Optimal Income Taxation: Mirrlees Meets Ramsey

Jonathan Heathcote
FRB Minneapolis and CEPR

Hitoshi Tsujiyama
Goethe University Frankfurt

ESSIM May 28 2014
The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.
How should we tax income?

- What structure of income taxation offers best trade-off between benefits of public insurance and costs of distortionary taxes?

- Proposals for a flat tax system with universal transfers
  - Friedman (1962)
  - Mirrlees (1971)
This Paper

We compare 3 tax and transfer systems:

1. **Affine tax system**: \( T(y) = \tau_0 + \tau_1 y \)
   - constant marginal rates with lump-sum transfers

2. **HSV tax system**: \( T(y) = y - \lambda y^{1-\tau} \)
   - increasing marginal rates without transfers
   - \( \tau \) indexes progressivity of the system

3. **Optimal tax system**
   - fully non-linear
Main Findings

- Best tax and transfer system in the HSV class better than the best affine tax system

- Welfare gains moving from the current tax system to the optimal one are tiny

- Planner only observes earnings = productivity × effort
- Think of planner choosing earnings $x$ and cons. $c$ for each unobservable productivity type $\alpha$
- Include incentive constraints, s.t. each type prefers the earnings level intended for their type
- Allocations are constrained efficient
- Trace out tax decentralization $T(x(\alpha)) = x(\alpha) - c(\alpha)$
Novel Elements of Our Analysis

1. Our model has a distinct role for private insurance
   - Standard decentralization of efficient allocations delivers all insurance through tax system ⇒ Very progressive taxes

2. We use a SWF that rationalizes amount of redistribution embedded in observed tax system
   - Analyzes typically assume utilitarian social welfare function ⇒ Strong desire for redistribution
Environment 1

- Static environment
- Heterogeneous individual labor productivity $w$
- Log productivity is sum of two independent stochastic components

$$\log w = \alpha + \varepsilon$$

- $\alpha$ no private insurance
- $\varepsilon$ private insurance
- planner sees neither component of productivity
  - (later introduce a third productivity component $\kappa$ that the planner can observe)
Environment 2

- Common preferences
  \[ u(c, h) = \log(c) - \frac{h^{1+\sigma}}{1 + \sigma} \]

- Production linear in aggregate effective hours
  \[ \int \int \exp(\alpha + \varepsilon) h(\alpha, \varepsilon) dF_\alpha dF_\varepsilon = \int \int c(\alpha, \varepsilon) dF_\alpha dF_\varepsilon + G \]
Planner’s Problems

- Seeks to maximize SWF denoted $W(\alpha, \varepsilon)$
- Only observes total income $y = \text{earnings plus private insurance income}$

First Stage
- Planner offers menu of contracts \( \{c(\tilde{\alpha}, \tilde{\varepsilon}), y(\tilde{\alpha}, \tilde{\varepsilon})\} \)
- Agents draw idiosyncratic $\alpha$ and report $\tilde{\alpha}$

Second Stage
- Agents buy private insurance against insurable shock $\varepsilon$
- Draw $\varepsilon$, receive insurance payments and report $\tilde{\varepsilon}$
- Work sufficient hours to deliver $y(\tilde{\alpha}, \tilde{\varepsilon})$
- Receive consumption $c(\tilde{\alpha}, \tilde{\varepsilon})$
First result: Cannot condition on $\tilde{\varepsilon}$

- Offered contracts take the form \( \{c(\tilde{\alpha}), y(\tilde{\alpha})\} \)

- Private insurance markets undercut planner’s ability to condition allocations on $\varepsilon$

- Planner cannot take over private insurance $\Rightarrow$ **Distinct roles for public and private insurance**
Planner’s Problem: Second Best

\[ \max_{c(\alpha), y(\alpha)} \int W(\alpha) U(\alpha, \alpha) dF_\alpha \]

s.t. \[ \int y(\alpha) dF_\alpha \geq \int c(\alpha) dF_\alpha + G \]

\[ U(\alpha, \alpha) \geq U(\alpha, \tilde{\alpha}) \quad \forall \alpha, \forall \tilde{\alpha} \]

where \[ U(\alpha, \tilde{\alpha}) \equiv \]

\[ \begin{cases} 
\max_{h(\varepsilon), B(\varepsilon)} \int \left\{ \log c(\tilde{\alpha}) - \frac{h(\varepsilon; \alpha, \tilde{\alpha})^{1+\sigma}}{1+\sigma} \right\} dF_\varepsilon \\
\text{s.t.} \quad \int Q(\varepsilon) B(\varepsilon; \alpha, \tilde{\alpha}) d\varepsilon = 0 \\
\quad \exp(\alpha + \varepsilon) h(\varepsilon; \alpha, \tilde{\alpha}) + B(\varepsilon; \alpha, \tilde{\alpha}) = y(\tilde{\alpha}) \quad \forall \varepsilon 
\end{cases} \]

price of insurance \[ \int_{E} Q(\varepsilon) d\varepsilon = \int_{E} dF_\varepsilon \]
Planner’s Problem: Ramsey

\[
\max_{\tau} \int W(\alpha) \left\{ \int u(c(\alpha, \varepsilon), h(\alpha, \varepsilon))dF_\varepsilon \right\} dF_\alpha
\]

s.t. \[
\int \int y(\alpha, \varepsilon)dF_\alpha dF_\varepsilon \geq \int \int c(\alpha, \varepsilon)dF_\alpha dF_\varepsilon + G
\]

where \( c(\alpha, \varepsilon) \) and \( h(\alpha, \varepsilon) \) are the solutions to

\[
\begin{aligned}
\max_{c(\alpha,\varepsilon), h(\alpha,\varepsilon), \mathcal{B}(\alpha,\varepsilon)} & \int \left\{ \log(c(\alpha,\varepsilon)) - \frac{h(\alpha,\varepsilon)^{1+\sigma}}{1+\sigma} \right\} dF_\varepsilon \\
\text{s.t.} & \int Q(\varepsilon)\mathcal{B}(\alpha,\varepsilon)d\varepsilon = 0 \\
& c(\alpha, \varepsilon) \leq y(\alpha, \varepsilon) - T(y(\alpha, \varepsilon); \tau) \quad \forall \varepsilon
\end{aligned}
\]

where \( y(\alpha, \varepsilon) \equiv \exp(\alpha + \varepsilon)h(\alpha, \varepsilon) + \mathcal{B}(\alpha, \varepsilon) \)
Baseline HSV Tax System: $T(y; \lambda, \tau) = y - \lambda y^{1-\tau}$

- Estimated on PSID data for 2000-2006
- Households with head / spouse hours $\geq 260$ per year
- Estimated value for $\tau = 0.151$, $R^2 = 0.96$
Baseline Wage Distribution

- Heavy Pareto-like right tail of labor earnings distribution (Saez, 2001)
- Assume Pareto tail reflects uninsurable wage dispersion
- $F_\alpha$: Exponentially Modified Gaussian $EMG(\mu, \eta^2, a)$
- $F_\varepsilon$: Normal $N\left(\frac{-v_{\varepsilon}}{2}, v_{\varepsilon}\right)$
- $\log(w) = \alpha + \varepsilon$ is itself EMG $\Rightarrow w$ is Pareto-Lognormal
- $\log(wh)$ is also EMG, given our utility function, market structure, and HSV tax system
Use micro data from the 2007 SCF to estimate $\alpha$ by maximum likelihood $\Rightarrow \alpha = 2.2$
Baseline Social Welfare Function

- Progressivity built into current tax system informative about society’s taste for redistribution

- Assume SWF takes the form

\[ W(\alpha) = \exp(-\theta \alpha) \]

- \( \theta \) controls taste for redistribution (e.g. \( \theta = 0 \) : utilitarian)

- Assume govt choosing a tax system in HSV class

\[ T(y) = y - \lambda y^{1-\tau} \]

- What value for \( \theta \) rationalizes observed choice for \( \tau \)?

- Empirically-Motivated SWF: \( \theta^{US} \) that solves \( \tau^*(\theta^{US}) = \tau^{US} \)
Social Welfare

- $\theta^{US}$ solves

$$-\eta^2 \theta^{US} + \frac{1}{a + \theta^{US}} = \frac{1}{a - 1 - \tau^{US}} + \eta^2 (1 - \tau^{US}) + \frac{1}{1 + \sigma} \left\{ 1 - \frac{1}{(1 - g^{US})(1 - \tau^{US})} \right\}$$

- $\theta^{US}$ is increasing in $\tau$ and $g$
- $\theta^{US}$ is decreasing in $\eta^2$ and $\sigma$

- Special case: If $F_\alpha$ is also Normal ($a \to \infty$),

$$\theta^{US} = -(1 - \tau^{US}) + \frac{1}{\eta^2} \frac{1}{1 + \sigma} \left\{ \frac{1}{(1 - g^{US})(1 - \tau^{US})} - 1 \right\}$$

- Use $\theta^{US}$ as baseline for welfare comparisons $\Rightarrow$ focus on relative efficiency of alternative tax systems
Calibration

- Frisch elasticity $= 0.5 \rightarrow \sigma = 2$
- Progressivity parameter $\tau = 0.151$ (HSV 2014)
- Govt spending $G$ s.t. $G/Y = 0.188$ (US, 2005)
- $\text{var}(\varepsilon) = 0.193$: estimated variance of insurable shocks (HSV 2013)
- $\text{var}(\alpha) = 0.273$: total variance of wages is 0.466
Numerical Implementation

- Maintain continuous distribution for $\varepsilon$
- Assume a discrete distribution for $\alpha$
- Baseline: 10,000 evenly-spaced grid points
- $\alpha_{\text{min}}$: $5 per hour (12\% \text{ of the average} = \$41.56)$
- $\alpha_{\text{max}}$: $3,075 per hour ($6.17m assuming 2,000 hours = 99.99\% \text{ percentile of SCF earnings distn.}$)
- Set $\mu$ and $\eta^2$ to match $E[e^{\alpha}] = 1$ and target for $\text{var}(\alpha)$ given $a = 2.2$
Wage Distribution
Quantitative Analysis: Benchmark

<table>
<thead>
<tr>
<th>Tax System</th>
<th>Tax Parameters</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>welfare</td>
</tr>
<tr>
<td>HSV$_{US}$</td>
<td>$\lambda$ 0.836 $\tau$ 0.151</td>
<td>–</td>
</tr>
<tr>
<td>Affine</td>
<td>$\tau_0$ −0.116 $\tau_1$ 0.303</td>
<td>−0.58</td>
</tr>
<tr>
<td>Cubic</td>
<td>$\tau_0$ −0.032 $\tau_1$ 0.126 $\tau_2$ 0.064 $\tau_3$ −0.003</td>
<td>0.05</td>
</tr>
<tr>
<td>Mirrlees</td>
<td></td>
<td>0.11</td>
</tr>
</tbody>
</table>
Quantitative Analysis: Benchmark

- Moving to affine tax system is welfare reducing
  ⇒ Increasing marginal rates more important than lump-sum transfers

- Moving to fully optimal system generates only tiny gains (0.1%)

- The optimal marginal tax rate is around 30%

- Almost no need for transfers
HSV Tax Function

Log Consumption

Hours Worked

Marginal Tax Rate

Average Tax Rate
Affine Tax Function

- Log Consumption
- Hours Worked
- Marginal Tax Rate
- Average Tax Rate

Mirrlees
Ramsey
Cubic Tax Function

Log Consumption

Hours Worked

Marginal Tax Rate

Average Tax Rate
Quantitative Analysis: Sensitivity

What drives the results?

1. Empirically-motivated SWF $\rightarrow$ Utilitarian SWF: $\theta = 0$

2. Eliminate insurable shocks: $\tilde{v}_\alpha = v_\alpha + v_\varepsilon$ and $\tilde{v}_\varepsilon = 0$

3. Wage distribution has thin Log-Normal right tail: $\alpha \sim N$
Sensitivity: Utilitarian SWF

- Utilitarian SWF $\Rightarrow$ stronger taste for redistribution
- Want higher tax rates and larger transfers
- Optimal HSV still better than optimal affine

<table>
<thead>
<tr>
<th>Tax System</th>
<th>Tax Parameters</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>welfare</td>
<td>$Y$</td>
</tr>
<tr>
<td>HSV$^US$</td>
<td>$\lambda : 0.836$</td>
<td>$\tau : 0.151$</td>
</tr>
<tr>
<td>HSV</td>
<td>$\lambda : 0.821$</td>
<td>$\tau : 0.295$</td>
</tr>
<tr>
<td>Affine</td>
<td>$\tau_0 : -0.233$</td>
<td>$\tau_1 : 0.452$</td>
</tr>
<tr>
<td>Mirrlees</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Sensitivity: No Insurable Shocks

- No insurable shocks ⇒ larger role for public redistribution
- Want higher tax rates and larger transfers
- Optimal HSV still better than optimal affine

<table>
<thead>
<tr>
<th>Tax System</th>
<th>Tax Parameters</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>welfare</td>
<td>$Y$</td>
</tr>
<tr>
<td>HSV$^{US}$</td>
<td>$\lambda : 0.836$</td>
<td>$\tau : 0.151$</td>
</tr>
<tr>
<td>HSV</td>
<td>$\lambda : 0.839$</td>
<td>$\tau : 0.192$</td>
</tr>
<tr>
<td>Affine</td>
<td>$\tau_0 : -0.156$</td>
<td>$\tau_1 : 0.360$</td>
</tr>
<tr>
<td>Mirrlees</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Utilitarian SWF + No insurable shocks
  ⇒ Lump-sum transfers more important
  ⇒ Optimal HSV worse than optimal affine
Sensitivity: Log-Normal Wage

- Log-normal distribution $\Rightarrow$ thin right tail
- Optimal HSV worse than optimal affine
- Optimal affine nearly efficient

<table>
<thead>
<tr>
<th>Tax System</th>
<th>Tax Parameters</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>welfare</td>
</tr>
<tr>
<td>HSV$^{US}$</td>
<td>$\lambda : 0.836$</td>
<td>$\tau : 0.151$</td>
</tr>
<tr>
<td>HSV</td>
<td>$\lambda : 0.826$</td>
<td>$\tau : 0.070$</td>
</tr>
<tr>
<td>Affine</td>
<td>$\tau_0 : -0.068$</td>
<td>$\tau_1 : 0.250$</td>
</tr>
<tr>
<td>Mirrlees</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Why Distribution Shape Matters

• Want high marginal rates at the top when (i) few agents face those marginal rates, but (ii) can capture lots of revenue from higher-income households.
Efficient Marginal Tax Rates: Sensitivity

![Graph showing the comparison of marginal tax rates across different benchmarks. The axes are labeled 'alpha' on the x-axis and 'Marginal Tax Rate' on the y-axis. The graph includes lines for Benchmark, Utilitarian, No Insurable Shocks, and Log-Normal, each distinguished by different colors and line styles.](image-url)
Extension: Coarse Grid

- Coarse grid ⇒ Mirrlees Planner can do much better
- Gives Mirrlees planner too much power if true distribution continuous

<table>
<thead>
<tr>
<th># of grid points</th>
<th>Welfare (relative to HSV)</th>
<th>Affine</th>
<th>Mirrlees</th>
<th>First Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td></td>
<td>-0.58</td>
<td>3.82</td>
<td>8.80</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>-0.58</td>
<td>1.17</td>
<td>8.79</td>
</tr>
<tr>
<td>1,000</td>
<td></td>
<td>-0.58</td>
<td>0.21</td>
<td>8.80</td>
</tr>
<tr>
<td>10,000</td>
<td></td>
<td>-0.58</td>
<td>0.11</td>
<td>8.80</td>
</tr>
<tr>
<td>100,000</td>
<td></td>
<td>-0.58</td>
<td>0.10</td>
<td>8.80</td>
</tr>
</tbody>
</table>
Extension: Coarse Grid
Extension: Type-Contingent Taxes

- Productivity partially reflects observable characteristics (e.g. education, age, gender)

- Some fraction of uninsurable shocks are observable: $\alpha \rightarrow \alpha + \kappa$

- Heathcote, Perri & Violante (2010) estimate variance of cross-sectional wage dispersion attributable to observables, $v_\kappa = 0.108$

- Planner should condition taxes on observables: $T(y; \kappa)$

- Consider two-point distribution for $\kappa$ (college vs high school)
Extension: Type-Contingent Taxes

- Significant welfare gains relative to non-contingent tax
- Conditioning on observables $\Rightarrow$ marginal tax rates of 20%

<table>
<thead>
<tr>
<th>System</th>
<th>Outcomes</th>
<th>wel.</th>
<th>Y</th>
<th>mar.</th>
<th>$TR/Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSV$^U_S$</td>
<td>$\lambda : 0.827, \tau : 0.151$</td>
<td>–</td>
<td>–</td>
<td>0.311</td>
<td>0.017</td>
</tr>
<tr>
<td>HSV</td>
<td>$\lambda^L : 0.988, \tau^L : 0.180$</td>
<td>1.34</td>
<td>4.64</td>
<td>0.212</td>
<td>0.043</td>
</tr>
<tr>
<td></td>
<td>$\lambda^H : 0.694, \tau^H : -0.059$</td>
<td></td>
<td></td>
<td></td>
<td>-0.061</td>
</tr>
<tr>
<td>Affine</td>
<td>$\tau_0^L : -0.140, \tau_1^L : 0.151$</td>
<td>1.39</td>
<td>5.20</td>
<td>0.199</td>
<td>0.126</td>
</tr>
<tr>
<td></td>
<td>$\tau_0^H : 0.095, \tau_1^H : 0.224$</td>
<td></td>
<td></td>
<td></td>
<td>-0.137</td>
</tr>
<tr>
<td>Mirrlees</td>
<td></td>
<td>1.46</td>
<td>5.20</td>
<td>0.200</td>
<td>0.103</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.136</td>
</tr>
</tbody>
</table>
Conclusions

- Moving from current HSV system to optimal affine system is welfare reducing
  - Increasing marginal rates more important than lump-sum transfers

- Moving from current HSV system to fully optimal system generates tiny welfare gains
  - Ramsey and Mirrlees tax schemes not far apart: can approximately decentralize SB with a simple tax scheme
  - Important to measure the gap in terms of allocations and welfare, not in terms of marginal tax rates

- Want to condition both transfers and tax rates on observables