Optimal Income Taxation: Mirrlees Meets Ramsey

Jonathan Heathcote
Minneapolis Fed

Hitoshi Tsujiyama
University of Minnesota and Minneapolis Fed

ASSA Meetings, January 2012

How should we tax earnings?

- Key question in public finance: how to maximize public insurance while minimizing production distortions?
- Tension between the two objectives:
  - Taxes highly distorting $\Rightarrow$ costly to redistribute $\Rightarrow$ want low taxes and transfers (US model)
  - Large gains from redistribution $\Rightarrow$ want high taxes and transfers (European model)
- We address the question of the optimal size of government following both the Ramsey and Mirrlees approaches
- We find redistribution is highly distortive ...
- ... but the gains from redistribution are also large

Contributions and Findings

Our contributions:

1. Calibrate fraction of wage variation due to publicly-non-observable differences in ability
2. Use standard macroeconomic preferences
3. Describe approximate decentralization of constrained-efficient allocations based on polynomial earnings tax function

Our findings:

1. Optimal marginal tax rate $\approx 50\%$: reduces aggregate output by $\approx 12\%$
2. A cubic earnings tax function approximately implements the second-best solution
3. Want to condition both transfers and tax rates on observables

Literature

- Mirrlees (1971), Diamond (1988), Saez (2001), Weinzierl (2010), Mankiw, Weinzierl and Yagan (2009), Diamond and Saez (2011)

The environment

- Standard static Mirrlees environment
- Log wage is the sum of two independently distributed components

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

- $F(\alpha)$ and $F(\varepsilon)$ are the respective distributions
- Agent's utility function is

$$U(c,h) = \frac{c^{1-\gamma}}{1-\gamma} - \psi \frac{h^{1+\sigma}}{1+\sigma}$$

- $W(\alpha, \varepsilon)$ is the social welfare function

Information structure

- First best: α and ε are both publicly observable
- Second best (Mirrlees): α is private information, ε is public, unrestricted non-linear taxes
- Third best (Ramsey): α is private, ε is public, restricted tax functions:
  1. Economy-wide earnings tax function
  2. ε-type-specific earnings tax function

Dynamics

We abstract from dynamics (Albanesi and Sleet, Fukushima 2010, Farhi and Werning 2010, Golosov, Troshkin and Tsyvinkski 2011)

- In a dynamic setting can make taxes history dependent - Mirrlees planner can do better
  - But Farhi and Werning find small gains to history-dependent taxation with utilitarian SWF
- Other challenges for dynamic models:
  1. Theoretically and computationally challenging $\Rightarrow$ hard to handle large number of values for productivity
  2. Hard to disentangle predictable wage changes from shocks
  3. Model for savings matters: Mirrlees planner can provide less insurance if agents can hide savings

Tax Authority Problem: First Best

$$\max_{c(\alpha,\varepsilon),h(\alpha,\varepsilon)} \quad \int\int U(c(\alpha,\varepsilon),h(\alpha,\varepsilon))W(\alpha,\varepsilon)dF(\alpha)dF(\varepsilon)$$

$$\text{s.t.} \quad \int\int c(\alpha,\varepsilon)dF(\alpha)dF(\varepsilon)+G \leq \int\int \exp(\alpha+\varepsilon)h(\alpha,\varepsilon)dF(\alpha)dF(\varepsilon)$$

where G is exogenous government consumption

Tax Authority Problem: Mirrlees

$$\max_{c(\alpha,\varepsilon),h(\alpha,\varepsilon)} \int\int U(c(\alpha,\varepsilon),h(\alpha,\varepsilon))W(\alpha,\varepsilon)dF(\alpha)dF(\varepsilon)$$

s.t.

$$\int\int c(\alpha,\varepsilon)dF(\alpha)dF(\varepsilon)+G \leq \int\int \exp(\alpha+\varepsilon)h(\alpha,\varepsilon)dF(\alpha)dF(\varepsilon)$$

$$U(c(\alpha,\varepsilon),h(\alpha,\varepsilon)) \geq U\left(c(\tilde{\alpha},\varepsilon),h(\tilde{\alpha},\varepsilon)\frac{\exp(\tilde{\alpha}+\varepsilon)}{\exp(\alpha+\varepsilon)}\right) \quad \forall\tilde{\alpha}\neq\alpha$$

Tax Authority Problem: Ramsey

- Consider earnings tax systems in class of polynomial functions:

$$T_\varepsilon^N(y) = \tau_\varepsilon^0 + \tau_\varepsilon^1 y + \tau_\varepsilon^2 y^2 + ... + \tau_\varepsilon^N y^N$$

- Let $\mathcal{C}(T_\varepsilon^N)$ define the set of competitive equilibrium allocations given T_ε^N

- Ramsey Problem:

$$\max_{T_\varepsilon^i \in \mathcal{T}_\varepsilon^i} \quad \int\int U(c(\alpha,\varepsilon), h(\alpha,\varepsilon))W(\alpha,\varepsilon)dF(\alpha)dF(\varepsilon)$$

$$\text{s.t.} \quad \{c(\alpha,\varepsilon), h(\alpha,\varepsilon)\} \in \mathcal{C}(T_\varepsilon^N)$$

Baseline Parameterization 1

- Log consumption: $\gamma = 1$
- Equal weights on c and h: $\psi = 1$
- Linear taxation: $T^i_\varepsilon(y) = \tau^1 y$
- Unit average productivity: $\int \exp(\alpha) dF(\alpha) = \int \exp(\varepsilon) dF(\varepsilon) = 1$

Given these assumptions:

$$
\begin{aligned}
h(\alpha, \varepsilon) &= 1 \\
y(\alpha, \varepsilon) &= \exp(\alpha + \varepsilon) \\
c(\alpha, \varepsilon) &= (1 - \tau^1) \exp(\alpha + \varepsilon) \\
\int \int \exp(\alpha + \varepsilon) h(\alpha, \varepsilon) dF(\alpha) dF(\varepsilon) &= 1 \\
G &= \tau^1
\end{aligned}
$$

Note: hours and output are independent of σ and τ^1

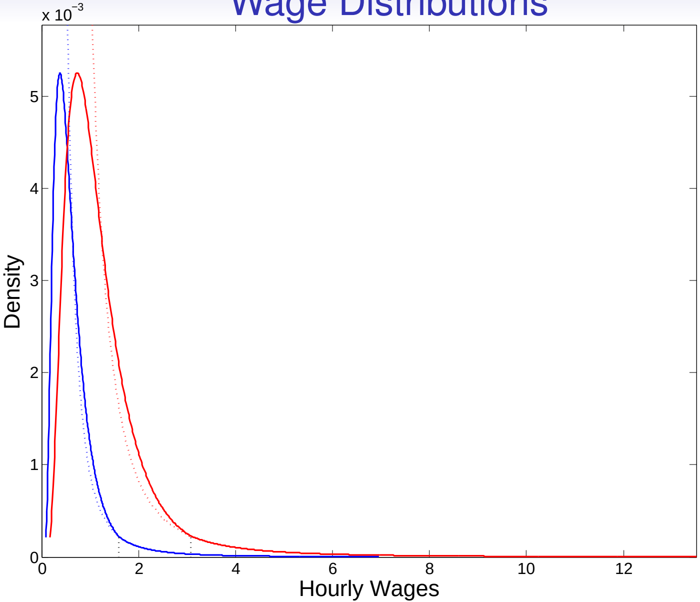
Parameterization 2

- Frisch elasticity $= 0.5$
  - $\sigma = 2$
- Gov C + Gov I $= 18.8\%$ of GDP in 2005
  - Corresponds to $\tau^1 = G = 0.188$
  - Hold G fixed in absolute terms across all tax schemes
- Variance log male wages 0.499 in 2005 (HPV 2010)
- Residual variance 0.389 (controlling for age, education, household composition)
  - $var(\alpha) = 0.389$
  - $var(\varepsilon) = 0.110$

Wage Distributions

- Assume 2 point equal-weight distribution for ε : $\exp(\varepsilon_H)/\exp(\varepsilon_L) = 1.94$
- Bounds for α : $\exp(\alpha) \in \left[\frac{\frac{1}{2}\times 5.15}{19.60}, \frac{200.56}{19.60}\right]$
  - $\$19.60$ is average hourly earnings in 2005 (BLS CES)
  - $\$5.15$ is Federal minimum wage in 2005
  - $\$200.56$ is earnings per hour (assuming 2000 hours) at 99.5^{th} percentile of 2005 earnings distribution (Piketty & Saez, Table B3)
- Distribution: Log-normal for $exp(\alpha) \leq x$, Pareto for $exp(\alpha) > x$
  - $x = 2.32$ (95% in log-normal range, Pareto above $45)
  - $a = 2.0$ (Pareto parameter estimated from Piketty & Saez)
- Grid for α : $1,000$ evenly spaced points (also consider $10,000$, 100, 15 and 2 points)

Wage Distributions



Computation

1. Mirrlees:
   - solve exactly on a discrete grid by forward iteration
   - much faster and more accurate than looking for approximate solution to Saez' functional equation
2. Ramsey: use Mirrlees allocations, run regression to find first guess for tax system

Calculations

- Take the linear tax system as the benchmark
- Relative to this evaluate:

1. Mirrless second best
2. Wide class of optimal functions conditional on restrictive functional forms:
    - lump-sum + linear
    - ... + quadratic term
    - ... + cubic term
3. Experiment with / without $\varepsilon-$type specific coefficients

Statistics

1. Welfare gain relative to proportional tax scheme
2. Welfare gain conditional on ε type
3. Efficiency gain (welfare gain for agent working average hours and consuming average consumption)
4. Output gain
5. Average earnings-weighted marginal tax rate
6. Share of output devoted to G versus τ^0

No Type-Specific Taxation

Tax system				welfare			
τ^0	τ^1	τ^2	τ^3	overall	level	ε_L	ε_H
0	0.188	-	-	0	0	0	0
0	0.053	0.108	-0.005	4.49	-3.74	8.07	1.03
-0.268	0.534	-	-	12.25	-2.59	32.16	-4.66
-0.267	0.533	0.000	-	12.25	-2.61	32.14	-4.64
-0.258	0.499	0.018	-0.001	12.36	-3.00	32.01	-4.36
Second Best (Mirrlees)				16.48	0.02	81.57	-25.28
First Best				58.60	39.36	164.81	-5.01

Welfare Results, Baseline Calibration, No type-contingent taxes

- Lump-sum component key to big welfare gains
- Quadratic term adds nothing to welfare
- Cubic term increases welfare gain, but still far from Second Best
- Relative to proportional tax scheme, Ramsey schemes imply large efficiency losses (3%)

Baseline Calibration, No type-contingent taxes

Tax system				Outcomes				
τ^0	τ^1	τ^2	τ^3	welfare	Y	mar. tax	G/Y	τ^0/Y
0	0.188	-	-	0	0	0.188	0.188	0
0	0.053	0.108	-0.005	4.49	-5.75	0.298	0.199	0
-0.268	0.534	-	-	12.25	-14.65	0.534	0.220	0.314
-0.258	0.499	0.018	-0.001	12.36	-14.71	0.534	0.220	0.302
Second Best (Mirrlees)				16.48	-11.39	0.496	0.212	0.244
First Best				58.60	22.23	0	0.152	0.808

Baseline Calibration, No type-contingent taxes

- Relative to proportional tax scheme, Ramsey schemes imply large output losses (15%)
- Optimal marginal tax rates around 50%
- Bulk of tax revenue used for transfers, not spending

Results, Type-Contingent Taxes

Tax system				Welfare			
τ^0	τ^1	τ^2	τ^3	Overall	Level	ε_L	ε_H
0	0.188			0	0	0	0
-0.258	0.499	0.018	-0.001	12.36	-3.00	32.01	-4.36
-	-0.195 0.385	-	-	5.55	0.00	47.06	-24.24
-0.413 -0.106	0.508	-	-	15.48	0.39	77.10	-24.70
-0.248	0.328 0.588	-	-	13.94	-2.24	51.20	-14.15
-0.389 -0.128	0.422 0.542	-	-	15.79	0.12	80.26	-25.63
-0.367 -0.111	0.318 0.499	0.079 0.019	-0.009 -0.001	16.00	-0.54	80.42	-25.41
Second Best (Mirrlees)				16.48	0.02	81.57	-25.28

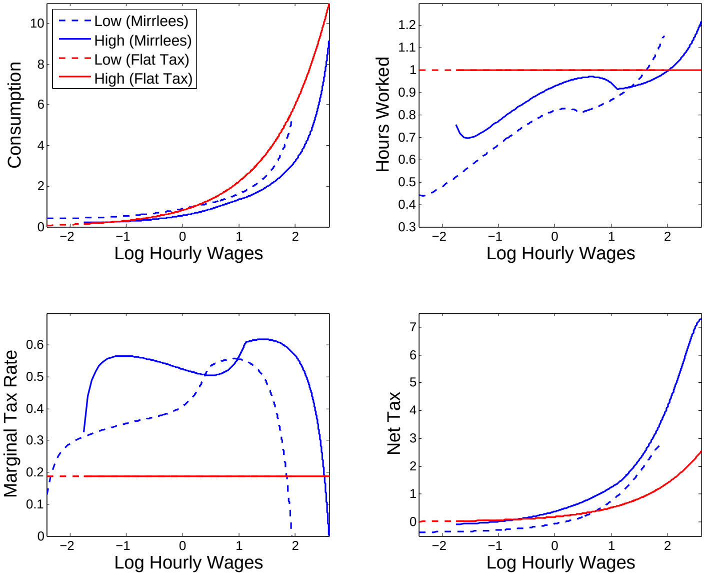
Results, Type-Contingent Taxes

- Significant welfare gains relative to non-contingent tax schemes
- With type-specific lump-sum and linear terms, approximately implement Mirrlees Second Best solution
- Gains from type-specific taxes largely from greater efficiency
- Want both τ^0 and τ^1 to be ε-type specific:
  - Redistribute across $\varepsilon-$ types via $\tau_H^1 > \tau_L^1$
  - If only linear taxes are available, set $\frac{1-\tau_H^1}{1-\tau_L^1} = \frac{\exp(\varepsilon_L)}{\exp(\varepsilon_H)}$
  - Induce higher hours from ε_H- type via $\tau_H^0 > \tau_L^0$

Results, Type-Contingent Taxes

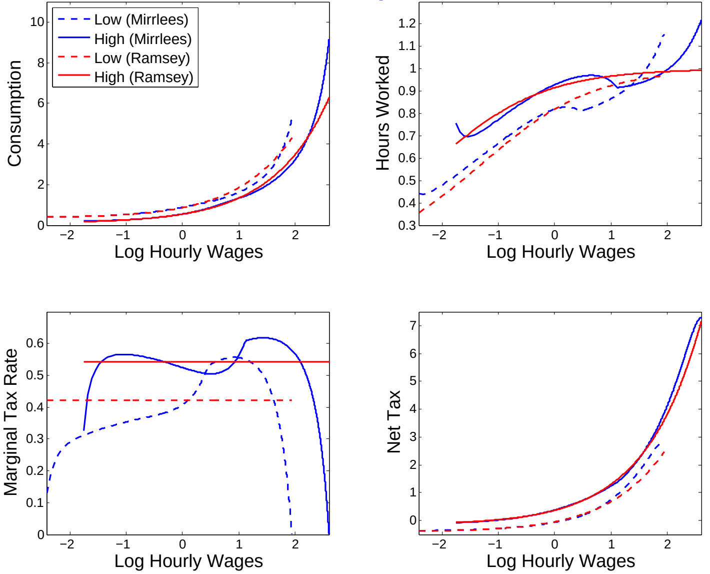
Tax system				Outcomes			
τ^0	τ^1	τ^2	τ^3	welfare	Y	mar. tax	τ^0/Y
0	0.188			0	0	0.188	0
-0.258	0.499	0.018	-0.001	12.36	-14.71	0.534	0.302
-	-0.195 0.385	-	-	5.55	0.00	0.188	0
-0.413 -0.106	0.508	-	-	15.48	-12.09	0.508	0.295
-0.248	0.328 0.588	-	-	13.94	-13.03	0.501	0.285
-0.389 -0.128	0.422 0.542	-	-	15.79	-11.79	0.506	0.293
-0.367 -0.111	0.318 0.499	0.079 0.019	-0.009 -0.001	16.00	-11.82	0.505	0.271
Second Best (Mirrlees)				16.48	-11.39	0.496	0.244

Flat Tax

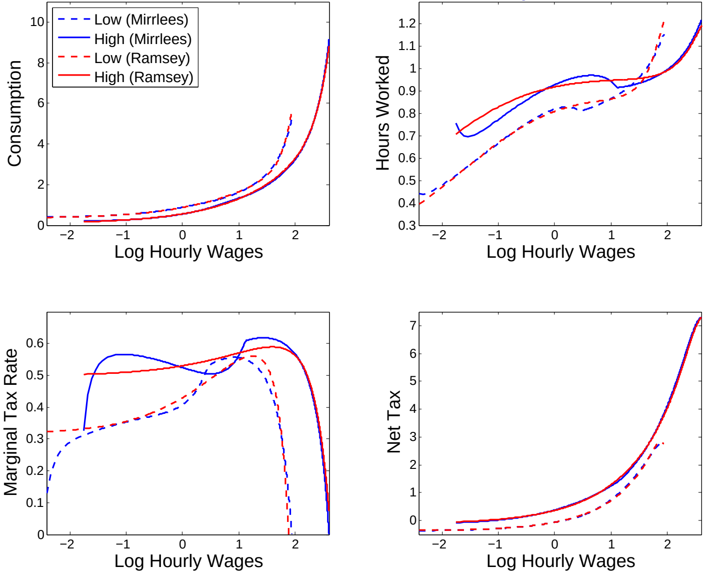


Low (Mirrlees)
High (Mirrlees)
Low (Flat Tax)
High (Flat Tax)
Consumption
Hours Worked
Marginal Tax Rate
Net Tax
Log Hourly Wages
Log Hourly Wages
Log Hourly Wages
Log Hourly Wages

Linear + Lump Sum Transfer



Cubic Tax-Transfer System



Sensitivity 1: Number of Grid Points for α

# of grid points	Welfare			
	Ramsey linear	Ramsey cubic	Mirrlees	First best
2	14.40	-	36.84	43.92
15	12.99	-	26.46	36.97
100	15.78	16.00	19.53	58.60
1,000	15.79	16.00	16.48	58.60
10,000	15.79	16.00	16.13	58.62

Sensitivity 1: Number of Grid Points for α

- Coarser grid $\Longrightarrow$ Mirrlees Planner can do better but Ramsey planner cannot
- Lump-sum + linear no longer gets you close to the second best
- Coarse approximations (Weinzierl, Fukushima, Golosov et al. etc) give Mirrlees planner too much power

Sensitivity 2: Shape of the Wage Distribution

Alternative Wage Distributions: Ramsey linear versus Mirrlees

	welfare		output		mar. tax	
	Ramsey	Mirrlees	Ramsey	Mirrlees	Ramsey	Mirrlees
Baseline	15.79	16.48	-11.79	-11.39	0.506	0.496
Log-normal	14.87	15.47	-11.10	-10.34	0.490	0.474
Wider-support	16.15	17.16	-12.03	-11.54	0.511	0.500
$v_\alpha = v_\alpha/2$	9.71	10.60	-5.76	-5.31	0.380	0.363
$v_\varepsilon = 0$	12.28	12.90	-14.69	-14.36	0.534	0.526

Sensitivity 2: Shape of the Wage Distribution

1. Shape of the top of the productivity distribution not that important
2. Should be raising tax rates as wage inequality increases
3. Sizable welfare gains for Mirrlees Planner when a component of productivity is observable

Sensitivity Analysis 3: Preferences

1. No risk aversion: $\gamma = 0$ (no redistribution motive, just want to finance G in the most efficient way)
2. More risk aversion: $\gamma = 2$
3. High Frisch elasticity: $\sigma = 1$
4. Inelastic labor supply (eliminate information friction, easy to achieve first best)

Sensitivity Analysis 3: Preferences

Alternative Preferences: Ramsey Linear versus Mirrlees

	welfare		output		mar. tax	
	Ramsey	Mirrlees	Ramsey	Mirrlees	Ramsey	Mirrlees
Baseline	15.79	16.48	-11.79	-11.39	0.506	0.496
$\gamma = 0$	1.04	1.16	7.53	8.77	0.058	0.037
$\gamma = 2$	29.05	29.89	-12.21	-11.82	0.620	0.610
$\sigma = 1$	15.24	15.95	-12.03	-11.41	0.450	0.439
$\sigma \to \infty$	30.94	30.94	0	0	1.00	1.00

Sensitivity Analysis 3: Preferences

1. Quasi-linear preferences $\Rightarrow$ low marginal taxes (zero if G is small enough) $\Rightarrow$ output gains
2. More risk aversion $\Rightarrow$ strong incentive to redistribute $\Rightarrow$ much higher tax rates
3. High Frisch elasticity $\Rightarrow$ only slightly lower tax rates (larger level effect)
4. Inelastic labor supply $\Rightarrow$ large welfare gains from equalizing consumption

How to tag? Type-specific transfers versus type-specific tax rates

- Because α and ε are multiplicative, need $\tau_H^1 > \tau_L^1$ to achieve redistribution across ε types
  - If only linear taxes are available, set $\frac{1-\tau_H^1}{1-\tau_L^1} = \frac{\exp(\varepsilon_L)}{\exp(\varepsilon_H)}$
- Absent desire for redistribution ($\gamma = 0$) ε-specific transfers more important than $\varepsilon-$specific tax rates
  - Can induce the high ε type to work more hours without introducing a wedge in the FOC
- In the baseline model, want both τ^0 and τ^1 type-specific

Sensitivity Analysis 4: Social Welfare Function

1. No redistribution across ε-types (only insurance against α)
2. Eliminate need for public consumption
3. Rawlsian Social Welfare Function

Sensitivity Analysis 4: Social Welfare Function

Alternative Welfare Functions: Ramsey Linear versus Mirrlees

	welfare		output		mar. tax	
	Ramsey	Mirrlees	Ramsey	Mirrlees	Ramsey	Mirrlees
Baseline	15.79	16.48	-11.79	-11.39	0.506	0.496
No ε transfers	7.72	8.34	-12.82	-12.54	0.492	0.485
$G = 0$	20.28	20.86	-16.27	-15.94	0.472	0.464
Rawlsian	602.94	663.87	-30.27	-25.21	0.786	0.688

Sensitivity Analysis 4: Social Welfare Function

1. Absent redistribution across $\varepsilon-$types, rates are slightly lower, welfare gains half as large
2. Eliminating G does not lower tax rates much - larger welfare gains from tax reform
3. With Rawlsian SWF, tax rates, output losses and welfare gains are very high

Conclusions

1. Debate on the size of government: higher taxes do reduce output dramatically, but the associated transfers imply large welfare gains
2. Ramsey and Mirrlees tax schemes are not very far apart: can approximately decentralize Mirrlees solution with a very simple tax scheme
   - important to measure the gap in terms of allocations and welfare, not in terms of marginal tax rates
3. Want to condition both transfers and tax rates on observables