Optimal Taxes and Transfers under Partial Information

By Stefan Homburg and Tim Lohse*

DISCUSSION PAPER NO. 298
MAY 2004
ISSN: 0949-9962

JEL-CLASSIFICATION: H21, I38
KEYWORDS: Optimal Taxation, Employment, Poverty, Welfare.

Revised version published in:

Jahrbücher für Nationalökonomie und Statistik **225** (2005), pp. 622 ff.

* We wish to thank two anonymous referees and the participants of the CESifo Area Conference on Employment and Social Protection, June 2004 in Munich for helpful remarks. Of course, we take sole responsibility for all remaining shortcomings.

1. INTRODUCTION

Until 2004, German long-term unemployed received a tax-financed benefit (*Arbeitslosenhilfe*) which exceeded social assistance for the disabled (*Sozialhilfe*). This has been changed by the recent reform known as "Hartz IV": Effective from 2005, long-term unemployed on the one hand (who are no more entitled to unemployment insurance) and the disabled on the other hand receive one and the same benefit, i.e. 345 Euro in cash plus accommodation and heating. During the reform debate, it was also suggested to reduce unemployment benefits below the level of social assistance. These observations raise the following question: Should unemployment benefits

- exceed social assistance, as until 2004, or
- fall short of social assistance, as proposed by Academic Advisory Council (2002), or
- should both benefits be equal, as is true from 2005 on?

The question rests on an important implicit assumption, namely, that the government is in a position to condition transfers upon respective working capabilities. Such a premise is also implicit in the new labor ethic which holds that "all who are able to work, should work" and that social assistance should be confined to the truly needy. Many pure theorists, however, would reject the latter postulate as naive because Mirrlees (1971) pointed out early that it can be socially optimal to have some persons who do not work. For this sort of unemployment, Seade (1977, p. 215) coined the term "bunching at the bottom". In evaluating social reforms, a significant part of the literature has used Mirrlees (1971) optimal tax model, interpreting transfers as negative taxes. The emerging *standard model*, as we will call it, does not shed any light on the above question since it assumes perfect ignorance about individual productivities on the side of the government.

This paper's objective is to develop a model which deviates from the standard model in one respect: We assume *partial information* in the sense that the government can distinguish the disabled from the productive, but cannot distinguish among the different productive types. Such an assumption appears sensible because the government can in fact use medical reports, for instance, in order to assess individual productivities. Anyway, the assumption is indispensable if one wishes to analyze the optimal relative treatment of unemployed and disabled persons. Without partial information, the new labor ethics were an empty concept.

The paper is organized as follows: In section 2 we review the standard model, leaving all assumptions but one unchanged. Introducing partial information yields an interesting property of optimal tax-transfer schemes. Section 3 illustrates the basic proposition by means of some numerical examples and warns that unemployment may be optimal even under partial information. Section 4 adds some caveats and concludes.

2. THE MODEL

Consider the finite variant of the standard optimal taxation model.[1] There are several types whose exogenous wage rates and fractions are denoted as w^h and $f^h > 0$, respectively, for $h = 0, 1 \ldots H$, where $H > 1$. We assume $0 = w^0 < w^1 < \ldots < w^H$, so that type 0 persons are *disabled*, whereas all other types are productive. A person consuming c^h and earning gross labor income y^h enjoys utility $u(c^h, y^h/w^h)$, where y^h/w^h represents effort (and y^0/w^0 vanishes by convention). The utility function is strictly increasing in consumption, strictly decreasing in effort and strictly concave. Moreover, we assume that its cross derivative vanishes and that all consumption quantities are strictly positive at the optimum. The government's decision problem reads:

$$
\begin{aligned}
&\max_{(c^h, y^h)_{h=0\ldots H}}! \quad EU = \sum_{h=0}^{H} u(c^h, \frac{y^h}{w^h})\, f^h \\
(1) \qquad & s.t. \quad i) \quad \sum_{h=0}^{H} (y^h - c^h)\, f^h \geq g, \\
& \qquad\;\; ii) \quad u(c^k, \frac{y^k}{w^k}) \geq u(c^h, \frac{y^h}{w^k}) \quad \text{for all } k, h > 0.
\end{aligned}
$$

Thus, the government maximizes expected utility of a person choosing a tax-transfer scheme from behind a veil of ignorance, subject to the budget constraint i), where g represents an exogenous revenue requirement,[2] and subject to the self-selection constraints ii) which ensure that no person can make himself better off by mimicking somebody else. The difference between the standard model and the present one is that the self-selection constraints refer to types k, $h > 0$ only, rather than to all types. This means that the productive cannot mimic the disabled, and vice versa. Solutions of this problem generate truthful reporting of abilities. Optimal taxes or transfers follow implicitly as $T(y^h) = y^h - c^h$, and the discrete marginal tax rates are defined, for all $y^h \neq y^{h-1}$, as

$$
(2) \qquad m^h = \frac{T(y^h) - T(y^{h-1})}{y^h - y^{h-1}}.
$$

The only general result of the standard model states that the marginal tax rates are strictly less than one. This can easily be seen: By the very definition of m^h, a marginal tax rate beyond one hundred per cent implies that the differences $c^h - c^{h-1}$ and $y^h - y^{h-1}$ have opposite signs, but no person $h > 1$ would accept a pair containing less consumption and more income (effort) as compared to another. This, as well as the two following features, is also true under partial information:

1 A recent textbook version is presented in Homburg (2005).

2 The revenue requirement may be positive in case of provision of public goods, or zero in case of a purely redistributive program. Its numerical value does not affect the qualitative results.

(3) $$y^h \geq y^{h-1} \quad \text{and} \quad c^h \geq c^{h-1} \quad \text{and} \quad u(c^h, \frac{y^h}{w^h}) = u(c^{h-1}, \frac{y^{h-1}}{w^h}) \quad \text{for all } h > 1\,.$$

The two inequalities state that income and consumption increase weakly in productivity. Bunching at strictly positive incomes is possible. However, as is well-known, bunching at the top will never occur. The right-hand equality states that the downward adjacent self-selection constraints are binding at an optimum, which is due to the fact that the government wishes to redistribute from top to bottom.[3] Together with the monotonicity, this "chain property" implies that all remaining self-selection constraints are automatically satisfied and hence can be neglected without loss of generality. It should be clear that the government's budget constraint also holds with equality at an optimum.

As in the standard approach, we can now maximize the objective function in (1) subject to the constraint (i) in (1) and the constraints (3). Comparing the two approaches immediately reveals that the use of partial information pays: Referring to type 1 persons as the *productive poor*, a constraint preventing the productive poor from mimicking the disabled is missing under partial information. As this constraint binds in the standard case, the outcome of the present model dominates the second-best optimum. The social value of partial information has already been pointed out by Akerlof (1978). Yet, owing to the intricacy of the Mirrlees model, neither Akerlof nor the literature thereafter (e.g. Immonen, Kanbur, Keen and Tuomala 1998) could derive any general results. The only insights in the field stem from examples and simulations. Using a different technique, we are able to identify an interesting property of optimal tax-transfer schemes under partial information.

Proposition: If the government solves the standard optimal tax problem, but can distinguish the disabled from the productive, then any optimum satisfies $c^0 > c^1$.

Proof: Since the government's budget constraint and the downward adjacent self-selection constraints hold as equalities and are linearly independent (Homburg 2003), we can infer necessary first-order conditions using the Lagrangean

(4) $$\mathcal{L} = \sum_{h=0}^{H} u(c^h, \frac{y^h}{w^h}) f^h + \lambda \sum_{h=0}^{H} \left(y^h - c^h - g\right) f^h + \sum_{h=2}^{H} \mu^h \left(u(c^h, \frac{y^h}{w^h}) - u(c^{h-1}, \frac{y^{h-1}}{w^h}) \right).$$

Differentiating with respect to consumption yields:

(5) $$\frac{\partial \mathcal{L}}{\partial c^0} = \frac{\partial u}{\partial c^0} f^0 - \lambda f^0 = 0\,,$$

(6) $$\frac{\partial \mathcal{L}}{\partial c^h} = \frac{\partial u}{\partial c^h} f^h - \lambda f^h + \mu^h \frac{\partial u}{\partial c^h} - \mu^{h+1} \frac{\partial u}{\partial c^h} = 0, \quad h > 0\,,$$

where $\mu^1 = \mu^{H+1} = 0$. Rearranging terms gives

3 Optimal policies redistribute from top to bottom (and not the other way round) if one assumes agent monotonicity and non-inferiority of leisure, cf. Homburg (2001). Both assumptions are fulfilled in case of a vanishing cross derivative.

(7) $$f^0 = \frac{\lambda f^0}{\partial u/\partial c^0},$$

(8) $$f^h + \mu^h - \mu^{h+1} = \frac{\lambda f^h}{\partial u/\partial c^h}, \quad h > 0.$$

Adding these equations over all h, the variables f^h sum up to one, the variables μ^h cancel out each other, and solving for λ shows that the shadow price of the budget constraint is the *harmonic mean* of the marginal utilities of consumption:

(9) $$\lambda = \frac{1}{\sum_{h=0}^{H} \frac{f^h}{\partial u/\partial c^h}}.$$

From (5) we have $\partial u/\partial c^0 = \lambda$ at the optimum, so that the marginal utility of the disabled equals the average marginal utility of all productive persons. However, $\partial u/\partial c^1 > \lambda$ because consumption increases in h for all productive persons, and there is no bunching at the top. Combined with the assumption of decreasing marginal utility, this implies $c^0 > c^1$. ■

The proposition states that the disabled should have more consumption than the productive poor. This may appear surprising but has a simple explanation.[4] Partial information separates the population into two subsets of persons who cannot mimic one another. Given this separation, resources should be used so as to equalize the marginal utilities of the persons belonging to the two subsets, respectively, implying that the marginal utility of the disabled should equal the harmonic mean of the productive persons' marginal utilities. Now, type 1 persons, the productive poor, enjoy the lowest consumption, hence the highest marginal utility of consumption among the productive. Since the marginal utility of consumption of type 0 equals the average, it necessarily falls short of the marginal utility of consumption of type 1. From the law of diminishing marginal utility we immediately obtain $c^0 > c^1$.

As an alternative explanation, increasing c^0 and c^1 marginally induces the same resource costs per person, but increasing c^1 entails an additional incentive cost, in that it tightens the self-selection constraint which prevents type 2 persons from mimicking the productive poor. Owing to the chain property, increasing c^1 makes it necessary to increase consumption, or to reduce effort, of all persons with higher productivities, so that the rich become better off at the expense of the poor.

Under the assumptions made, our proposition holds perfectly generally, irrespective of whether or not the optimal policy eliminates bunching at the bottom. But if it does, the productive poor have more gross income and less consumption as compared with the

4 A similarly surprising result has been derived by Corneo (2003, p. 146) in a somewhat different context: Corneo shows that persons with a higher unemployment risk will enjoy higher consumption if unemployment benefits are set optimally.

disabled, and applying definition (2) of a marginal tax rate mechanically, the latter exceeds one hundred per cent. Of course, this irritating finding is not in accordance with the usual notion of a "marginal tax rate". It simply derives from the fact that marginal tax rates have no disincentive effects at the very bottom of the income distribution, if the government can distinguish the disabled from the productive poor. Moreover, the "marginal tax rate" m^1 exceeds one hundred per cent only if the productive poor actually work. This is not necessarily the case because, contrary to a further widespread view, treating the disabled and the productive differently does not suffice to rule out bunching at the bottom, as we will show. It may still be optimal to have some productive persons idle and then, following the usual logic, the marginal tax rate at the bottom will fall short of one hundred per cent.

3. An Example

Consider an economy with five types, a utility function $u(c^h, y^h/w^h) = \ln c^h + \ln (500 - y^h/w^h)$, a uniform productivity distribution, and per-capita revenue $g = 100$. Table 1 depicts the standard second-best optimum. Marginal tax rates are high at the bottom of the income scale and decrease monotonically. The local marginal rate – the difference between the marginal rate of substitution and the marginal rate of transformation – vanishes at the very top of the income scale, as is well known, but the discrete marginal rate defined in (2) is positive (and converges to zero as the income distribution grows dense). More importantly, there is induced unemployment since the productive poor are bunched together with the disabled at zero income.

w	***c***	***y***	***m***	***u·f***
0	543	0	--	2,50
2	543	0	--	2,50
4	728	509	64%	2,50
8	1433	2.226	59%	2,53
10	1994	3.006	28%	2,58
Σ				12,62

Table 1: Standard Optimum.

Table 2 shows the corresponding optimum under partial information. The government's ability to distinguish between the disabled and the productive induces it to let the former enjoy average consumption (this is true only for log-linear utility functions). As a consequence, expected utility rises from 12,62 to 12,67, but the policy change does not present a Pareto-improvement: the utilities of the productive poor and of the middle classes have been diminished. Marginal tax rates are roughly as high as in the standard case, and have been supplemented by a horrific "marginal tax rate" of 516 per cent at the bottom of the income scale. The policy's aim to get the productive to work has been reached in that induced unemployment does no longer exist.

w	***c***	***y***	***m***	***u·f***
0	1137	0	--	2,65
2	511	151	516%	2,46
4	706	661	62%	2,47
8	1394	2.309	58%	2,52
10	1937	3.063	28%	2,57
Σ				12,67

Table 2: Optimum under Partial Information.

The final Table 3 warns, however, that optimal taxes do not always rule out bunching. The table has been constructed using the same assumptions as above, the only difference being that the smallest positive wage rate has been reduced from two to one. Again, the productive poor are bunched together with the disabled at income zero and obtain less consumption, because their consumption, in the sense pointed out above, is more costly. Therefore, even if the government can separate the disabled from the productive, pooling some of them at zero income – rather than separating them with respect to working hours – may still be optimal. It is not by accident that the marginal tax rate at the bottom falls short of one hundred per cent: Whenever an index $\hat{h} > 1$ such that $y^h = 0$ for all $h < \hat{h}$ and $y^h > 0$ for all $h \geq \hat{h}$ exists, the marginal tax rate at the smallest positive income is given by $m^{\hat{h}}$ and is below one hundred per cent, since type $\hat{h}$ must be prevented from mimicking his left-hand neighbor.

w	***c***	***y***	***m***	***u·f***
0	1113	0	--	2,65
1	463	0	--	2,47
4	693	665	65%	2,47
8	1377	2.322	59%	2,51
10	1920	3.080	28%	2,56
Σ				12,66

Table 3: Bunching under Partial Information.

4. Conclusion and Evaluation

This paper has analyzed the structure of optimal tax-transfer-systems in case of partial information about people's abilities. The significance of such an informational assumption was already pointed out by Akerlof (1978), but neither he nor the literature following him could derive any general result. This is due to the complexity of the Mirrlees tax model which yields very few insights, unless one makes very stringent assumptions both on preferences and the productivity distribution. Yet, our main result shows that the government should grant higher transfers to the disabled than to the productive unemployed. This finding may explain why persons with certain characteristics signaling disability – such as age, illness, or dependent children – can expect better treatment in many welfare states as compared to unemployed persons without such characteristics. The intuition behind the result is not that idle productive persons are more expensive than idle disabled persons. Rather, increasing consumption of the productive entails an additional incentive cost, which is absent when increasing consumption of the disabled.

An alternative suggestion of how to reform the welfare state was provided by the Academic Advisory Council at the German Federal Ministry of Economics and Labor (2002). The Council suggested i) leaving the social assistance unchanged in case of disabled persons, ii) cutting it substantially in case of productive persons and iii) lowering the marginal tax rates on low incomes in order to induce less productive types to join the labor force. To summarize, unemployed persons should receive different levels of transfers, depending on whether or not they are disabled, and marginal tax rates on low incomes should be reduced. Our model sustains the recommendation of the Academic Advisory Council that transfers to the productive poor should be *lower* than the social assistance for the disabled. *Identical* transfers for both groups turned out to be suboptimal, as the above proposition elucidates. In this sense, Germany has missed the optimal reform. However, the model does not support the Council's further recommendation which aims at eliminating bunching at the bottom through low marginal tax rates. Bunching at the bottom, i.e. unemployment of productive individuals, may still be optimal under partial information, and optimal marginal tax rates at the bottom may still be high in order to prevent the rich from mimicking the poor.

In evaluating the model, three caveats come to mind.[5] Firstly, we have assumed that the disabled can be distinguished from the productive perfectly and without costs. If the distinction involves administrative costs, other aspects become important which have been analyzed by Boadway, Marceau and Sato (1999). Secondly, our discrete setting conveys a clear-cut conceptual distinction between the disabled and the productive. Considering a limiting process where the productivity distribution becomes dense will obscure the borderline because, then, persons with an arbitrarily low productivity are likely to exist in the right-hand neighborhood of the disabled. The distinction between the disabled and the productive is no more a positive issue but a normative one: the government must decide which persons qualify as disabled. Nevertheless, our above analysis can easily be generalized to cover this case: Suppose the government fixes a

5 Moreover, all objections against the standard approach still apply. Here, we only discuss caveats that trace back to our new assumption of partial information.

threshold $\hat{h}$ such that all types with productivities below $w^{\hat{h}}$, which can be identified by hypothesis, qualify as disabled. At the optimum, these types obtain the common bundle (c^0, 0), and type $\hat{h}$ obtains ($c^{\hat{h}}$, $y^{\hat{h}}$), where $c^{\hat{h}} < c^0$. Again, the "marginal tax rate" at the very bottom of the income distribution will be positive and will exceed one hundred per cent if and only if there is no bunching at the bottom.

Thirdly, and most importantly, productivities have been assumed exogenous. This premise seems innocuous in the standard model – which can easily be extended to include educational decisions – but not in case of partial information. Given the harsh treatment of the productive poor, these may be apt to become disabled. Strategies as drinking or drug addiction come into mind. Perhaps in a generalized model, which takes account of endogenous productivities, a lower tax pressure at the bottom would be optimal.

References

Academic Advisory Council (2002) *Reform des Sozialstaats für mehr Beschäftigung im Bereich gering qualifizierter Arbeit*. Bundesministerium für Wirtschaft, Dokumentation Nr. 512.

Akerlof, G. A. (1978) The Economics of "Tagging" as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning, *American Economic Review* **68** (1), pp. 8–19.

Boadway, R., N. Marceau and M. Sato (1999) Agency and the design of welfare systems. *Journal of Public Economics* **73** (1), pp. 1–30.

Corneo, G. (2003) *Öffentliche Finanzen: Ausgabenpolitik*. Tübingen: Mohr Siebeck.

Hartz-Commission (2002) *Moderne Dienstleistungen am Arbeitsmarkt*. Berlin: mimeo.

Homburg, St. (2001) The Optimal Income Tax: Restatement and Extensions. *Finanzarchiv* **58** (4), pp. 363–395.

Homburg, St. (2003) An Axiomatic Proof of Mirrlees' Formula. *Public Finance/Finances Publiques* **53** (3-4), 1998/2003, pp. 285–295.

Homburg, St. (2005) *Allgemeine Steuerlehre*. 4th edition München : Vahlen.

Immonen, R., R. Kanbur, M. Keen and M. Tuomala (1998) Tagging and taxing: The Optimal Use of Categorical and Income Information in Designing Tax/transfer Schemes. *Economica* **65**, pp. 179–192.

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

Seade, J. K. (1977) On the Shape of Optimal Tax Schedules. *Journal of Public Economics* **7**, pp. 203–235.

Corresponding Author: Tim Lohse
Chair of Public Finance, School of Economics and Management, University of Hannover, Königsworther Platz 1, D–30167 Hannover, Germany. www.fiwi.uni-hannover.de.
Telephone: (+49) (511) 762-5176, Fax: (+49) (511) 762-5656, Email: Lohse@fiwi.uni-hannover.de.