A generalization of the Atkinson–Stiglitz (1976) theorem on the undesirability of nonuniform excise taxation

Martin F. Hellwig

Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn, Germany

ARTICLE INFO

Article history:
Received 23 July 2009
Received in revised form 16 September 2009
Accepted 23 April 2010
Available online 9 May 2010

Keywords:
Optimal taxation
Atkinson–Stiglitz Theorem

JEL classification:
H21

ABSTRACT

The Atkinson–Stiglitz (1976) theorem on the undesirability of nonuniform excise taxation when all agents have homogeneous, separable preferences is extended to allow for nonseparability with respect to endogenous variables that will be subject to distortions.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Atkinson and Stiglitz (1976) have shown that, under certain homogeneity and separability assumptions on preferences, an optimal system of taxes for public-sector funding or redistribution relies on direct taxation only. Kaplow (2006) and Laroque (2005) have extended this result to show that, under the assumptions of Atkinson and Stiglitz, any feasible, incentive-compatible allocation is Pareto-dominated by an allocation that can be implemented by direct taxation, without recourse to distortionary indirect taxes. This note uses the argument of Kaplow (2006) and Laroque (2005) to show that the result still holds under a weaker separability assumption on preferences.1

2. The setup

I consider a model in which agents care about outcomes that are specified as n-dimensional vectors \((x_1, \ldots, x_n) \in \mathbb{R}^n\). An agent's preferences over outcomes depend on his "type", a parameter \(t\) that lies in an interval \([\ell_0, t_1]\). His preferences are represented by a utility function \(u\), which is assumed to take the separable form

\[
u(x_1, \ldots, x_n; t) = v(\varphi(x_1, \ldots, x_n), x_n; t);
\]

where \(v\) and \(\varphi\) are continuously differentiable functions, with derivatives satisfying \(v_{x_i} > 0\) and \(\varphi_{x_i} > 0\) for all \(i\). The function \(v\) is also assumed to satisfy the single-crossing condition that \(v_{x_i}\), the marginal rate of substitution between \(\varphi\) and \(x_n\), is everywhere increasing in \(t\).

In Atkinson and Stiglitz (1976), Kaplow (2006), and Laroque (2005), the outcome variables \(x_1, \ldots, x_n \geq 0\) would be consumption levels of different goods, \(x_n = -y\), where \(y\) is the level of output that the agent produces, and \(t\) a productivity parameter, with the interpretation that \(1 = \frac{x_n}{y}\) is the amount of labour the agent needs to put in to produce the output \(y\). In their work, \(u\) takes the form

\[
u(x_1, \ldots, x_n, t) = v(\varphi(x_1, \ldots, x_n, t), x_n, t);
\]

which clearly is a special case of Eq. (1). Besides allowing for a more general form of the dependence of utility on \(t\), Eq. (1) differs from Eq. (2) by allowing for the possibility that the subutility \(\varphi\) may depend on all outcome variables, rather than just \(x_1, \ldots, x_n\).

3. Incentive compatibility

An agent's type is assumed to be his private information. An allocation is a mapping \(\tau \mapsto (x_1(\tau), \ldots, x_n(\tau))\) that indicates how outcomes depend on types. An allocation is incentive-compatible on the interval \([\ell_0, t_1]\) if

\[
u(x_1(\tau), \ldots, x_n(\tau), \tau) \geq u(x_1(\tau'), \ldots, x_n(\tau'), \tau)
\]

for all \(\tau\) and \(\tau'\) in \([\ell_0, t_1]\). Under the separability assumption (1), this condition is equivalent to the requirement that

\[
u(\varphi(t), x_n(t), t) \geq u(\varphi(t'), x_n(t'), t)
\]

which is satisfied if \(v_{x_n} > 0\).
for all \( t \) and \( t' \) in \( [t_0, t_1] \), where, for each \( t \),
\[
   w(t) := \psi(x_1(t), \ldots, x_n(t)).
\]

Incentive compatibility depends only on \( x_0(t) \) and on the subutility \( w(t) \). Thus, one obtains:

**Lemma 1.** Let \((x_1(\cdot), \ldots, x_n(\cdot))\) be an incentive-compatible allocation. For each \( t \), let \( w(t) \) be given by Eq. (5), and let \( \hat{x}_1(t), \ldots, \hat{x}_{n-1}(t) \) be such that
\[
   \psi(\hat{x}_1(t), \ldots, \hat{x}_{n-1}(t), x_n(t)) = w(t).
\]

Then the allocation \((\hat{x}_1(\cdot), \ldots, \hat{x}_{n-1}(\cdot), x_n(\cdot))\) is also incentive-compatible.

The observation that incentive compatibility reduces to condition (4) brings the analysis into the domain of known results about incentive compatibility. Given that \( v \) satisfies a single-crossing conditions, well-known arguments yield:

**Lemma 2.** An allocation \((x_1(\cdot), \ldots, x_n(\cdot))\) is incentive-compatible on \([t_0, t_1]\) if and only if \( x_n(\cdot) \) is a nondecreasing function and, for all \( t \) and \( t' \) in \([t_0, t_1]\), one has
\[
   V(t) - V(t') = \int_{t'}^{t} v_t(w(t), x_n(t), \tau) \, d\tau,
\]
where
\[
   V(t) := v(w(t), x_n(t), t).
\]

**4. Feasibility and dominance**

Let \( F \) be the cross-section distribution of agents’ types, where \( F \) has support \([t_0, t_1]\). An allocation is said to be *feasible* if the aggregates \( \int x_i(t) \, dF(t) \), \( i = 1, \ldots, n \), satisfy the resource constraint
\[
   \sum_{i=1}^n p_i \int x_i(t) \, dF(t) \leq K,
\]
where \( p_1, \ldots, p_n \geq 0 \) and \( K \) are constant. If the inequality in Eq. (9) is strict, the allocation is said to be *strictly feasible*.

**Lemma 3.** Let \((x_1(\cdot), \ldots, x_n(\cdot))\) be feasible and incentive-compatible on \([t_0, t_1]\) and assume that
\[
   \sum_{i=1}^n p_i x_i(t) > \min_{x_1, \ldots, x_{n-1}} \sum_{i=1}^n p_i x_i \text{ s.t. } \psi(x_1, \ldots, x_{n-1}, x_n(t)) \geq w(t),
\]
for a nonnegative set of types, where \( w(t) \) is given by Eq. (5). Then there exists an allocation \((\hat{x}_1(\cdot), \ldots, \hat{x}_{n-1}(\cdot), x_n(\cdot))\) with \( \hat{x}_1(\cdot) = x_1(\cdot) \) that is strictly feasible and incentive-compatible on \([t_0, t_1]\) and that generates the same payoffs as the allocation \((x_1(\cdot), \ldots, x_n(\cdot))\).

**Proof.** For each \( t \), if
\[
   \sum_{i=1}^n p_i x_i(t) = \min_{x_1, \ldots, x_{n-1}} \sum_{i=1}^n p_i x_i \text{ s.t. } \psi(x_1, \ldots, x_{n-1}, x_n(t)) \geq w(t),
\]
let \( \hat{x}_1(t) = x_1(t), \ldots, \hat{x}_{n-1}(t) \) be such that
\[
   \sum_{i=1}^n p_i \hat{x}_i(t) \leq \sum_{i=1}^n p_i x_i(t)
\]
and, moreover, \( \psi(\hat{x}_1(t), \ldots, \hat{x}_{n-1}(t), x_n(t)) = w(t) \). Then, by construction the allocation \((\hat{x}_1(\cdot), \ldots, \hat{x}_{n-1}(\cdot), x_n(\cdot))\) is strictly feasible and payoff-equivalent to \((x_1(\cdot), \ldots, x_n(\cdot))\). By Lemma 1, it is also incentive-compatible on \([t_0, t_1]\).

If the surplus of the allocation \((\hat{x}_1(\cdot), \ldots, \hat{x}_{n-1}(\cdot), x_n(\cdot))\) can be redistributed to the participants without upsetting incentive compatibility, the allocation \((x_1(\cdot), \ldots, x_n(\cdot))\) in Lemma 3 is actually Pareto-dominated. This is the point of:

**Lemma 4.** If \((x_1(\cdot), \ldots, x_n(\cdot))\) satisfy the assumptions of Lemma 3, there exists an \((\bar{x}_1(\cdot), \ldots, \bar{x}_n(\cdot))\) that is feasible and incentive-compatible on \([t_0, t_1]\) and that provides every type with a payoff that is strictly greater than his payoff under the allocation \((x_1(\cdot), \ldots, x_n(\cdot))\).

**Proof sketch.** Given the allocation \((x_1(\cdot), \ldots, x_n(\cdot))\), let \( V(\cdot) \) and \( w(\cdot) \) be given by Eqs. (8) and (5) and, for any \( \Delta \geq 0 \), consider the integral equation
\[
   W(t, \Delta) = W(t_0, \Delta) + \int_{t_0}^{t} v_t \left( \bar{w}(t, \Delta), x_n(\tau), \tau \right) d\tau,
\]
with the initial condition
\[
   W(t_0, \Delta) = V(t_0) + \Delta.
\]

Then, as \( \Delta \rightarrow 0 \), the allocation \((\bar{x}_1(\cdot), \ldots, \bar{x}_n(\cdot))\) converges to the strictly feasible allocation \((x_1(\cdot), \ldots, x_{n-1}(\cdot), x_n(\cdot))\) and, if \( \Delta \) is sufficiently small, the allocation \((\bar{x}_1(\cdot), \ldots, \bar{x}_n(\cdot))\) itself must be feasible. By Lemma 3, it is also incentive-compatible.

If one combines these lemmas with the taxon principle of Hammond (1979) and Guesnerie (1995), one obtains the following extension of the Atkinson–Stiglitz theorem.

**Theorem 5.** Assume that \( \psi \) is everywhere increasing in \( t \) and let \((x_1(\cdot), \ldots, x_n(\cdot))\) be Pareto-optimal in the set of feasible, incentive-compatible allocations. Then, there exists a tax schedule \( T(\cdot) \) such that, for almost all \( t \in [t_0, t_1] \), \((x_1(t), \ldots, x_n(t))\) maximizes \( u(x_1, \ldots, x_n, t) \) under the constraint that
\[
   \sum_{i=1}^n p_i x_i + T(x_n) \leq 0.
\]

**Proof.** For any \( x_n \) and \( w \), let \( \bar{r}_n(x_n, w) := \arg \max \{ E(x_n, w) | x_n, w \} \) be a solution to the problem of minimizing \( \sum_{i=1}^{n-1} p_i x_i \) under the constraint that \( \psi(x_1, \ldots, x_n) \geq w \), and let \( E(x_n, w) = \sum_{i=1}^{n-1} p_i \bar{r}_i(x_n, w) \). By Lemmas 3 and 4 and the Pareto-optimality of the allocation \((x_1(\cdot), \ldots, x_n(\cdot))\), one must have
\[
   \sum_{i=1}^{n-1} p_i \bar{r}_i(t) = E(x_n(t), w(t))
\]
for almost all \( t \in [t_0, t_1] \), where, again \( w(t) \) is given by Eq. (5). For any \( t \in [t_0, t_1] \), let
\[
   \tau(t) = -E(x_n(t), w(t)) - p_n x_n(t).
\]
I claim that, for any \( t \) and \( t' \), \( x_0(t) = x_0(t') \) implies \( \tau(t) = \tau(t') \). For suppose that \( x_0(t) = x_0(t') \) and \( \tau(t) \neq \tau(t') \). If \( \tau(t) > \tau(t') \), then, by Eq. (16), one must have \( w(t) < w(t') \). By the monotonicity of \( v \), it follows that \( v(w(t), x_0(t)) < v(w(t'), x_0(t)) = v(w(t'), x_0(t')) \), contrary to the assumption that the allocation \( \{x_1(\cdot), \ldots, x_n(\cdot)\} \) is incentive-compatible. The assumption that \( x_0(t) = x_0(t') \) and \( \tau(t) > \tau(t') \) thus leads to a contradiction and must be false. A precisely symmetric argument eliminates the possibility that \( x_0(t) = x_0(t') \) and \( \tau(t) < \tau(t') \).

Because \( x_0(t) = x_0(t') \) implies \( \tau(t) = \tau(t') \), there exists a function \( T(\cdot) \), defined on the range of \( x_0(\cdot) \) such that, for any \( t \in [t_0, t_1] \), one has \( \tau(t) = T(x_0(t)) \). Thus, if \( x_0 = x_0(t') \) for some \( t' \in [t_0, t_1] \) and \( (x_1, \ldots, x_n) \) satisfies Eq. (14), one has

\[
\sum_{i=1}^{n-1} p_i x_i \leq E(x_0(t'), w(t')).
\]

and, hence,

\[
u(x_1, \ldots, x_n, t) \leq v(w(t'), x_0(t'), t),
\]

regardless of \( t \). By incentive compatibility, it follows that

\[
u(x_1, \ldots, x_n, t) \leq v(w(t), x_0(t), t) = u(x_1(t), \ldots, x_n(t), t)
\]

for any such \( (x_1, \ldots, x_n) \) and any \( t \).

To complete the proof, it suffices to note that, for any \( x_0 \) that does not belong to the range of the function \( x_0(\cdot) \), incentive compatibility is obtained by setting \( T(x_0) \) large enough so that no type would like to choose a vector \( (x_1, \ldots, x_n) \) which gives rise to this tax.

**Remark 6.** The feasibility condition (9) can be replaced by the more general condition \( \mathcal{G}(\int x_1(t) dF(t), \ldots, \int x_n(t) dF(t)) \leq 0 \) if convexity assumptions on preferences and technology make it possible to work with a price vector \( (p_1, \ldots, p_n) \), as in the second welfare theorem.

**Remark 7.** The extension of the Atkinson-Stiglitz theorem that is provided by **Theorem 5 is not available for Deaton’s (1979) theorem on the undesirability of nonuniform excise taxation when income taxes are affine and preferences over consumption goods are separable from labour–leisure choices, quasi-homothetic, and identical across agents. Although the basic argument of Kaplow (2006) and Laroque (2005) can also be used to prove Deaton’s theorem,\(^3\) in the more general case, when \( x_0 \) is an argument of \( \varphi \), there is no reason to believe that, if one applies the taxation principle to the Pareto-dominating allocation that is obtained from Lemmas 3 and 4, the resulting tax schedule will be affine in \( x_0 \).

**Acknowledgments**

I am grateful for discussions with Guy Laroque and for helpful comments from Felix Bierbrauer, Christoph Engel, and Andreas Nicklisch.

**References**


---

\(^3\) This is shown in Hellwig (2009). In that paper, Deaton’s theorem is erroneously quoted as assuming homotheticity, and the argument is given for that case. In fact, Deaton (1979) has the more general assumption of quasi-homotheticity; the argument in Hellwig (2009) applies to this more general case as well.