimensional type taxpayers, given in Tables 1 – 5 as well as general conclusions in optimal income taxation literature.

Third, all high productivity types are better off in the heterogeneous world. We are not surprised that the low-elasticity high-productivity types are better off since their low elasticity was exploited in the world where all had low labor supply elasticity. In a heterogeneous world, the average elasticity of labor is higher, and so there should be lower taxes on high-productivity workers. The surprise is that the high-elasticity, high-productivity workers

also gain by hiding in a heterogeneous world. The reason, as seen in Table 7, is that these workers do respond to incentives and find that it is tempting to join the pool at $w = 4$. This case also is an example of where the binding constraints are not local, since the highest income type is tempted to pretend to be workers with much less income.

Fourth, heterogeneity reduces redistribution. This is related to point three above, and is highlighted in Figures 1 and 2⁶ where we see that the tax schedule and the average tax rates are almost uniformly lower in the heterogeneous world than in any of the individual Mirrless economies. Hence, redistribution in the heterogeneous world is not just the average of redistribution in the simpler worlds, but instead is substantially less.

⁶In figures 1 and 2, the thin lines represent taxes when all have same elasticity of labor supply. The thick line is the 2D heterogeneity case.

Table 6. $\eta = (1, 1/2, 1/3, 1/5, 1/8)$, $w = (1, 2, 3, 4, 5)$

(i, j)	c_{ij}	y_{ij}	$\Delta TR_{i,j}$	$MTR_{i,j}$	$ATR_{i,j}$	l_{ij}/l_{ij}^*	c_{ij}/c_{ij}^*	Utility	
								Judd-Su	Mirrlees
(1, 1)	1.68	0.42		0.28	-2.92	0.42	1.68	0.4294	.3641
(1, 2)	1.77	0.62	0.51	0.32	-1.86	0.62	1.77	0.4952	.3138
(1, 3)	1.79	0.65	0.54	0.51	-1.75	0.65	1.79	0.5378	.6601
(1, 4)	1.83	0.77	0.66	0.50	-1.37	0.77	1.83	0.5700	.7830
(1, 5)	1.86	0.86	0.62	0.43	-1.16	0.86	1.86	0.5940	.8760
(2, 1)	1.86	0.86		0.60	-1.16	0.43	0.93	0.5308	.3751
(2, 2)	2.03	1.39	0.68	0.50	-0.45	0.69	1.01	0.5973	.6180
(2, 3)	2.07	1.50	0.60	0.56	-0.38	0.75	1.03	0.6512	.7189
(2, 4)	2.16	1.74	0.62	0.46	-0.24	0.87	1.08	0.7006	.8181
(2, 5)	2.20	1.83	0.55	0.46	-0.20	0.91	1.10	0.7413	.9085
(3, 1)	2.20	1.83		0.55	-0.20	0.61	0.73	0.6053	.5496
(3, 2)	2.47	2.49	0.59	0.43	0.00	0.83	0.82	0.7157	.7269
(3, 3)	2.47	2.49		0.53	0.00	0.83	0.82	0.7878	.8158
(3, 4)	2.55	2.68	0.59	0.52	0.04	0.89	0.85	0.8520	.9057
(3, 5)	2.62	2.85	0.54	0.42	0.07	0.95	0.87	0.8965	.9672
(4, 1)	3.36	4.00	0.36	0.16	0.15	1.00	0.84	0.7127	.7090
(4, 2)	3.36	4.00	—	0.16	0.15	1.00	0.84	0.8794	.8664
(4, 3)	3.36	4.00	—	0.15	0.15	1.00	0.84	0.9627	.9402
(4, 4)	3.36	4.00	—	0.15	0.15	1.00	0.84	1.0461	1.0080
(4, 5)	3.36	4.00		0.15	0.15	1.00	0.84	1.1017	1.0476
(5, 5)	4.00	5.14	0.44	0	0.22	1.02	0.80	1.2439	1.1487
(5, 4)	4.11	5.24	-0.10	-0.05	0.21	1.04	0.82	1.1928	1.1331
(5, 3)	4.34	5.43	-0.17	-0.12	0.20	1.08	0.86	1.1188	1.0877
(5, 2)	4.49	5.56	-0.17	-0.11	0.19	1.11	0.89	1.0428	1.0286
(5, 1)	4.87	5.87	-0.22	-0.15	0.17	1.17	0.97	0.8933	.8901

Table 7. Binding IC $[(i, j), (i', j')]$

(i, j)	(i', j')	(i, j)	(i', j')
		(4, 1)	(3, 2), (3, 3), (3, 5), (4, 2), (4, 3), (4, 4), (4, 5)
(1, 2)	(1, 1)	(4, 2)	(4, 1), (4, 3), (4, 4), (4, 5)
(1, 3)	(1, 2)	(4, 3)	(4, 1), (4, 2), (4, 4), (4, 5)
(1, 4)	(1, 3)	(4, 4)	(4, 1), (4, 2), (4, 3), (4, 5)
(1, 5)	(1, 4), (2, 1)	(4, 5)	(4, 1), (4, 2), (4, 3), (4, 4)
(2, 1)	(1, 4), (1, 5)	(5, 1)	(4, 1), (4, 2), (4, 3), (4, 4), (4, 5)
(2, 2)	(1, 5), (2, 1)	(5, 2)	(4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (5, 1)
(2, 3)	(2, 2)	(5, 3)	(5, 2)
(2, 4)	(2, 3)	(5, 4)	(5, 3)
(2, 5)	(2, 4), (3, 1)	(5, 5)	(5, 4)
(3, 1)	(2, 3), (2, 5)		
(3, 2)	(2, 5), (3, 1), (3, 3)		
(3, 3)	(3, 2)		
(3, 4)	(3, 2), (3, 3)		
(3, 5)	(3, 4)		

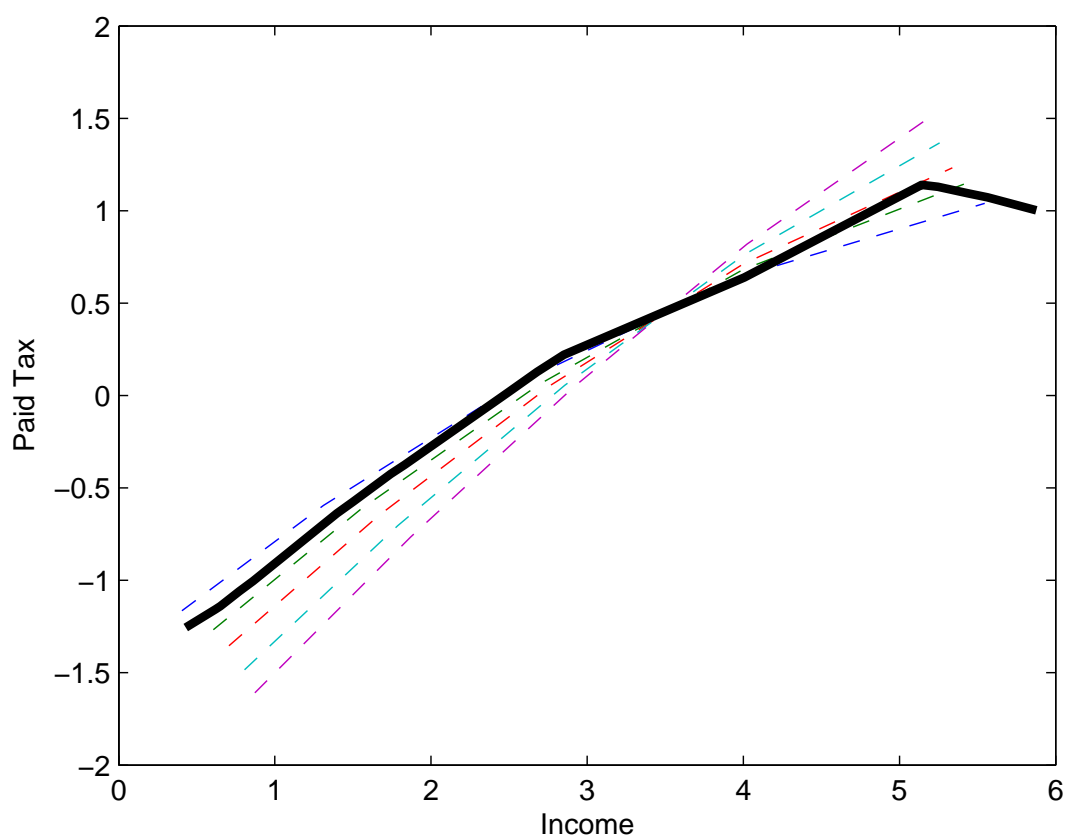


Figure 1: Income v.s. Paid Tax for the two-dimensional heterogeneity example.

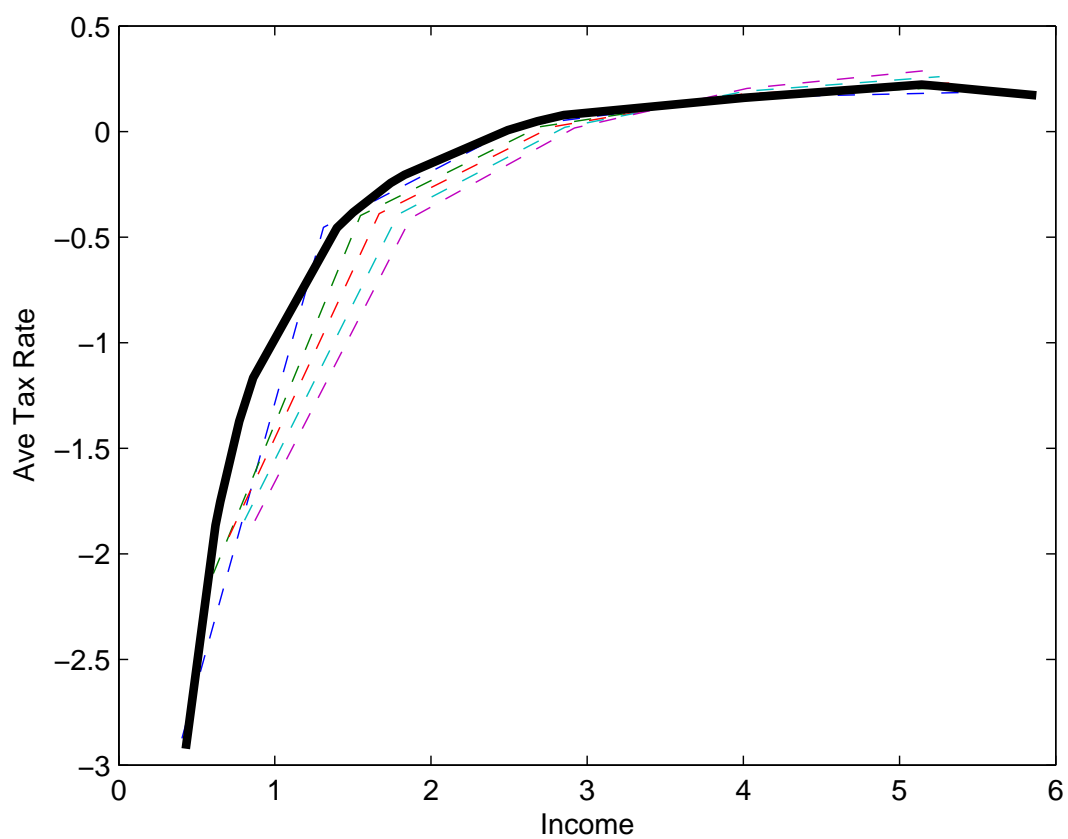


Figure 2: Income v.s. Average Tax Rate for the two-dimensional heterogeneity example.

3.2 Heterogeneity in Wages, Basic Needs, and Labor Supply Elasticity

We next consider a case of three-dimensional heterogeneity.

$$\begin{aligned}
& \max_{(y_{ijk}, c_{ijk})_{i,j,k}} \sum_{i,j,k=1}^N \lambda_{ijk} u^{jk}(c_{ijk}, y_{ijk}/w_i) \\
& \text{s.t.} \quad u^{jk}(c_{ijk}, y_{ijk}/w_i) - u^{jk}(c_{i'j'k'}, y_{i'j'k'}/w_i) \geq 0 \quad \forall (i,j,k), (i',j',k') \\
& \quad \sum_{i,j,k=1}^N c_{ijk} \leq \sum_{i,j,k=1}^N y_{ijk} \\
& \quad \sum_{i,j,k=1}^N c_{ijk} \geq 0,
\end{aligned} \tag{11}$$

where $u^{jk}(c, l) = \log(c - a_k) - l^{\eta_j+1}/(\eta_j + 1)$, and (c_{ijk}, y_{ijk}) is the allocation for the (i, j, k) -type taxpayer. We choose the following parameters: $N = 3$, $w_i \in \{2, 3, 4\}$, , and $\lambda_{ijk} = 1/(N^3)$. We use the zero tax solution (c^*, y^*) as a starting point for the NLP solver, SNOPT, and compute the solutions for $\eta_j \in \{1/2, 1, 2\}$, and $a_k \in \{0, 1, 2\}$.

Figures 3 and 4 illustrate the solution. The dotted lines are average tax rates if wage is only heterogeneity. Thin lines represent cases of two-dimensional heterogeneity, both $w - \eta$ heterogeneity and $w - a$ heterogeneity. The solid line is one economy with 3D heterogeneity - taxpayers differ in wages, basic needs, and labor supply elasticity. The patterns are clear. The most redistributive economies are those where wage is the only heterogeneity. As we add heterogeneity in either a or η , redistribution is less. The final case where there is heterogeneity in wages, basic needs, and elasticity, has the least redistribution.

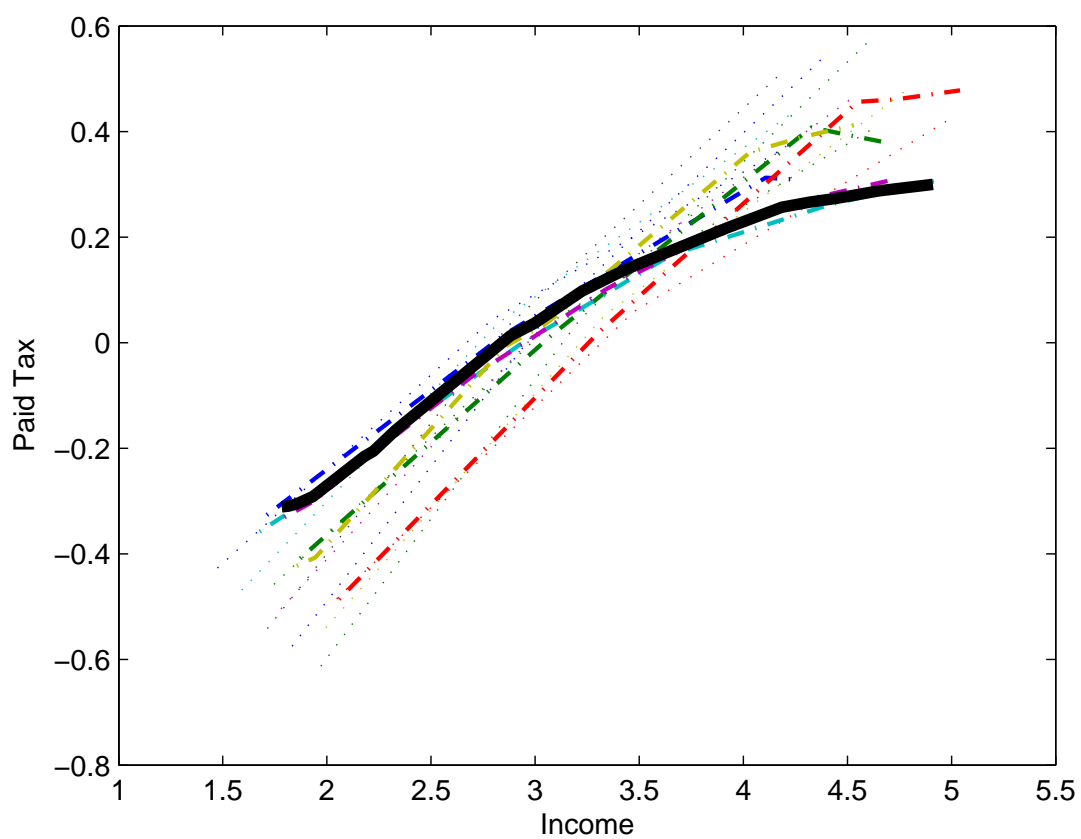


Figure 3: Income v.s. Paid Tax for the three-dimensional heterogeneity example.

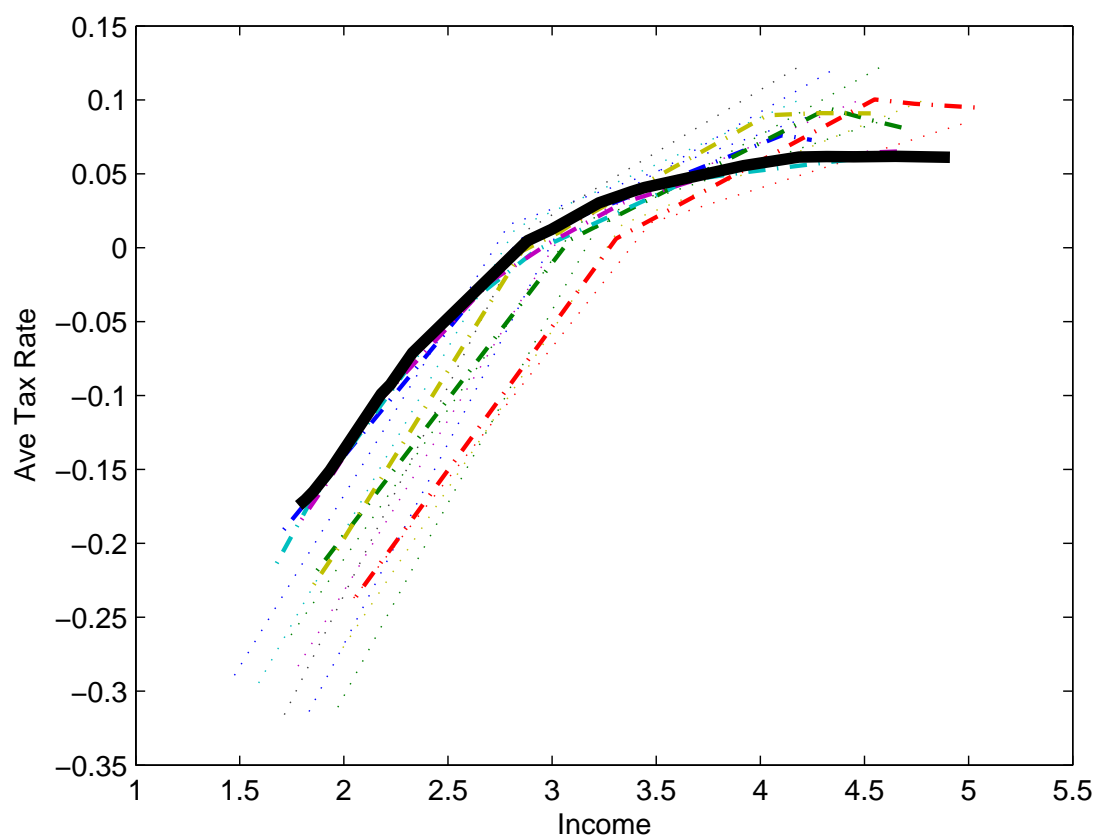


Figure 4: Income v.s. Average Tax Rate for the three-dimensional heterogeneity example.

3.3 Wage-Age Heterogeneity

Our final example is a case where people differ by productivity and age. If we apply the Mirrlees approach to dynamic models, the result will be age-dependent income-consumption allocations. This is not the case for actual tax systems. Instead, tax liabilities generally depend only on the current year's income. We will examine the optimal tax policy under the assumption that tax liabilities depend only on current income. Implicit in this are many limitations on the government policy. In particular, we will not allow the government to use current consumption nor current assets in determining taxes.

The way to do this is to impose more constraints on the income-consumption bundles. In our example, we will assume that there are three types of workers who live for three periods. From the point of view of a government that cannot see age, there are nine types of people and the optimal tax policy will offer nine consumption-income bundles. Each individual will choose a life-cycle pattern from that menu. There are 729 possible paths to choose from. The incentive compatibility constraint will say that each individual will freely choose that one life-cycle path that the government has prescribed for him. Since there are three types of people, there are 2187 incentive constraints.

We assume that there are three types of workers, and that each type has deterministic pattern of life-cycle wages. We use the following numerical specification:

Wage History			
	Period		
Type	t_1	t_2	t_3
1	1	3	2
2	2	4	4
3	2	5	4
4	3	5	6

Kocherlakota, and Golosov et al. have applied the Mirrless approach to this model, where each individual announces his type when born, and then is given a life-cycle pattern of income and taxes that he must follow. We argued above that this kind of history dependence does not describe existing tax policies. Therefore, we look at three cases: Mirrlees taxation, an optimal (nonlinear) tax on workers that are consistent with the ability of each worker to, and the optimal linear tax policy.

We first see that as we add more restrictions, going from Mirrlees to optimal age-independent policy to linear tax, the amount of redistribution falls and the tax burden on the high productivity type falls. Table 8 shows the life-cycle results for each type. Consider type 1 agents. They work much more as we move from Mirrlees to nonlinear ageless taxation, and work a bit more when we go to optimal linear taxation. Similar, both the subsidy and utility fall. On the other hand, the high productivity type agents gain in utility terms. They earn less, but pay less taxes.

Table 8: Aggregate Outputs for Each Type

Type	Total Income			Total Tax Paid			Total Utility		
	Mirrlees	Nonlin.	Linear	Mirr.	Nlin.	Linear	Mirr.	Nlin.	Linear
1	4.72	5.43	5.65	-2.40	-1.36	-0.96	1.79	1.40	1.23
2	9.60	10.02	9.70	-0.03	0.07	-0.07	2.22	2.20	2.23
3	11.88	11.19	10.83	0.51	0.36	0.18	2.43	2.46	2.49
4	15.48	14.35	13.90	1.91	0.93	0.85	2.82	3.01	3.03

Table 9 gives a more detailed description of the three tax policies. One notable feature is that in the age-free case, no person paying taxpayers (as opposed to receiving a subsidy) is ever facing a zero effective tax rate. The highest income is earned by the highest productivity type at age three and faces a tax rate of 0.01. Perhaps this is numerical error, but the diagnostics we have performed so far indicates that this is not zero. In any case, the high productivity type faces nonzero marginal tax rates at ages one and two. This points to an interesting feature of this ageless problem. In this case, each taxpayer is not only tempted by the possibility of pretending to be one of the other types, but also is possibly tempted to be someone who is a different type at different ages. The enormous increase in incentive constraints makes it less likely that some real person, as opposed to some artificial type, faces a zero tax rate.

Table 9: Life-cycle patterns of income, taxes, and MTR

		Mirrlees approach			Nonlinear tax			Linear tax		
Type	age	y	Tax	MTR	y	Tax	MTR	y	Tax	MTR
1	1	0.31	-0.79	0.25	0.32	-1.01	0.25	0.42	-0.64	0.07
1	2	3.15	-0.79	0.16	3.55	0.24	0.10	3.46	0.02	0.15
1	3	1.25	-0.79	0.25	1.54	-0.59	0.12	1.75	-0.34	0.03
2	1	1.05	-0.01	0.15	1.05	-0.73	0.12	1.12	-0.48	0.08
2	2	4.32	-0.01	0.13	4.48	0.39	0.07	4.28	0.20	0.12
2	3	4.22	-0.01	0.15	4.48	0.39	0.07	4.28	0.20	0.12
3	1	1.05	0.17	0.00	1.02	-0.73	0.07	1.12	-0.48	0.00
3	2	6.59	0.17	0.00	6.29	0.79	0.09	6.10	0.60	0.13
3	3	4.22	0.17	0.00	3.85	0.29	0.12	3.59	0.05	0.20
4	1	1.99	0.63	0.00	1.54	-0.59	0.23	1.75	-0.34	0.15
4	2	5.52	0.63	0.00	4.90	0.47	0.12	4.83	0.32	0.15
4	3	7.96	0.63	0.00	7.90	1.05	0.01	7.30	0.87	0.11

4 Conclusions

We have examined some simple cases of optimal taxation in economies with multiple dimensions of heterogeneities. These examples show that many results from the basic one-dimensional Mirrlees model no longer hold. The examples also indicate that redistribution is less in economies with multidimensional heterogeneity, probably because income is a noisier signal of a taxpayer's type. This last result is very provisional until we do a robust examination of this issue across alternative parameter values for tastes.

The other main point is that these problems are difficult to solve numerically, but they can be solved if one recognizes the critical numerical difficulties and uses the appropriate software.

5 References

Atkinson, A.B. and J.E. Stiglitz (1976), The Design of Tax Structure: Direct versus Indirect Taxation, *Journal of Public Economics* 6, 55-75.

Blomquist, S., Christiansen, V. (2004). Taxation and Heterogeneous Preferences. Uppsala Universitet Working Paper 9, Department of Economics.

Blomquist, S. and L. Micheletto (2003). Age Related Optimal Income Taxation, Working Paper 2003:7, Department of Economics, Uppsala University.

Boadway R.; Marchand M.; Pestieau P., del Mar Racionero M. (2002): Optimal Redistribution with Heterogeneous Preferences for Leisure, *Journal of Public Economic Theory* 4 (no. 4), 475-498.

Boadway, R., Cuff, K., Marchand, M., 2000. Optimal income taxation with quasi-linear preferences revisited. *Journal of Public Economic Theory* 2, 435– 460.

Brito, D.L., Oakland, W., 1977. Some properties of the optimal income-tax. *International Economic Review* 18, 407–423.

Cuff, K. (2000). Optimality of Workfare with Heterogeneous Preferences. *Canadian Journal of Economics* 33, 149-174.

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

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

Ebert, U. (1988). Optimal Income Taxation: On the Case of Two-Dimensional Populations. Discussion Paper A169.

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

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

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

Laffont, J.-J., Maskin, E., Rochet, J.-C., 1987. Optimal nonlinear pricing with two-dimensional characteristics. In: Groves, T., Radner, R., Reiter, S. (Eds.), *Information, Incentives, and Economic Mechanisms*. University of Minnesota Press, Minneapolis, pp. 256–266.

Luo, Z.-Q., Pang, J.-S., Ralph, D. (1996): *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge, UK.

Marchand, M., Pestieau, P. (2003). Optimal redistribution when workers are indistinguishable. *Canadian Journal of Economics*, 36, 911–922.

McAfee, R.P., McMillan, J., 1988. Multidimensional incentive compatibility and mechanism design. *Journal of Economic Theory* 46, 335–354.

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

Mirrlees, J.A., 1976. Optimum tax theory: A synthesis. *Journal of Public Economics* 7, 327–358.

Mirrlees, J.A., 1997. Optimal marginal taxes at low incomes. Unpublished manuscript, Faculty of Economics, University of Cambridge.

Mirrlees, J.A. (1986). The theory of optimal taxation. In Arrow, K.J. and Intriligator, M.D. (eds.), *Handbook of Mathematical Economics*. Vol. III, Elsevier Science Publishers, B.V., Amsterdam.

Outrata, J., Kocvara, M., Zowe, J. (1998): *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory, Applications, and Numerical Results*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Rochet, J.-C., Stole, L.A., 2000. The economics of multidimensional screening. In: Dewatripont, M., Hansen, L., Turnovsky, S. (Eds.), *Advances in Economics and Econometrics: Eighth World Congress*. Cambridge University Press, Cambridge.

Sandmo, A. (1993) : Optimal Redistribution when Tastes Differ, *Finanz Archiv* 50(2), 149–163.

Sandmo, A., 1998. Redistribution and the marginal cost of public funds. *Journal of Public Economics* 70, 365–382.

Seade, J.K., 1977. On the shape of optimal tax schedules. *Journal of Public Economics* 7, 203–235.

Shapiro, J., 1999. Income maintenance programs and multidimensional screening. Unpublished manuscript, Department of Economics, Princeton University.

Sheshinski, E., 1971. On the theory of optimal income taxation. Discussion Paper No. 172, Harvard Institute for Economic Research, Harvard University.

Stiglitz, J.E. (1982). Self-Selection and Pareto Efficient Taxation. *Journal of Public Economics* 17, 213–240.

Tarkianen, R. and M. Tuomala (1999). Optimal Nonlinear Income Taxation with a Two-Dimensional Population; A Computational Approach. *Computational Economics* 13, 1-16.

Tuomala, M. (1990). Optimal Income Tax and Redistribution. Clarendon Press, Oxford, U.K.

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

Wilson, R. (1993). *Nonlinear Pricing*. Oxford University Press, New York.

Wilson, R. (1995). Nonlinear pricing and mechanism design. In Amman, H., Kendrick, D. and Rust, J. (eds.), *Handbook of Computational Economics*. Vol. 1, Elsevier Science Publishers B.V., Amsterdam.