Optimal Income Taxation: Mirrlees Meets Ramsey

Jonathan Heathcote
FRB Minneapolis and CEPR

Hitoshi Tsujiyama
Goethe University Frankfurt

Yeshiva University, November 30, 2015
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**: $T(y) = y - \lambda y^{1-\tau}$
   - function introduced by Feldstein (1969), Persson (1983), and Benabou (2000)
   - increasing marginal rates without transfers
   - $\tau$ indexes progressivity: $1 - \tau = \frac{1 - T'(y)}{1 - \frac{T(y)}{y}}$

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

- Best tax and transfer system in the HSV class typically better than the best affine tax system
  - More valuable to have marginal tax rates increase with income than to have lump-sum transfers
- Optimal tax system as well as associated welfare gains hinge on the choice of SWF, may be tiny gains

- Agents differ wrt unobservable productivity $\alpha$
- Planner only observes earnings $x = \exp(\alpha) \times h$
- Think of planner choosing $(c, x)$ for each $\alpha$ type
- 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. We use various SWFs in sensitivity
   - Utilitarian social welfare function as a benchmark
     ⇒ Strong desire for redistribution
   - Alternative SWF that rationalizes amount of redistribution embedded in observed tax system

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

• Standard static Mirrlees plus partial private insurance (quantitatively important)

• Heterogeneous individual labor productivity with two independent stochastic components

\[ \log w = \alpha + \varepsilon \]

• \( \varepsilon \) is privately-insurable, \( \alpha \) is not

  • Agents belong to large families
  • \( \alpha \) common across all members of a family
  • \( \varepsilon \) idiosyncratic, within-family insurance, same distribution in each family

• Planner sees neither component of productivity
Environment 2

- Common preferences

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

- Production linear in aggregate effective hours

\[ \int \int \exp(\alpha + \varepsilon)h(\alpha, \varepsilon)dF_\alpha dF_\varepsilon = \int \int c(\alpha, \varepsilon)dF_\alpha dF_\varepsilon + G \]
Limitations of the Framework

1. The extent of private risk-sharing is exogenous with respect to alternative tax systems

2. Labor supply is the only margin distorted by taxes

   - Skill investment also potentially important: Heathcote, Storesletten and Violante, 2013; Guvenen, Kuruscu and Ozkan, 2012; Krueger and Ludwig, 2013
Planner’s Problems

- Seeks to maximize SWF denoted $W(\alpha)$

- Only sees total family income
  \[ Y(\alpha) = \int \exp(\alpha + \varepsilon) h(\alpha, \varepsilon) dF_\varepsilon \]

- First Stage
  - Planner offers menu of contracts \{ $C(\tilde{\alpha}), Y(\tilde{\alpha})$ \}
  - Family heads draw idiosyncratic $\alpha$ and report $\tilde{\alpha}$

- Second Stage
  - Family members draw idiosyncratic $\varepsilon$
  - Family head tells each member how much to work
  - Total earnings must deliver $Y(\tilde{\alpha})$ to the planner
  - Must divide consumption $C(\tilde{\alpha})$ between family members
Nature of the Solution

• Planner cannot condition individual allocations on $\varepsilon$, given free within-family transfers
  
  • equally cheap for any family member to deliver income to the planner, and equally valuable to receive consumption

• Planner cannot take over private insurance

⇒ Distinct roles for public and private insurance
Planner’s Problem: Second Best

\[
\begin{align*}
\max_{C(\alpha), Y(\alpha)} & \quad \int W(\alpha) U(\alpha, \alpha) dF_\alpha \\
\text{s.t.} & \quad \int Y(\alpha) dF_\alpha \geq \int C(\alpha) dF_\alpha + G \\
& \quad U(\alpha, \alpha) \geq U(\alpha, \tilde{\alpha}) \quad \forall \alpha, \forall \tilde{\alpha}
\end{align*}
\]

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

\[
\begin{align*}
\left\{ \max_{\{c(\alpha,\tilde{\alpha},\varepsilon), h(\alpha,\tilde{\alpha},\varepsilon)\}} \int \left\{ \log(c(\alpha, \tilde{\alpha}, \varepsilon)) - \frac{h(\alpha, \tilde{\alpha}, \varepsilon)^{1+\sigma}}{1+\sigma} \right\} dF_\varepsilon \right. \\
\left. \quad \text{s.t.} \quad \int c(\alpha, \tilde{\alpha}, \varepsilon) dF_\varepsilon = C(\tilde{\alpha}) \\
& \quad \int \exp(\alpha + \varepsilon) h(\alpha, \tilde{\alpha}, \varepsilon) dF_\varepsilon = Y(\tilde{\alpha}) \right. 
\end{align*}
\]

Mirrlees solution can be decentralized with taxes at the family or individual level
Planner’s Problem: Ramsey

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

\[\text{s.t. } \int \int c(\alpha, \varepsilon) dF_\alpha dF_\varepsilon + G = \int \int \exp(\alpha + \varepsilon) h(\alpha, \varepsilon) dF_\alpha dF_\varepsilon\]

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

\[
\begin{aligned}
\max & \quad \left\{ \log c(\alpha, \varepsilon) - \frac{h(\alpha, \varepsilon)^{1+\sigma}}{1+\sigma} \right\} dF_\varepsilon \\
\text{s.t.} & \quad \int c(\alpha, \varepsilon) dF_\varepsilon = Y(\alpha) - T(Y(\alpha); \tau) \\
& \quad Y(\alpha) = \int \exp(\alpha + \varepsilon) h(\alpha, \varepsilon) dF_\varepsilon
\end{aligned}
\]
Social Preferences

- Assume SWF takes the form $W(\alpha; \theta) = \exp(-\theta \alpha)$
  - $\theta$ controls taste for redistribution
  - $W(\alpha; \theta)$ function could be micro-founded as a probabilistic voting model

- Nests standard SWFs used in the literature:
  - $\theta = 0$: Utilitarian [our benchmark]
  - $\theta = -1$: Laissez-Faire Planner
  - $\theta \to \infty$: Rawlsian
Empirically Motivated SWF

- Progressivity built into current tax system informative about politico-economic demand for redistribution

- Assume planner (political system) choosing tax system in HSV class: \( T(y) = y - \lambda y^{1-\tau} \)

- Assume planner has SWF in class \( W(\alpha; \theta) = \exp(-\theta \alpha) \)

- What value for \( \theta \) gives observed \( \tau \) as solution to Ramsey problem?

  - Let \( \tau^*(\theta) \) denote welfare-maximizing choice for \( \tau \) given \( \theta \)
  - Empirically Motivated SWF \( W(\alpha; \theta^*) \) s.t. \( \tau^*(\theta^*) = \tau^{US} \)
  - related to inverse optimum problem

- Ramsey planner with \( \theta = \theta^* \) choosing a tax and transfer scheme in the HSV class would choose exactly \( \tau^{US} \)
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 \( \tau = 0.161, R^2 = 0.96 \)
Calibration: 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 $EMG(\mu_{\alpha}, \sigma_{\alpha}^2, \lambda_{\alpha})$
- $F_{\varepsilon}$: Normal $N(\frac{-\sigma_{\varepsilon}^2}{2}, \sigma_{\varepsilon}^2)$
- $\log(w) = \alpha + \varepsilon$ is itself EMG $\Rightarrow w$ is Pareto log-normal
- $\log(wh)$ is also EMG, given our utility function, private insurance model, and HSV tax system
- Normal variance coefficient in the EMG distribution for log earnings: $\sigma_y^2 = \left(\frac{1+\sigma}{\sigma+\tau}\right)^2 \sigma_{\varepsilon}^2 + \sigma_{\alpha}^2$. 
Use micro data from the 2007 SCF to estimate $\alpha$ by maximum likelihood $\Rightarrow \lambda_\alpha = 2.2$ and $\sigma^2_y = 0.4117$
Calibration

- Frisch elasticity $= 0.5 \Rightarrow \sigma = 2$
- Progressivity parameter $\tau = 0.161$ (HSV 2014)
- Govt spending $G$ s.t. $G/Y = 0.188$ (US, 2005)
- Variance of normal component of SCF earnings + external evidence on importance of insurable shocks $\Rightarrow \sigma^2_\varepsilon = \sigma^2_\alpha = 0.1407$
  - Consistent with HSV 2014
  - Total variance of log wages is 0.488
Bottom of Wage Distribution

- Difficult to measure distribution of offered wages at the bottom, given selection into participation

- Low and Pistaferri (2015) estimate distribution of latent offered wages within a structural model in which workers face disability risk and choose participation

<table>
<thead>
<tr>
<th>Percentile Ratios</th>
<th>Model</th>
<th>LP</th>
</tr>
</thead>
<tbody>
<tr>
<td>P5/P1</td>
<td>1.48</td>
<td>1.48</td>
</tr>
<tr>
<td>P10/P5</td>
<td>1.24</td>
<td>1.20</td>
</tr>
<tr>
<td>P25/P10</td>
<td>1.44</td>
<td>1.40</td>
</tr>
</tbody>
</table>
Numerical Implementation

• Maintain continuous distribution for $\varepsilon$

• Assume a discrete distribution for $\alpha$

• Baseline: 10,000 evenly-spaced grid points

• $\alpha_{\text{min}}$: $2 per hour (5\% \text{ of the average} = $41.56$

• $\alpha_{\text{max}}$: $3,075 per hour ($6.17m assuming 2,000 hours = 99.99\text{th percentile of SCF earnings distn.}$

• Set $\mu_\alpha$ and $\sigma^2_\alpha$ to match $E[e^{\alpha}] = 1$ and target for $\text{var}(\alpha)$ given $\lambda_\alpha = 2.2$
Wage Distribution

\[ \text{Density} \times 10^{-3} \]

\[ \text{Wage (exp}(\alpha + \varepsilon)) \]
Quantitative Analysis

- U.S. tax system approximated by HSV with $\tau = 0.161$

- Focus on three optimal systems:
  1. HSV tax function: $T(y) = y - \lambda y^{1-\tau}$
  2. Affine tax function: $T(y) = \tau_0 + \tau_1 y$
  3. Mirrless tax function (second best allocation)
Quantitative Analysis: Benchmark

<table>
<thead>
<tr>
<th>Tax System</th>
<th>Tax Parameters</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSV\textsuperscript{US}</td>
<td>(\lambda: 0.839) (\tau: 0.161)</td>
<td>– (T'(y)) (TR/Y)</td>
</tr>
<tr>
<td>HSV</td>
<td>(\lambda: 0.817) (\tau: 0.330)</td>
<td>2.08 (-7.22) 0.466 0.063</td>
</tr>
<tr>
<td>Affine</td>
<td>(\tau_0: -0.259) (\tau_1: 0.492)</td>
<td>1.77 (-8.00) 0.492 0.279</td>
</tr>
<tr>
<td>Mirrlees</td>
<td></td>
<td>2.48 (-7.99) 0.491 0.213</td>
</tr>
</tbody>
</table>
Quantitative Analysis: Benchmark

- Optimal HSV better than optimal affine
  \[\Rightarrow\] Increasing marginal rates more important than lump-sum transfers
- Moving to fully optimal system generates substantial gains (2.5%)
- The optimal marginal tax rate is around 50%
HSV Tax Function

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

Graphs showing the relationships between HSV Tax Function and the parameters Log Consumption, Hours Worked, Marginal Tax Rate, and Average Tax Rate, with the parameter alpha (\(\alpha\)) on the x-axis and various scales on the y-axis.
Affine Tax Function

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

Mirrlees
Affine
Quantitative Analysis: Sensitivity

What drives the results?

1. Utilitarian SWF $\theta = 0$

   $\Rightarrow$ Various SWFs including Empirically motivated SWF

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: $\alpha \sim N$
Social Welfare

- Consider alternative SWFs:
  - $\theta = -1$: Laissez-Faire Planner
  - $\theta \to \infty$: Rawlsian

- Empirically motivated SWF: $W(\alpha; \theta^*)$ s.t. $\tau^*(\theta^*) = \tau^{US}$
  - Closed form expression for $\theta^*$!
    \[
    \sigma^2_\alpha \theta^* - \frac{1}{\lambda_\alpha + \theta^*} = -\frac{1}{\lambda_\alpha - 1 + \tau} - \sigma^2_\alpha (1 - \tau) + \frac{1}{1 + \sigma} \left\{ \frac{1}{(1 - g)(1 - \tau)} - 1 \right\}
    \]

- Simple in Normal case ($\lambda_\alpha \to \infty$)
  \[
  \theta^* = -(1 - \tau) + \frac{1}{\sigma^2_\alpha} \frac{1}{1 + \sigma} \left\{ \frac{1}{(1 - g)(1 - \tau)} - 1 \right\}
  \]

- $\theta^*$ increasing in $\tau$ and $g$
- $\theta^*$ declining in $\sigma$ and $\sigma^2_\alpha$
- $\theta^*$ increasing in $\lambda_\alpha$ (holding fixed $var(\alpha) = \sigma^2_\alpha + \frac{1}{\lambda_\alpha^2}$)
Social Welfare Functions

Relative Pareto Weight ($\exp(-\theta \alpha)$)

- **Laissez-Faire:** $\theta = -1$
- **Utilitarian:** $\theta = 0$
- **Empirically Motivated:** $\theta^* = -0.576$
Optimal tax system very sensitive to assumed SWF

Welfare gains moving from the current tax system to the optimal one can be tiny

Affine system works well when preference for redistribution is either very strong or very weak:
- In the first case, want large lump-sum transfers
- In the second, want lump-sum taxes

For intermediate tastes for redistribution ($\theta \in [-0.88, 0.16]$), HSV is better than affine
HSV vs Affine with Various SWFs

Taste for Redistribution ($\theta$)

-0.878 0.160

Welfare Gains (%)

HSV

Mirrlees

Affine
Sensitivity: No Insurable Shocks

- No insurable shocks $\Rightarrow$ larger role for public redistribution
- Want higher tax rates and larger transfers
- Optimal HSV worse than optimal affine
  $\Rightarrow$ Distinguishing insurable shocks from uninsurable shocks is important

<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&lt;sup&gt;US&lt;/sup&gt;</td>
<td>$\lambda : 0.842$ $\tau : 0.161$</td>
<td>–</td>
</tr>
<tr>
<td>HSV</td>
<td>$\lambda : 0.804$ $\tau : 0.383$</td>
<td>4.17</td>
</tr>
<tr>
<td>Affine</td>
<td>$\tau_0 : -0.283$ $\tau_1 : 0.545$</td>
<td>5.34</td>
</tr>
<tr>
<td>Mirrlees</td>
<td></td>
<td>5.74</td>
</tr>
</tbody>
</table>
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>$\lambda : 0.828$ $\tau : 0.161$</td>
<td>$-$</td>
</tr>
<tr>
<td>HSV</td>
<td>$\lambda : 0.813$ $\tau : 0.285$</td>
<td>0.88</td>
</tr>
<tr>
<td>Affine</td>
<td>$\tau_0 : -0.230$ $\tau_1 : 0.451$</td>
<td>2.19</td>
</tr>
<tr>
<td>Mirrlees</td>
<td></td>
<td>2.28</td>
</tr>
</tbody>
</table>
Why Distribution Shape Matters

- Want high top marginal rates when (i) few agents face those marginal rates, but (ii) can capture lots of revenue from higher-income households.

![Graph showing why distribution shape matters](image-url)
U-shaped Tax Rates with High $G$

Graphs showing the relationship between Log Consumption, Hours Worked, Marginal Tax Rate, and Average Tax Rate for different values of $g$. The graphs illustrate how changes in $g$ affect these economic indicators.
U-shaped Tax Rates with High $\theta$

Emp. Motivated
Utilitarian
$\theta = 0.5$
$\theta = 1$
Rawlsian

Log Consumption
Hours Worked
Marginal Tax Rate
Average Tax Rate

$\alpha$

$\alpha$

$\alpha$

$\alpha$
## Extension: Polynomial Tax Functions

<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>$\lambda$</td>
<td>$\tau$</td>
</tr>
<tr>
<td></td>
<td>0.839</td>
<td>0.161</td>
</tr>
<tr>
<td>Affine</td>
<td>$\tau_0$</td>
<td>$\tau_1$</td>
</tr>
<tr>
<td></td>
<td>–0.259</td>
<td>0.492</td>
</tr>
<tr>
<td>Cubic</td>
<td>$\tau_0$</td>
<td>$\tau_1$</td>
</tr>
<tr>
<td></td>
<td>–0.212</td>
<td>0.370</td>
</tr>
<tr>
<td>Mirrlees</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Affine Tax Function

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

Graphs showing the comparison between the Mirrlees and Affine tax functions for different values of $\alpha$. The graphs illustrate how log consumption, hours worked, marginal tax rate, and average tax rate change with respect to $\alpha$. The charts highlight the differences in tax policies under the two models.
Cubic Tax Function

Log Consumption vs. Hours Worked

Marginal Tax Rate vs. Average Tax Rate

Mirrlees vs. Cubic Tax Function
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 42%

<table>
<thead>
<tr>
<th>System</th>
<th>Outcomes</th>
<th>$\lambda$</th>
<th>$\tau$</th>
<th>$T'(y)$</th>
<th>$TR/Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSV$^{US}$</td>
<td>wel.</td>
<td>$0.834$</td>
<td>$0.161$</td>
<td>$0.319$</td>
<td>$0.015$</td>
</tr>
<tr>
<td>HSV</td>
<td>$\lambda^L : 1.069$, $\tau^L : 0.480$</td>
<td>$6.21$</td>
<td>$-2.80$</td>
<td>$0.416$</td>
<td>$0.147$</td>
</tr>
<tr>
<td></td>
<td>$\lambda^H : 0.595$, $\tau^H : 0.073$</td>
<td>$-0.019$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affine</td>
<td>$\tau^L_0 : -0.403$, $\tau^L_1 : 0.345$</td>
<td>$6.15$</td>
<td>$-2.53$</td>
<td>$0.421$</td>
<td>$0.420$</td>
</tr>
<tr>
<td></td>
<td>$\tau^H_0 : -0.032$, $\tau^H_1 : 0.452$</td>
<td>$0.008$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mirrlees</td>
<td>$6.54$</td>
<td>$-2.53$</td>
<td>$0.418$</td>
<td>$0.368$</td>
<td></td>
</tr>
</tbody>
</table>

Note: HSV and Affine are two different models, and HSV$^{US}$ is a specific case of HSV.
Conclusions

- Debate on the structure of labor income taxation: how to balance redistribution versus distortions to labor supply

- Optimal HSV system better than optimal affine system
  
  - Increasing marginal rates more important than lump-sum transfers

- Welfare gains moving from the current tax system to the optimal one hinge on the choice of SWF, may be tiny

- 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