OPTIMAL INCOME TAXES AND TRANSFERS

Henrik Jacobsen Kleven

London School of Economics

Lecture Notes for MSc Public Economics (EC426)
AGENDA

1. The fundamental welfare theorems and the information constraint in redistribution policy.

2. The Mirrlees optimal nonlinear income tax problem:
   (a) The optimal high-income tax rate.
   (b) The general optimal nonlinear income tax.

3. Commodity vs income taxation [Atkinson-Stiglitz].

4. Optimal transfer programs:
   (a) Tagging [Akerlof].
   (b) In-work benefits (EITC).
   (c) In-kind benefits and ordeals.
THE TWO FUNDAMENTAL THEOREMS OF WELFARE ECONOMICS

First Theorem: under (i) perfect competition, (ii) no externalities/internalities, and (iii) perfect information, the competitive equilibrium is Pareto efficient.

Second Theorem: under (i) convex preferences and technology, (ii) no externalities/internalities, and (iii) perfect information, any Pareto efficient allocation can be achieved as a competitive equilibrium with appropriately selected endowments.
EFFICIENCY VERSUS EQUALITY

Is there a trade-off between efficiency and equality? Not according to the second welfare theorem, which says that any Pareto efficient allocation can be achieved as an equilibrium.

A strong assumption is made: the government can observe and redistribute exogenous endowments using individual lump-sum taxes/transfers ["tax on DNA"].

In practice, redistribution takes place through taxes, not on endowments, but on choices. Such taxes are distortionary and leads to Pareto inefficiency ⇒ a trade-off between efficiency and equality.
A first-best redistribution scheme is based on innate ability (the endowment in the second welfare theorem), which is immutable for the individual.

But ability is known only to the individual; the gov’t observes instead earnings (asymmetric information).

Because earnings is a choice variable, earnings-based redistribution induces high-ability individuals to reduce earnings and masquerade as low-ability individuals.

Given the information constraint, optimal income tax theory analyzes the second-best redistribution scheme.
THE MIRRLEES PAPER

Mirrlees (1971) is the first rigorous treatment of optimal income taxation.

It is the central paper in modern public finance. Nobel prize in 1996.

Widely known paper because it started the literature on asymmetric information (moral hazard and adverse selection). Hugely influential in IO, auction theory, mechanism design, contract theory, etc.
THE MIRRLEES SETTING: INDIVIDUALS

A continuum of individuals with heterogeneous and exogenous skill $w$ distributed according to $f(w), F(w)$.

Perfect competition and linear technology $\rightarrow$ skill = wage rate.

Individuals have identical preferences $u(c,h)$.

Budget is $c = wh - T(wh)$.

$T(.)$ is a general tax function (embodying transfers).

Maximize $u(wh - T(wh), h)$ with respect to $h$.

FOC: $w(1 - T'(wh))u'_c + u'_h = 0 \Rightarrow h = h(w)$. 

THE MIRRLEES SETTING: GOVERNMENT

Additively separable Bergson-Samuelson social welfare function $\Psi(.)$ where $\Psi' > 0$, $\Psi'' \leq 0$.

Government problem:

$$\max_{T(.)} W = \int_{\bar{w}} \Psi (u (wh - T(wh), h)) f(w) dw$$

subject to

$$\int_{\bar{w}} T(wh) f(w) dw \geq R \quad [\text{GBC, multiplier } \mu],$$

and

$$w(1 - T'(wh))u'_c + u'_h = 0 \quad [\text{IC}].$$

This formulation assumes that the Spence-Mirrlees single crossing condition is satisfied. See Salanie (2003) for details.
RESULTS IN THE EARLY LITERATURE

Mirrlees (1971) solved the general problem using the Hamiltonian approach.

Formulas are complex and hard to interpret. Very few results on the shape of optimal tax schedules.

Three results in the early literature:

1. Optimal marginal tax rate is zero at the top [Sadka (1976); Seade (1977)].

2. Optimal marginal tax rate is zero at the bottom if the lowest skill is positive and everybody works [Seade (1977)].

3. Optimal marginal tax rates are always between 0 and 1 [Seade (1977)].
RECENT LITERATURE

Piketty (1997), Diamond (AER 1998), and Saez (RES 2001):

• Relate optimal tax formulas to labor supply elasticities.

• Link theory to data on income distributions and empirical elasticities → statements about the optimal marginal tax rate profile.

• Perturbation approach to deriving optimal tax formulas (considering small tax reforms around the optimum).

Before considering the full optimal tax schedule, we will consider an easier subproblem: what is the optimal high-income tax rate?
OPTIMAL HIGH-INCOME TAX RATE

Assume a constant marginal tax rate $\tau$ above a given income level $\bar{z}$.

Individuals $i = 1, \ldots, N$ in the top bracket where individual $i$ has earnings $z^i \geq \bar{z}$. Mean income is $z_m \equiv \sum_i z^i / N$.

Assume no income effects so that $z^i = z^i (1 - \tau)$.

Earnings elasticity $\varepsilon^i \equiv \frac{d z^i / z^i}{d (1-\tau)/(1-\tau)}$. Assume $\varepsilon^i = \bar{\varepsilon}$ at the top.

Denote by $\mu$ the marginal value of public funds, by $\alpha^i$ the social marginal utility of income to individual $i$, and define $g^i \equiv \alpha^i / \mu$. Assume $g^i = \bar{g}$ at the top.
OPTIMAL HIGH-INCOME TAX RATE

Raise $\tau$ slightly by $d\tau$. Three effects on social welfare:

(1) Mechanical revenue effect:

$$dM = \sum_i (z^i - \bar{z}) \cdot d\tau = [z_m - \bar{z}] \cdot N \cdot d\tau.$$  

(2) Behavioral revenue effect:

$$dB = \sum_i \tau \cdot dz^i = -\bar{e} \cdot \frac{\tau}{1 - \tau} \cdot z_m \cdot N \cdot d\tau.$$  

(3) Direct welfare effect:

$$dW = -\bar{g} \cdot dM.$$  

At the optimum, we must have $dM + dB + dW = 0$. 

OPTIMAL HIGH-INCOME TAX RATE

The optimal top marginal tax rate:

\[
\frac{\tau}{1 - \tau} = \frac{1}{\bar{\varepsilon}} \cdot \left[ \frac{z_m - \bar{z}}{z_m} \right] \cdot [1 - \bar{g}].
\]

where \( \frac{z_m - \bar{z}}{z_m} \in [0, 1) \) depends on the income distribution, and reflects the relative strength of mechanical and behavioral effects.

The top marginal tax rate \( \tau \) is

1. decreasing in the **social welfare weight** on the rich \( \bar{g} \).
2. decreasing in the **earnings elasticity** at the top \( \bar{\varepsilon} \).
3. increasing in the **income distribution** variable \( \frac{z_m - \bar{z}}{z_m} \).
NO DISTORTION AT THE TOP

To obtain the tax rate at the upper bound of the income distribution, let the threshold \( \tilde{z} \) be equal to the upper-bound income.

In this case, we have \( z_m = \tilde{z} \) so that \( \frac{z_m - \tilde{z}}{z_m} = 0 \) \( \Rightarrow \) \( \bar{\tau} \) is zero at the top.

This result holds even when the gov’t does not value the marginal consumption of the rich \( (\bar{g} = 0) \).

Intuition: close to the upper bound, the mechanical welfare gain of raising taxes \( (1 - \bar{g})dM \) is negligible relative to the behavioral welfare loss \( dB \).
PRACTICAL RELEVANCE OF ZERO TOP RATE

Examine \((z_m - \bar{z})/z_m\) in empirical earnings distributions. For the US, Saez (2001) shows that \(z_m/\bar{z} \approx 2\) (so that \((z_m - \bar{z})/z_m \approx 0.5)\) from $150,000 to $30 million.

Distributions with constant \((z_m - \bar{z})/z_m\) are Pareto distributions. In general, upper tails of empirical distributions are roughly Pareto.

Pareto distribution: Prob(income \(\geq z\)) = \(k/z^a\) where \(a > 1\) measures the thinness of the upper tail. We have \((z_m - \bar{z})/z_m = 1/a)\).

No-distortion-at-the-top result is not practically relevant as \((z_m - \bar{z})/z_m\) starts dropping to zero only at extreme incomes. The top-bracket rate in a piecewise linear system would not be affected by the result.
RATIO $Z_m / \bar{Z}$ IN THE UNITED STATES, 1992-93

Source: Saez (2001)
SOAKING THE RICH

With a Pareto tail and in the special case where $\bar{g} = 0$, we have

$$\bar{\tau} = \frac{1}{1 + a \bar{\varepsilon}}$$

which is the top Laffer rate (see also lecture 2).

A generalization of the well-known formula for the flat/linear revenue maximizing tax rate $1/[1 + \varepsilon]$.

If the gov’t does not value the marginal consumption of the rich, it will collect as much revenue as possible from them.

This would be optimal for a Rawlsian social planner who cares only about the worst-off individuals in society.
GENERAL OPTIMAL NON-LINEAR INCOME TAX

General Mirrlees problem: determine the tax function $T(z)$ at each $z$.

No income effects as in Diamond (1998). Elasticity $\varepsilon \equiv \frac{\partial z}{\partial (1-\tau)} \frac{1-\tau}{z}$.

Normalize total population to 1. Earnings distributed according to distribution function $F(z)$ and density function $f(z)$.

Denote by $g(z) \equiv \alpha(z)/\mu$ the social marginal value of consumption (in terms of the value of public funds) for taxpayers at income $z$. Define $G(z) \equiv \int_{z}^{\infty} g(s) f(s) ds/[1 - F(z)]$.

Consider a perturbation around the optimum: increase marginal tax rate $\tau \equiv T'(z)$ by a small amount $d\tau$ in a small interval $(z, z + dz)$. 
PERTUBATION AROUND THE OPTIMAL TAX SCHEDULE

Tax $T(z)$

Earnings $z$

$\text{slope } \tau + d\tau$

$\text{slope } \tau$

$\text{d}T$

$z$, $z + dz$
EFFECTS OF SMALL TAX REFORM

(1) Mechanical revenue effect:

\[ dM = dz \cdot d\tau \cdot (1 - F(z)). \]

(2) Behavioral revenue effect:

\[ dB = -\varepsilon \cdot \frac{d\tau}{1 - \tau} \cdot z \cdot \tau \cdot f(z)dz. \]

(3) Direct welfare effect:

\[ dW = -\int_{z}^{\infty} g(s) \cdot dz \cdot d\tau \cdot f(s)ds = -G(z) \cdot dM. \]

At the optimum, we must have \( dM + dB + dW = 0. \).
OPTIMAL MARGINAL TAX RATES

A general characterization of the optimal tax structure:

\[
\frac{\tau(z)}{1 - \tau(z)} = \frac{1}{\varepsilon(z)} \cdot \left[ \frac{1 - F(z)}{zf(z)} \right] \cdot [1 - G(z)].
\]  

(1)

This is the Diamond (1998) optimal tax formula (which he derives using the Hamiltonian method).

Three elements determine marginal tax rates:

1. The earnings elasticity \( \varepsilon(z) \).

2. The shape of the income distribution captured by \( \frac{1 - F(z)}{zf(z)} \).

3. Social marginal welfare weights \( G(z) \).
OPTIMAL MARGINAL TAX PROFILE

Consider the hazard ratio \( r(\zeta) \equiv [1 - F(\zeta)]/[zf(\zeta)] \):

1. At the bottom, \( r(\zeta) \) is very high \( \Rightarrow \tau(\zeta) \) very high at the bottom.
2. At the top, for a Pareto distribution, we have \( r(\zeta) = 1/a \).
3. Empirically, \( r(\zeta) \) is strongly decreasing at the bottom, weakly increasing at the middle, and constant at the top.

Consider redistributive tastes:

1. \( G(\zeta) \) is decreasing in \( \zeta \) and tends to \( \bar{g} \) at the top
   \( \Rightarrow \tau \) increasing other things equal.
2. Rawlsian case: \( G(\zeta) = 0 \) and eq. (1) gives the Laffer rate at each \( \zeta \).
HAZARD RATIO $[1-F(z)]/[zf(z)]$ IN THE US, 1992-93

Source: Saez (2001)
HAZARD RATIO $[1-F(z)]/[zf(z)]$ IN THE UK, 2003-04

Source: Brewer-Saez-Shepard (2009)
OPTIMAL MARGINAL TAX RATES THE UK
(ELASTICITY SENSITIVITY)

Source: Brewer-Saez-Shepard (2009)
OPTIMAL MARGINAL TAX RATES THE UK
(REDISTRIBUTIVE TASTE SENSITIVITY)

Source: Brewer-Saez-Shepard (2009)
TAX-TRANSFER TREATMENT OF FAMILIES

The Mirrlees literature considers individuals with one unobserved ability parameter. This is a one-dimensional screening problem.

In reality, redistribution takes place across families/couples with unobserved abilities of both husband and wife. This is a multi-dimensional screening problem.

Such problems are very hard to analyze. First step towards a theory of nonlinear income taxation of couples: Kleven-Kreiner-Saez (2007, 2009).

Earlier work on linear income taxation of couples (a Ramsey-style problem), which avoids multi-dimensional screening issues.
INCOME VS COMMODITY TAXATION

We have seen two models:

1. Many-person Ramsey model calls for differentiated linear tax rates on goods to redistribute income.

2. Mirrlees nonlinear income tax model calls for progressive nonlinear taxation of income.

Should we have a mix of both, or only one of them, in order to maximize social welfare?

This question was first analyzed by Atkinson-Stiglitz (1976), who found a condition under which we don’t need commodity taxes.
ATKINSON-STIGLITZ RESULT

Utility $u^i = u^i(c_1, ..., c_N, z)$ and budget $p_1c_1 + ... + p_Nc_N \leq z - T(z)$ where $p_k = q_k + t_k$ and $T(.)$ is a general nonlinear income tax.

Under weak separability of goods and labor plus homogeneous preferences for goods, i.e. $u^i = u^i(v(c_1, ..., c_N), z)$, there is no need for commodity taxation.

In this case, the consumption pattern depends only on net-of-tax income, not on type $i$. Hence, any redistribution achieved by commodity taxes can be achieved by an income tax and no distortion of commodity prices.

Recent simple proofs of this result: Laroque (2005), Kaplow (2006).
DEVIATIONS FROM ATKINSON-STIGLITZ

A positive tax on good $k$ is optimal if:

1. Good $k$ is a leisure complement. At a given income level, high-ability types have more leisure than low-ability types. Hence, a tax on leisure complements can achieve a redistribution across skill levels at a given level of after-tax income.
   **Example:** vacation trips. **Anti-example:** work uniforms.

2. High-skilled types have higher taste for good $k$ (conditional on income). Again, taxing those goods can achieve a redistribution across skill at a given level of after-tax income.
   **Example:** modern art museums. **Anti-example:** cigarettes.
APPLICATION OF ATKINSON-STIGLITZ: CAPITAL VS LABOR TAXATION

For simplicity, consider the standard 2-period model where individuals work in period 1 and live off savings in period 2.

Utility \( u^i = u^i(c_1, c_2, z) \) where \( c_k \) is consumption in period \( k \).

Life-time budget constraint \( c_1 + \frac{c_2}{1+r(1-t_c)} = z - T(z) \), where \( t_c \) is a tax on capital.

Notice that a capital tax works like a differentiated commodity tax with a higher tax on future than on present consumption.

This framework is a special case of the Atkinson-Stiglitz framework.
CAPITAL VS LABOR TAXATION

Standard specification in macro:

\[
u^i = u(c_1) + \rho \cdot u(c_2) - v(z/w_i),
\]

(2)

where \(\rho\) is the discount factor and \(w_i\) is the wage rate (skill) of type \(i\).

Eq. (2) satisfies the Atkinson-Stiglitz condition \(\Rightarrow\) **optimal capital tax is zero**: we should tax only labor income.

The controversial part of (2) is not so much separability, but that \(\rho, u(.)\) are common across skills. This implies that savings provide no signal of skill conditional on earnings and is not helpful for redistribution.

Empirical evidence: savings propensities are correlated with education (skill) conditional on income. This would call for a positive capital tax.
INCOME TRANSFERS IN THE MIRRLEES MODEL

The Mirrlees model characterizes the income tax *net of transfers* at each income level, $T(z)$.

For a gov’t with redistributive preferences, we would have $T(0) < 0$.

We can think of the $T$-schedule as the combination of (i) an out-of-work transfer $-T(0)$, and (ii) a pattern of marginal tax rates $T'(z)$ showing how the transfer is taxed away as earnings increase.

Our analysis of (ii) shows that $T'(z)$ is very high at the bottom: a redistribution scheme with out-of-work cash transfers that are taxed away rapidly as earnings increase [means-tested cash transfers].
LITERATURE ON TRANSFER PROGRAMS

Can we do better than means-tested cash transfers to those out of work?

1. Means-tested vs categorical programs (tagging) [Akerlof 1978].

2. Out-of-work vs in-work benefits [Saez 2002].


TAGGING

If we can identify individual characteristics that are

1. observable to the government

2. negatively correlated with ability

3. immutable for the individual (unresponsive to incentives)

then targeting benefits to such characteristics is optimal.

Criteria 1 makes this form of targeting feasible, criteria 2 ensures that it redistributes from high- to low-ability, and criteria 3 ensures that there is no efficiency cost associated with this redistribution.
TAGGING IN PRACTICE

We are looking for characteristics that are (1) observable, (2) correlated with earnings capacity, and (3) immutable.

Are there characteristics satisfying all three criteria in practice?

Potential candidates: (i) disability, (ii) race, (iii) gender, (iv) age, (v) height, (vi) beauty, (vii) single motherhood.

Single motherhood is widely used as a tagging device in many countries. It satisfies 1 and 2, but does it satisfy 3? It has been accused by conservatives of destroying the traditional family.
IS SINGLE MOTHERHOOD CAUSED BY WELFARE INCENTIVES?

Evidence suggests that the effect is either very small or non-existent:

1. **U.S. time series evidence** shows that single motherhood has been increasing since the 70s, whereas welfare benefits have been declining. But it is difficult to draw causal interpretations from time series.

2. **U.S. state/time variation** in welfare benefits and single motherhood. Single motherhood does not grow (significantly) more in states that are raising benefits relative to others (Moffitt, JEL 1992; Blank, JEL 2002).
WELFARE BENEFITS AND SINGLE MOTHERHOOD
OVER TIME IN THE UNITED STATES

Source: Gruber (2007)
IN-WORK BENEFITS

Transfers which are conditional on labor force participation.

They are typically means-tested and categorical (targeted to low-income single mothers).

Main examples:
– **Earned Income Tax Credit (EITC)** in the US
– **Working Families Tax Credit (WFTC)** in the UK

We saw in the labor supply lecture that these programs have been successful at increasing labor force participation [Eissa-Liebman 1996; Meyer-Rosenbaum 2001].
IS THE EITC AN OPTIMAL POLICY?

Define the EITC conceptually as **negative tax rates at the bottom**

Can negative marginal tax rates be optimal? What does Mirrlees (1971) have to say about this?

Consider the optimal tax structure in equation (1):

- Average social welfare weight in the population equals 1, $G(0) = 1$.
- Since $G(z)$ is declining, we then have $G(z) \leq 1 \ \forall z \Rightarrow \tau(z) \geq 0$.

This **rules out an EITC**.
A TAX REFORM AROUND A SCHEDULE WITH NEGATIVE TAX RATES

Tax $T(z)$

Earnings $z$

slope $\tau + d\tau$

slope $\tau < 0$
INTUITION FOR MIRRLEES NO-EITC RESULT

If $\tau$ were negative in an interval $(z, z + dz)$, then a small reform which increases $\tau$ in this interval and distributes the proceeds lump sum will

1. improve equity, and

2. reduce hours worked which improves efficiency (with negative $\tau$, people are working too much).

$\Rightarrow$ Schedule with negative $\tau$ cannot be optimal.
ACCOUNTING FOR EXTENSIVE RESPONSES

Mirrlees (1971) is based on a **convex model** with only **intensive** labor supply responses.

The empirical labor supply literature shows that **non-convexities** and **extensive** responses are important.

Does the non-negativity of optimal tax rates survive once we account for extensive responses?

This has been analyzed by Saez (2002) and the answer is "not necessarily".
EXTENSIVE MODEL: SETUP

Saez (2002) sets out a reduced-form discrete model. I will instead work with a fully specified, micro-founded model with fixed work costs.

To simplify, I ignore intensive responses and normalize hours worked for workers to 1.

There is heterogeneity in two dimensions: ability $n$ and fixed costs of working $q$.

Ability is distributed on $(\bar{n}, \bar{n})$ according to density $f(n)$. Fixed costs are distributed on $(0, \infty)$ according to the conditional density $p(q|n)$. 
EXTENSIVE MODEL: SETUP

Individual utility if working is given by $v(n, q) = n - T(n) - q$.

Individual utility if not working $v(0) = -T(0)$.

The participation constraint $v(n, q) \geq v(0)$ implies

$$q \leq n - [T(n) - T(0)] \equiv \bar{q}(n).$$

(3)

The participation rate is $\int_0^{\bar{q}} p(q|n) \, dq = P(\bar{q}|n)$, and the participation elasticity is

$$\eta(n) \equiv \frac{\partial P(\bar{q}|n)}{\partial \bar{q}} \frac{\bar{q}}{P(\bar{q}|n)}.$$  

(4)
EXTENSIVE MODEL: GOVERNMENT

Government chooses $T(0)$ and $\{T(n)\}_{n=1}^{\bar{n}}$ to maximize

$$W = \int \int \int \Psi(v) p(q|n) f(n) dqdn$$

subject to individual optimization (3) and a government budget constraint ($\lambda$). We assume $\Psi' > 0$ and $\Psi'' < 0$.

A key concept is the social marginal welfare weight:

$$g(n, 1) = \frac{\int \Psi' \left( v(n, q) \right) p(q|n) dq}{\lambda P(\bar{q}|n)}$$

$$g(0) = \frac{\Psi' \left( v(0) \right)}{\lambda}. $$

The average welfare weight in the population $E[g]$ equals 1.
EXTENSIVE MODEL: TAX REFORM PERTUBATION APPROACH

Say that the tax system has been optimized.

Consider a reform increasing the tax burden by a small amount \( dT \) for all participants in a small earnings band \([n, n + dn]\).

The additional tax proceeds are distributed lump-sum.

Given that the tax system has already been optimized, the reform will not change social welfare.
EXTENSIVE MODEL: OPTIMAL TAX RULE

The direct effect on social welfare:

\[ dW = dT \cdot (1 - g(n, 1)) \cdot P(\bar{q}|n) f(n) \, dn. \]

The fiscal externality from participation responses:

\[ dB = -dT \cdot (T(n) - T(0)) \cdot p(\bar{q}|n) f(n) \, dn. \]

At the optimum, we must have \( dW + dB = 0 \) \( \Rightarrow \)

\[ \frac{a(n)}{1 - a(n)} = \frac{1 - g(n, 1)}{\eta(n)}, \]

where \( a(n) \equiv (T(n) - T(0))/n \) is the participation tax rate.
INTERPRETATION OF OPTIMAL TAX RULE

If $g(n, 1) > 1$ at the bottom, the gov’t cares more for the working poor than for the average individual.

Then tax subsidies $a(n) < 0$ at the bottom (evaluated at $g(n, 1)$) financed by lump sum taxes (evaluated at $E[g]$) create equity gains.

But tax subsidies also distort participation (people work too much!) and create an efficiency loss.

But starting at $a = 0$ where participation is undistorted, the first dollar of subsidy creates only a second-order efficiency loss ($dB = 0$ at $T(n) = T(0)$) but a first-order equity gain $\Rightarrow$ some subsidization is optimal.
SOCIAL PREFERENCES AND THE EITC

The optimality of an EITC at the bottom \( a(n) < 0 \) for low \( n \) is driven by social preferences for the working poor \( g(n, 1) > 1 \) for low \( n \).

The result is not driven by a large extensive elasticity (in fact, \( \eta \uparrow \Rightarrow \text{EITC} \downarrow \)), but relies crucially on the absence of intensive responses.

If the gov’t has Rawlsian preferences, it cares only about the worst-off individuals—the non-workers. Then \( g(n, 1) = 0 \ \forall n \) and the EITC is not optimal.
SOCIAL PREFERENCES AND THE EITC

EITC is optimal for the "standard" profile.
SOCIAL PREFERENCES AND THE EITC

- g-weights
- g(0)
- 1
- Rawlsian profile
- EITC is not optimal
- earnings $n$
SOCIAL PREFERENCES AND THE EITC

Profile with higher weights on the working poor than on the poor.
Consider ways of improving the efficiency of transfer programs by putting restrictions on recipients.

General theme: if such restrictions are less costly to intended recipients than to potential imposters, then they will induce self-revelation and hence improve targeting.

Restrictions can be placed on:
(a) The choice of income [related to means-testing and Mirrlees]
(b) The choice of consumption [related to Atkinson-Stiglitz]
(c) The allocation of time
A SIMPLE MODEL

Two individuals: high-ability \( w_H \) and low-ability \( w_L \). One common utility function \( u(c_i, h_i) \) for \( i = L, H \). Consumption \( c_i \) equals earned income \( w_i h_i \equiv y_i \) plus a net transfer.

The gov’t has a concave social welfare function and wishes to redistribute from \( H \) to \( L \), but it cannot observe abilities.

The tax-transfer schedule has to be based on income and takes the following form: if income is less than \( \bar{y} \), receive transfer \( T \). If income is greater than \( \bar{y} \), pay tax \( T \).
If the scheme is too generous, $H$ masquerades as $L$ by choosing income $\bar{y}$. In this case, there would be no redistribution, only inefficient choices.

The optimal tax-transfer scheme must satisfy incentive compatibility

\[
u (y^*_H - T, \frac{y^*_H}{w_H}) \geq u (\bar{y} + T, \frac{\bar{y}}{w_H}) \tag{IC}
\]

where $y^*_H$ is the optimal choice of $H$ conditional on not masquerading.

The optimal tax system redistributes as much as possible, while ensuring that (IC) is satisfied.
OPTIMAL TRANSFER ELIGIBILITY

A natural solution is $\bar{y} = y_L^*$. This reduces $\bar{y}$ as far below $y_H^*$ as possible (to prevent moral hazard) without hurting $L$.

But this is not optimal because, from (IC), if we reduce $\bar{y}$ further, we relax the incentive compatibility constraint and can increase $T$.

As we move $\bar{y}$ slightly below $y_L^*$, the cost to $L$ is second-order but the cost to $H$ (if he masquerades) is first-order. Room to increase $T$ (a first-order gain for $L$) without $H$ masquerading.

This income restriction sacrifices some productive efficiency to increase targeting efficiency.
THE OPTIMALITY OF RESTRICTING RECIPIENTS’ INCOME

utility

pre-tax income

$y_H^*$

$y_L^*$ = $\bar{y}$

$\bar{y}'$

$u_H^*$ (get transfer)

$u_H$ (pay tax)

$u_L^*$

first-order utility loss

second-order utility loss
IN-KIND BENEFITS AND GOODS SUBSIDIES

Can target efficiency be improved further by restricting other choices than income? What about restricting consumption choices by linking redistribution to specific goods?

There is a strong a priori argument against doing this: consumer sovereignty. The utility gain from receiving a bundle of goods is never higher than the cash value of those goods.

But, although this argument applies to each recipient individually, it may not apply for a program as a whole.
EXTENDING THE MODEL

Allow for two consumption goods, $x$ and $z$, the prices of which are $p_x$ and $p_z$.

The utility function is $u (x_i, z_i, h_i, w_i)$ for $i = L, H$. We allow for ability to impact utility directly.

The tax-transfer schedule imposes a tax $T$ above income $\bar{y}$ and gives a transfer $T$ at or below $\bar{y}$.

This system has been optimized as prescribed above, i.e. $\bar{y} < y_L^*$ and $H$ marginally prefers not to masquerade.
DEMAND IF MASQUERADING

If $H$ masquerades, his budget is $p_x x_H + p_z z_H = \bar{y} + T$ and his utility is $u\left(x_H, z_H, \frac{\bar{y}}{w_H}, w_H\right)$. Utility is maximized under the budget with respect to $x_H$ and $z_H$.

Optimal demand for $x$ is then $x_H^* = x_H \left(p_x, p_z, \bar{y} + T, \frac{\bar{y}}{w_H}, w_H\right)$.

For the low-ability individual, we have $x_L^* = x_L \left(p_x, p_z, \bar{y} + T, \frac{\bar{y}}{w_L}, w_L\right)$.

Demand for $x$ (and $z$) may depend on ability either through the argument $\frac{\bar{y}}{w_i}$ (substitutability with leisure) or through the argument $w_i$ (direct preference-dependence on ability).
OPTIMALITY OF IN-KIND TRANSFER

If $x^*_L > x^*_H$, so that demand varies with ability conditional on income, then $x$ is an indicator good.

If such a good exists, we can improve targeting efficiency by linking redistribution to this good. The argument is a first-order vs second-order argument as before.

If we convert part of the cash transfer into a transfer of $x$ at a level slightly above $x^*_L$, we hurt $L$ a little bit but we hurt $H$ more (if he masquerades). Room to increase $T$ (a first-order gain for $L$) without $H$ masquerading.
THE OPTIMALITY OF IN-KIND TRANSFERS

utility

$u_H(x_H^*)$
$u_H(x_L^*)$
$u_L^*$

$u_H(x_H^*)$

$u_H(x_L^*)$

$u_L^*$

quantity of $x$

$u_H$ (masquerade)
THE OPTIMALITY OF IN-KIND TRANSFERS

utility

$u_H(x_H^*)$

$u_L(x_L^*)$

$u_L^*$

$u_H$ (masquerade)

quantity of $x$

first-order utility loss

second-order utility loss
WHEN ARE IN-KIND BENEFITS OPTIMAL?

1. If differences in demand across rich and poor are due only to income effects (luxuries vs. necessities), then in-kind transfers cannot improve target efficiency. Example: Fiat versus Ferrari

2. If, conditional on income, demand is higher for low-ability individuals, then in-kind transfers can improve target efficiency.

Does any in-kind programs in the real world satisfy the Nichols-Zeckhauser condition?
ORDEALS

An ordeal is a pure deadweight cost on recipients.

Examples of ordeals are:


2. Tedious and complex administrative procedures. Common in many social programs (Currie 2006).

3. Demeaning qualification tests.

ORDEALS

Ordeals can serve a screening function if (i) the utility gain from transfers is lower for the non-deserving, and/or if (ii) the utility cost of the ordeal is higher for the non-deserving.

Two types of recipients in a cash welfare program: low-ability individuals and high-ability but "lazy" individuals.

Introduce a time loss in the program, e.g. work requirements or waiting lines. The time loss will be more costly to the high-ability/lazy individuals.
COST OF ORDEALS: INCOMPLETE TAKE UP

A problem with many social programs, especially in the US, is that not all eligibles take up benefits. Take-up rates can be far lower than 100% in some programs.

Three possible explanations: (i) welfare stigma, (ii) imperfect information, (iii) transaction costs associated with take up.

The third explanation is related to the presence of ordeals. Currie (2006) suggests that transaction costs due to ordeals may be an empirically important factor for incomplete take up.