Taxation of Top Incomes

1 Overview

1. Standard Responses (Diamond-Saez JEP 2011)
4. Superstars (Scheuer-Werning 2015)

2 Standard Responses

- Diamond-Saez (JEP 2011)
- $u_\theta(c, y) = c - h_\theta(y)$, e.g. $h_\theta(y) = h(y/\theta)$ where $\theta$ is a skill type
- $c = y - T(y)$ after-tax income
- agents solve $\max_y y - T(y) - h_\theta(y)$ with FOC $h'_\theta(y) = 1 - T'(y)$
- optimal labor supply only depends on marginal tax rate (MTR)
- focus on top incomes: constant marginal tax rate $\tau$ above a given income threshold $\bar{y}$
- hence FOC gives $y_\theta(1 - \tau)$, income as function of top bracket MTR
- aggregate over all top bracket tax payers to get $y(1 - \tau)$, the average income reported by top earners as function of $1 - \tau$
- aggregate elasticity of top incomes w.r.t. $1 - \tau$ is
  \[
  \xi_1 = \frac{dy}{d(1 - \tau)} \frac{1 - \tau}{y}
  \]
  captures “real” effects of changes in $\tau$
• government solves

\[
\max_{T(y)} \int G(u_\theta) dF(\theta)
\]

s.t.

\[
\int T(y_\theta) dF(\theta) \geq T_0
\]

• look at revenue maximizing top MTR \( \tau \), i.e. put zero welfare weight on top earners: \( G'(u) \rightarrow 0 \) for \( u \rightarrow \infty \) (peak of the Laffer curve for the top bracket)

• total revenue from top bracket is \( T = \tau(y - \bar{y}) \)

• solve

\[
\max_\tau \tau(y(1 - \tau) - \bar{y})
\]

with FOC

\[
y - \bar{y} - \tau \frac{dy}{d(1 - \tau)} = 0
\]

mechanical effect of increase in \( \tau \) as well as behavioral effect

• can rewrite this as

\[
\frac{\tau}{1 - \tau} \varepsilon_1 = \frac{y - \bar{y}}{y}
\]

• top of the income distribution is well approximated by Pareto distribution \( H(y_\theta) = 1 - (\bar{y}/y_\theta)^\alpha \)

• this distribution has the property that

\[
y = \mathbb{E}[y_\theta] = \frac{\alpha \bar{y}}{\alpha - 1} \Rightarrow \frac{y - \bar{y}}{y} = \frac{1}{\alpha}
\]

• Hence, optimal top MTR is a simple function of parameters

\[
\tau^* = \frac{1}{1 + \alpha \varepsilon_1}.
\]

• E.g. can estimate empirically

\[
\alpha = \frac{y}{y - \bar{y}} = \frac{\mathbb{E}[y_\theta | y_\theta > \bar{y}]}{\mathbb{E}[y_\theta | y_\theta > \bar{y}] - \bar{y}}
\]

• in recent years in US, \( \alpha \approx 1.5 \)
remains ε₁. With ε₁ = .25, τ⁺ = 73%. With ε = .5, τ⁺ = 57%. With ε₁ = 1, τ⁺ = 40%.

Top MTR in US (including all taxes) is 42.5%, higher in Europe.

3 “Trickle Down”

• Stiglitz (JPubE 1982): two unobservable types i = L, H, high and low skill, equal mass
• both types choose how much to work li
• total output is given by F(lₐ, lₜ) with CRS, wages are given by
  \[ w_i = \frac{\partial F}{\partial l_i} \tag{1} \]
• if only income \( y_i = w_i l_i \) is observable, the Pareto problem becomes
  \[ \max_{c_i, l_i} \psi_L u(c_L, l_L) + \psi_H u(c_H, l_H) \]
  s.t. the resource constraint
  \[ \sum_i c_i \leq F(l_L, l_H), \]
  the incentive constraint
  \[ u(c_H, l_H) \geq u \left( c_L, \frac{w_L l_L}{w_H} \right) \]
  (assuming that we want to redistribute from H to L) and (1)
• Lagrangian
  \[ \mathcal{L} = \psi_L u(c_L, l_L) + \psi_H u(c_H, l_H) + \lambda \left[ F(l_L, l_H) - \sum_i c_i \right] + \mu \left[ u(c_H, l_H) - u \left( c_L, \frac{w_L l_L}{w_H} \right) \right] + \sum_i \eta_i \left[ w_i - F_i(l_L, l_H) \right] \]
• we are interested in the implicit top marginal tax rate, so take FOCs w.r.t. \( c_H \) and \( l_H \):
  \[ [\psi_H + \mu] u_c(c_H, l_H) = \lambda \]
  \[ [\psi_H + \mu] u_l(c_H, l_H) = -\lambda F_H(l_L, l_H) + \sum_i \eta_i F_{iH}(l_L, l_H) \]
• the FOCs w.r.t. \( w_i \) are
\[
-\mu u_l \left( c_L, \frac{w_L l_L}{w_H} \right) \frac{l_L}{w_H} + \eta_L = 0
\]
and
\[
\mu u_l \left( c_L, \frac{w_L l_L}{w_H} \right) \frac{w_L l_L}{w_H^2} + \eta_H = 0,
\]
which implies \( \eta_L < 0, \eta_H > 0 \)

• the implicit marginal tax rate ("wedge") \( \tau_H \) on type \( H \) is such that
\[
w_H (1 - \tau_H) = -\frac{u_l(c_H, l_H)}{u_c(c_H, l_H)} \Rightarrow \tau_H = 1 + \frac{u_l(c_H, l_H)}{u_c(c_H, l_H)w_H} < 1
\]

• substituting the above FOCs yields
\[
\tau_H = 1 + \frac{-\lambda F_H(l_L, l_H) + \sum_i \eta_i F_i H(l_L, l_H)}{\lambda F_H(l_L, l_H)} = \frac{\sum_i \eta_i F_i H(l_L, l_H)}{\lambda F_H(l_L, l_H)}
\]

• recall \( \eta_L < 0 \) and \( \eta_H > 0 \). Moreover, \( F_{HL} > 0 \) and \( F_{HH} < 0 \) under CRS and concavity (which implies complementarity), so we have \( \tau_H < 0 \).

• Top earners are subsidized at the margin because their labor raises the wages of lower earners. This moves the two wages closer together and thereby relaxes the binding incentive constraint, allowing for additional redistribution.

• compare to Mirrlees (1971) model with effectively linear technology
\[
F(l_L, l_H) = \theta_L l_L + \theta_H l_H
\]
Then wages \( w_i = \theta_i \) are fixed, \( F_{iH} = 0 \) for \( i = L, H \) and \( \tau_H = 0 \). Key difference here comes from general equilibrium effects through endogenous wages, "trickle down"

• so far, individuals’ sectoral choices are fixed: type \( L \) can only do low-skill work, type \( H \) only high-skill, both choose only intensive margin

• More generally, individuals can choose both the sector they work in (e.g. blue-versus white-collar) and how much to work in the chosen sector. They have a two-dimensional skill type (one skill for each sector), like in the Roy (1951) model widely used in the labor economics literature.

• Rothschild/Scheuer (QJE 2013) solve for the optimal income tax policy in such a model, but with a continuous, two-dimensional skill distribution. Key: trickle
down/GE effects still push towards lower top marginal tax rate than in a standard Mirrlees model with fixed wages, but less so than in a world without occupational choice (as in Stiglitz 1982).

- Intuition: Subsidizing the high-skill sector leads to more effort there, leading to lower high-skill wages and higher low-skill wages (as before). With occupational choice and continuous skill distributions, some individuals who were just indifferent between the two sectors will switch from the high- to the low-skill sector as a result. This works against the original GE effect, making trickle down less effective.

- Scheuer (AEJ: Policy 2014): One occupation is entrepreneurs/self-employed, the other is workers (hired by entrepreneurs). Trickle-down (job creation) motives for optimal policy disappear once we have targeted taxes for each occupation (e.g. separate income tax for entrepreneurs, corporate income tax).

4 Rent-seeking

4.1 Rothschild-Scheuer (REStud 2016)

- Simple example (fully general in paper)

- individuals can choose between 2 activities
  - productive/traditional: wages reflect social marginal product
  - rent-seeking: workers compete for a fixed rent $\mu$, social marginal product of effort is zero but wages are proportional to $\mu/E$, where $E$ is total aggregate effort in that sector

- each individual has skill for each sector $\theta, \phi$)

- paper solves full 2-dimensional screening problem with continuous type distribution, but consider example with just 2 types:
  - productive workers: $\theta = \phi = 1$ (can do both activities)
  - rent-seekers: $\theta = 0, \phi = \phi_R$ (can only rent-seek)

- total rent-seeking effort is $E = \phi_R e_R + \lambda_p e_P$, where $\lambda_p$ is the fraction of productive workers who go to the rent-seeking sector

- workers may...
– be indifferent if $E = \mu$, since then $\mu / E = 1$
– all work in traditional sector if $E > \mu$
– all work in rent-seeking sector if $E < \mu$

- preferences $u(c, e) = c - h(e)$
- can show that optimum involves interior equilibrium, so

$$E = \phi_R e_R + \lambda_p e_P = \mu \Rightarrow \lambda_p(e_R, e_P) = \frac{\mu - \phi_R e_R}{e_P}$$

- output is sum of rents and production, i.e.

$$\mu + (1 - \lambda_p(e_R, e_P))e_P = e_P + \phi_R e_R$$

- utilitarian optimum maximizes output net of effort cost

$$W = e_P + \phi_R e_R - h(e_P) - h(e_R)$$

- suppose the government can control effort but not occupational choice, e.g. through a nonlinear income tax schedule

- FOCs for effort are

$$h'(e_P) = 1 \text{ and } h'(e_R) = \phi_R,$$

which coincides with the agents facing no distortionary taxes

- can check that, indeed, $\lambda_p \in (0, 1)$ under conditions on parameters, so equilibrium is interior

- 0 tax on rent-seekers even though completely unproductive (Pigouvian tax would be 100%)

- taxing rent-seekers would attract workers into rent-seeking sector ($e_R$ would fall but $E = \mu$ regardless, so $\lambda_p$ would have to increase)

- rent-seekers are “indirectly” productive by crowding out workers, making sure that many of them remain in productive sector

- key: GE effects (sectoral shifts) can be very important in shaping optimal top MTRs
• if tax top earners at very high rate, discourage their effort and therefore reduce total E

• but this raises returns to rent-seeking $\mu/E$ and attracts individuals into the rent-seeking sector who were previously in the productive sector

• to limit this socially wasteful sectoral shift, should not tax the top earners too high even if all top earners are rent-seekers and completely unproductive

• See Rothschild-Scheuer (NBER working paper 2014) for fully general framework with $N$ activities and arbitrary patterns of externalities between them

4.2 Piketty-Saez-Stantcheva (AEJ: Policy 2014)

• Bargaining

• e.g. CEOs could be overpaid if they can influence their own pay

• e.g. suppose each type $\theta$ receives a fraction $\eta$ of his/her actual output $z$, allow for $\eta > 1$

• individuals can put effort into increasing $z$ or $\eta$, with preferences

$$u_\theta(c, z, \eta) = c - h_\theta(z) - k_\theta(\eta)$$

• call $b = \eta z - z$ bargained earnings

• $\mathbb{E}[b]$, average bargained earnings, must come at the expense of somebody. Assume that any gain from bargaining comes at the expense of everybody else uniformly

• nonlinear income tax on total earnings $y = \eta z = z + b$

• now we need to distinguish the “real” economic elasticity

$$\varepsilon_1 = \frac{dz}{d(1-\tau)} \frac{1-\tau}{z}$$

from the total elasticity

$$\varepsilon = \frac{dy}{d(1-\tau)} \frac{1-\tau}{y}$$
and the bargaining response
\[ \varepsilon_2 = \frac{db}{d(1-\tau)} \frac{1-\tau}{y} = s\varepsilon = \varepsilon - \varepsilon_1 \]

with
\[ s = \frac{db/d(1-\tau)}{dy/d(1-\tau)} \]

being the fraction of the behavioral response due to bargaining

- now the optimal top MTR becomes
\[ \tau^* = \frac{1 + \alpha \varepsilon_2}{1 + \alpha \varepsilon} = 1 - \frac{\alpha (z/y)\varepsilon_1}{1 + \alpha \varepsilon}, \]

where the numerator captures the Pigouvian (corrective) motive for taxation, making top earners internalize the bargaining externality that they impose on others

- of course, \( \varepsilon_1 \) versus \( \varepsilon_2 \) is hard to measure. The authors take a macro approach and try to get \( \varepsilon_2 \) as a residual.

- First, they run
\[ \log(\text{top 1\% income share}_{it}) = \beta_0 + \beta_1 \log(1 - \text{top MTR}_{it}) + \epsilon_{it} \]

from time series/cross-country regression

- from this, find total elasticity \( \varepsilon \approx .3 \)

- next, run
\[ \log(\text{real GDP}_{it}) = \beta_0 + \beta_1 \log(1 - \text{top MTR}_{it}) + \beta_2 t + \epsilon_{it} \]

to get “true” economic elasticity \( \varepsilon_1 \approx 0 \)

- hence, they infer \( \varepsilon_2 = \varepsilon - \varepsilon_1 \approx 0.3 \) really big

- as a result, the optimal top MTR can be close to 100%

- strong OLS assumption that any deviation of GDP growth from trend is uncorrelated with top MTR. But top MTR is endogenous to GDP, could be moving with third variables (e.g. recession in 80s when MTRs went down)
• also, residual approach has limitations since anything that does not get captured in regression for \( \varepsilon_1 \) gets mechanically accounted for as residual \( \varepsilon_2 \)

## 5 Superstars

See posted slides on Scheuer-Werning (2015)