Robust Bounds on Optimal Tax Progressivity

Anmol Bhandari[1] Jaroslav Borovička[2] Yuki Yao[3]

November 2025

[1]University of Minnesota and NBER

[2]New York University and NBER

[3]University of Kent

Large literature on optimal tax design

- theoretical framework: Ramsey (1927), Mirrlees (1971)
- applications: Diamond and Saez (2011), Golosov et al. (2016), Heathcote et al. (2017)

Finding: Optimal tax scheme should be much more progressive than the current U.S. tax system.

Key predictions depend on hard to measure objects

- distribution of earning potentials (labor productivity)
- distribution of preferences (labor supply elasticity)

Optimal tax design acknowledging uncertainty about distribution of individual characteristics

- build on decision theory under uncertainty to model welfare consequences of statistical uncertainty about type distributions with Mirrlees (1971)
- quantify uncertainty using information from **historical** data on incomes and elasticities

Key sources of uncertainty

- tails of productivity and preference distribution with scarce information relative to welfare implications

Main finding

- concerns for uncertainty call for substantially **lower** tax progressivity for high incomes
- taxes for below-incomes remain essentially unchanged

Framework

A continuum of households indexed with types type $s \sim F(s)$.

Households choose effort subject to an income tax function $T(y)$.

A utilitarian government chooses $T(y)$ to maximize social welfare.

- trades off redistributive motives and efficiency
- faces uncertainty about the type distribution $F(s)$

Next, we will

- start with 1-dimensional uncertainty: productivity distribution
- extend to multidimensional uncertainty: productivity and labor supply elasticity distribution

Given a labor income tax function $T(y)$, household of type z solves

$$\max_{c,n} U(c, n; z)$$

subject to the budget constraint

$$c = \underbrace{zn}_{y = zn} - T(zn).$$

Solution yields indirect utility function $\mathcal{U}(z; T)$.

The government is concerned that distribution $F(z)$ may be misspecified.

- it considers alternative distributions $\widetilde{F}(z)$ that are statistically close to $F(z)$

A measure of statistical distance is the relative entropy (Kullback–Leibler divergence)

$$\mathcal{E}(F, \widetilde{F}) = \int m(z) \log m(z)\, dF(z)$$

- $m(z) = \frac{d\widetilde{F}(z)}{dF(z)}$ is the Radon–Nikodým derivative of $\widetilde{F}$ with respect to F

We also characterize the problem for a broader class of Cressie and Read (1984) power divergence.

- these alter the penalization of diviations of alternative distributions in the tails

For a given benchmark F and entropy bound κ, the set of statistically close distributions is

$$\mathcal{F}(F,\kappa) = \left\{\widetilde{F} : \mathcal{E}(F,\widetilde{F}) \leq \kappa\right\}$$

- the set $\mathcal{F}(F,\kappa)$ is large and the government does not put a prior on that set

Design a tax function that performs well under any of the distributions in the set $\mathcal{F}(F,\kappa)$.

- Hansen and Sargent (2001a,b), and the broader literature on decision-making under ambiguity

A utilitarian government solves the problem

$$\max_{T} \int \psi(z)\mathcal{U}(z;T) \, dF(z) + V(G)$$

subject to

$$\int T(\mathcal{Y}(z;T)) \, dF(z) = G.$$

- $\psi(z)$ is a Pareto/Negishi weighting function, normalized to $\mathbb{E}[\psi] = 1$
- net tax revenue $T(\mathcal{Y}(z;T))$ redistributes and pays for government expenditures G

A robust utilitarian government solves the max-min problem

$$\max_{T} \min_{\widetilde{F}\in\mathcal{F}} \int \psi(z)\mathcal{U}(z;T) \qquad d\widetilde{F}(z) + V(G)$$

subject to

$$\int T(\mathcal{Y}(z;T)) \qquad d\widetilde{F}(z) = G.$$

- max-min: given tax function T, adverse nature searches for the 'worst-case' distribution in $\mathcal{F}$
- optimal tax function performs well relative to any distribution in $\mathcal{F}$

A robust utilitarian government solves the max-min problem

$$\max_{T} \min_{m:\widetilde{F}\in\mathcal{F}} \int \psi(z)\mathcal{U}(z;T)m(z)dF(z) + V(G)$$

subject to

$$\int T(\mathcal{Y}(z;T))m(z)dF(z) = G.$$

- utilitarian concern: low weight $m(z)$ on households with high contribution to welfare
- budgetary concern: low weight $m(z)$ on households with high contribution to the budget

It is easier to work with the Lagrangian formulation.

- let θ be the Lagrange multiplier associated with the entropy constraint

Reformulate the robust government problem as

$$\max_{T} \min_{\substack{m>0 \\ E[m]=1}} \int \psi(z)\mathcal{U}(z;T)m(z)\,dF(z) + V(G) + \theta \int m(z) \log m(z)\,dF(z)$$

subject to

$$\int T(\mathcal{Y}(z;T))m(z)\,dF(z) = G.$$

Theoretical analysis

Mechanism design approach

The robust government solves

$$\max_{c,y} \min_{m:\widetilde{F}\in\mathcal{F}} \int \psi(z)\, U\left(c(z), \frac{y(z)}{z}\right) m(z)\, dF(z) + V(G)$$

subject to incentive compatibility constraints

$$U\left(c(z), \frac{y(z)}{z}\right) \geq U\left(c(z'), \frac{y(z')}{z}\right) \qquad \forall z, z'$$

and the government budget constraint

$$\int (y(z) - c(z))\, m(z)\, dF(z) = G.$$

The robust government solves

$$\min_{m:\widetilde{F}\in\mathcal{F}} \max_{c,y} \int \psi(z)\, U\left(c(z), \frac{y(z)}{z}\right) m(z)\, dF(z) + V(G)$$

subject to incentive compatibility constraints

$$U\left(c(z), \frac{y(z)}{z}\right) \geq U\left(c(z'), \frac{y(z')}{z}\right) \qquad \forall z, z'$$

and the government budget constraint

$$\int (y(z) - c(z))\, m(z)\, dF(z) = G.$$

The minimax theorem allows switching of the order of optimization

Worst-case distribution in the entropy case

The worst-case distribution is given by $\widetilde{f}(z) = m(z) f(z)$ with

$$m(z) = \bar{m} \exp\left(-\frac{1}{\theta(\kappa)}\left[\psi(z)\mathcal{U}(z) + \mu T(y(z))\right]\right)$$

where $\theta(\kappa)$ is a Lagrange multiplier on the entropy constraint.

- utilitarian concern: lower weight on households with high welfare contribution $\psi(z)\mathcal{U}(z)$
- budgetary concern: lower weight on households who generate high tax revenue $T(y(z))$

Theoretical characterization of top marginal tax rates in a simple (but informative) case.

- quasilinear household utility

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

- 'Rawlsian' welfare weights: $\psi(z) = 0$ in the right tail
- benchmark distribution $F(z)$ Pareto with shape parameter α
- Relative entropy as divergence measure

Optimal marginal tax schedule is given by a modified Diamond (1998)–Saez (2001) formula

$$\frac{T'(y(z))}{1 - T'(y(z))} = \underbrace{(1+\gamma)}_{(A)} \underbrace{\frac{1 - \widetilde{F}(z)}{z\widetilde{f}(z)}}_{(B)}$$

- (*A*): adverse effect of taxes on labor supply via labor supply elasticity
- (*B*): tradeoff between labor supply distortion at z and revenue from taxing types above z

Optimal tax schedule

Without misspecification concerns ($\kappa = 0$)

$$\frac{T'(y(z))}{1-T'(y(z))} = \underbrace{(1+\gamma)}_{(A)} \underbrace{\frac{1-F(z)}{zf(z)}}_{(B)} = (1+\gamma)\frac{1}{\alpha}$$

- taxes at the top are nonzero and quantitatively possibly large: $\gamma = 2$, $\alpha = 2 \implies T'(y) \to 60\%$
- intuition: the tax revenue from types above z outweighs the labor supply distortion at z

With misspecification concerns ($\kappa > 0$)

$$\frac{T'(y(z))}{1 - T'(y(z))} = \underbrace{(1+\gamma)}_{(A)} \underbrace{\frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)}}_{(B)}$$

- household at a given z less consequential from a budget perspective than household above z
- $\widetilde{f}(z)$ tilted less than $1 - \widetilde{F}(z) \implies (B)$ decreases

With misspecification concerns, the tax schedule and distribution $\widetilde{F}(z)$ are determined jointly.

- Distribution $\widetilde{F}(z)$ pins down tax schedule by

$$\frac{T'(y(z))}{1-T'(y(z))} = (1+\gamma)\frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)}$$

- Tax schedule determines distribution $\widetilde{F}(z)$ by

$$\begin{aligned} T(y(z)) &= T(y(\underline{z})) + \int_{y(\underline{z})}^{y(z)} T'(\eta)d\eta \\ m(z) &= \bar{m}\exp\left(-\frac{\mu}{\theta}T(y(z))\right) \end{aligned}$$

The optimal tax schedule is a fixed point of this argument.

Theorem 1.1

Assume preferences are quasilinear and $\kappa > 0$. Then the marginal tax rate vanishes to zero at the top:

$$\lim_{z \to \infty} T'(y(z)) = 0. \tag{1.1}$$

Moreover, if the right tail of z is Pareto distributed with shape parameter α, then

$$\lim_{y \to \infty} \frac{d \log T'(y)}{d \log y} = -\frac{1}{2}. \tag{1.2}$$

- the limit and rate of decay are independent of other parameters of the model
- results hold for arbitrarily small amounts of uncertainty κ

Generalizations: Beyond the quasilinear Rawlsian case

Results carry over to

- general (isoelastic) separable utility

$$U(c,n) = \frac{c^{1-\rho}}{1-\rho} - \chi \frac{n^{1+\gamma}}{1+\gamma}$$

- general welfare weights (e.g., utilitarian planner)
- general class of power divergence functions of Cressie and Read (1984).

For example, for a utilitarian planner with $\psi(z) \equiv 1$ and isoelastic utility, we have

$$\begin{aligned} \lim_{y\to\infty} T'(y) &= 0, \\ \lim_{y\to\infty} \frac{d\log T'(y)}{d\log y} &= \min\left(\rho - 1, -\frac{1}{2}\right). \end{aligned}$$

- the distortion

$$m(z) = \bar{m}\exp\left(-\frac{1}{\theta}\left[\mathcal{U}(z) + \mu T(y(z))\right]\right)$$

may be dominated by the utilitarian concern when utility from consumption is close to linear

We study a general class of power divergence functions of Cressie and Read (1984).

$$\mathcal{E}_\eta\left(F, \widetilde{F}\right) = \mathbb{E}\left[\phi_\eta(m)\right] = \mathbb{E}\left[\frac{m^{1+\eta} - 1}{\eta(1+\eta)}\right].$$

- $\eta = 0$ yields entropy as a special case
- $\eta > 0$ is more tolerant to pronounced thinning of $\widetilde{F}$ (divergence convex increasing in m)
- $\eta < 0$ penalizes pronounced thinning of $\widetilde{F}$ more strongly (divergence convex decreasing in m)

This allows to impose control over how thin-tailed alternative distributions can be relative to F.

Top marginal tax rates for the quasi-linear case

- when $\eta \geq 0$, then the marginal tax rate at the top satisfies

$$\lim_{y \to \infty} T'(y) = 0$$

- when $\eta < 0$, then the marginal tax rate at the top is given by

$$\lim_{y \to \infty} T'(y) = \tau_\eta = \frac{1+\gamma}{1+\gamma+\widetilde{\alpha}} \qquad \text{with } \widetilde{\alpha} = \alpha - \frac{1+\gamma}{\gamma}\frac{1}{\eta} > \alpha$$

In the $\eta < 0$ case, the worst-case density is asymptotically Pareto with a thinner shape parameter $\widetilde{\alpha}$.

- We use this case to impose a notion of 'prior' knowledge on the permitted shapes of the tail.

Asymptotic results point at lower marginal taxes at the top.

- key driving force are budgetary concerns: the tradeoff between raising income and distortions
- strength of the effect depends on specific calibration

We now explore the quantitative impact.

- calibrate the amount of uncertainty the planner plausibly faces
- study the impact across the whole type distribution
- investigate alternative setups with uncertainty over productivity and labor supply elasticity distribution

Quantitative application

Preferences and technology

- isoelastic preferences: $U(c,n)=\frac{c^{1-\rho}}{1-\rho}-\upsilon\frac{n^{1+\gamma}}{1+\gamma}$ with $\rho=1$, $\upsilon=1$, $\gamma=2$
- government spending $V(G)=\bar{v}G$

Benchmark distribution F

- $\log z$ has exponentially modified Gaussian (EMG) distribution (Heathcote and Tsujiyama (2021))
- left tail of z distribution is lognormal (parameters μ, σ)
- right tail approximately Pareto (parameter α)

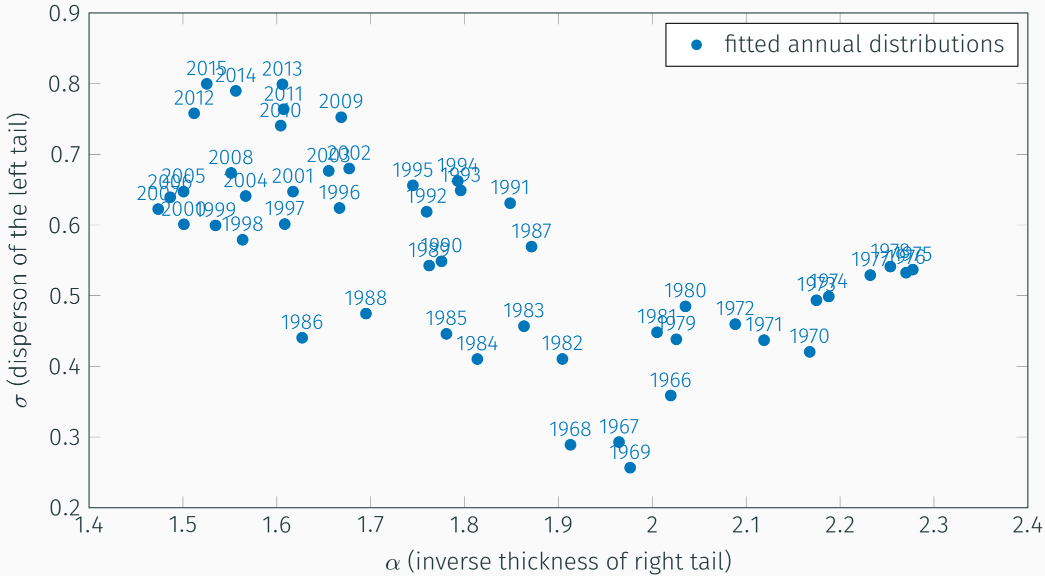
Entropy bound κ

- use time-series variation in observed income distributions (World Income Database)

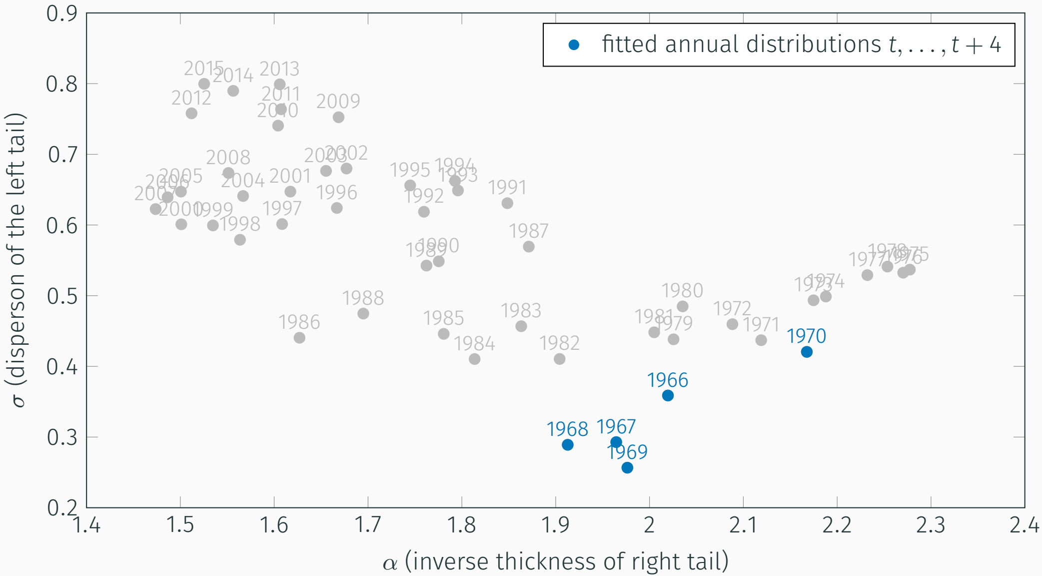
1. For each year t, we fit the EMG distribution to obtain parameters $(\mu_t, \sigma_t, \alpha_t)$.
2. For each 5-year window $\{t, \ldots, t+4\}$, we construct $\mathcal{F}(F_t, \kappa_t)$ as the set that
   - includes all fitted EMG distributions from years $\{t, \ldots, t+4\}$
   - has the smallest entropy radius κ_t
3. Baseline calibration uses the lowest $\kappa \in \{\kappa_t\}$.

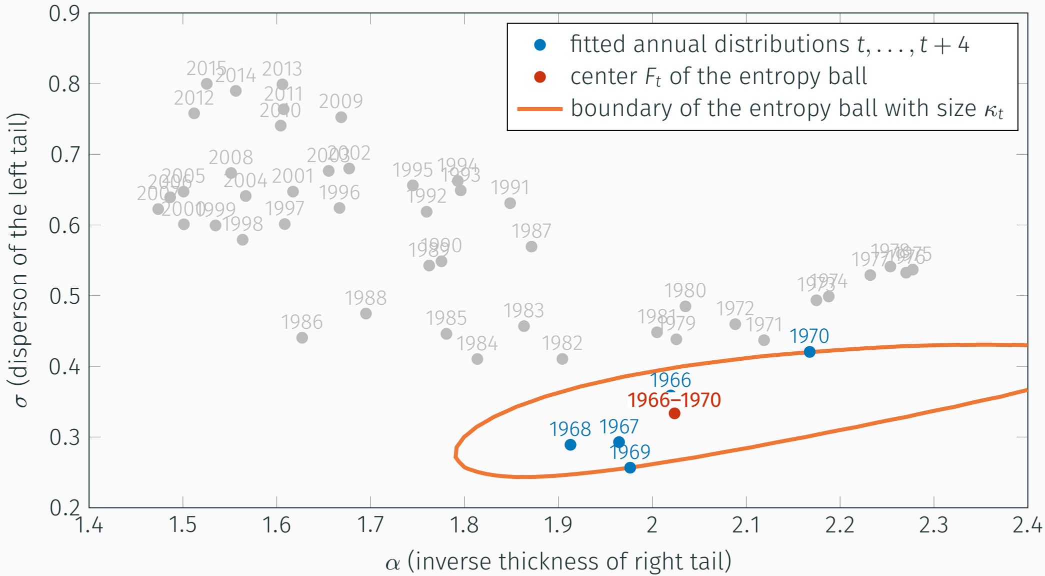
The set $\mathcal{F}(F_t, \kappa_t)$ is rich:

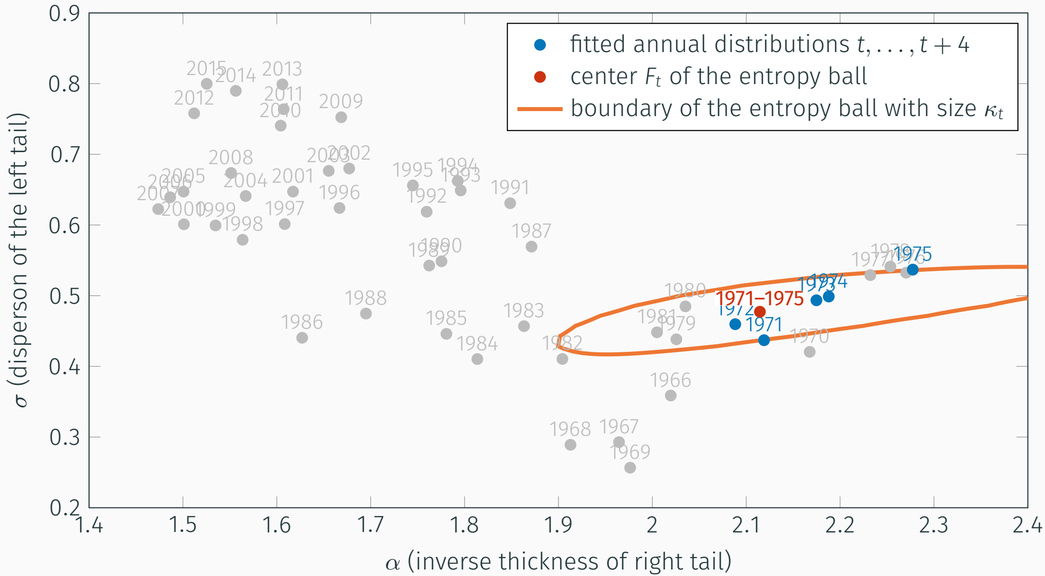
- it contains all distributions that are close to F_t
- not only the parameterized EMG family

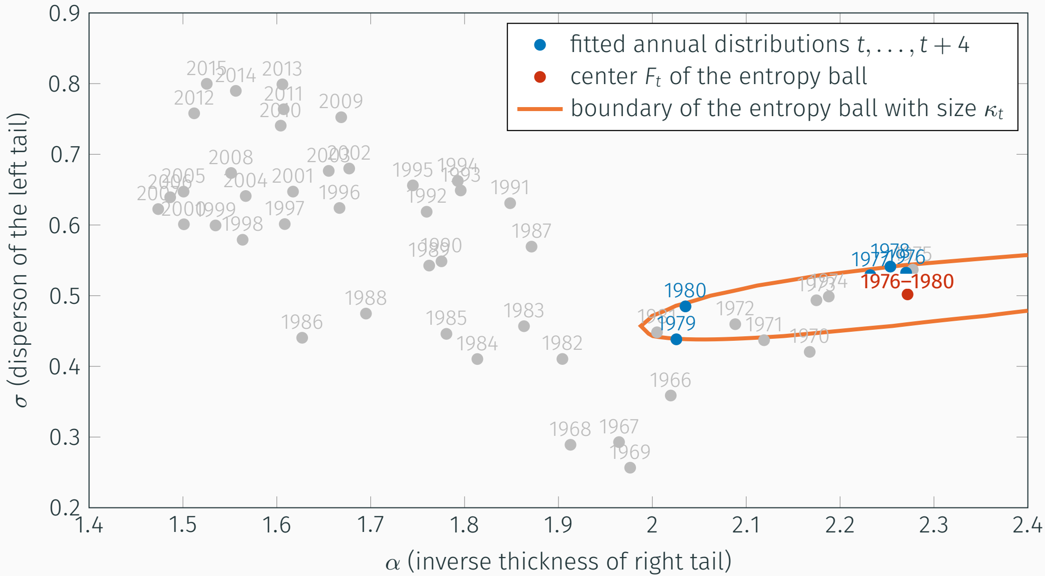

fitted annual distributions
σ (disperson of the left tail)
α (inverse thickness of right tail)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4
2015
2014
2013
2012
2011
2009
2008
2005
2004
2001
1995
1991
1992
1996
1997
1998
1987
1988
1986
1985
1984
1983
1982
1980
1972
1971
1970
1979
1966
1968
1967
1969

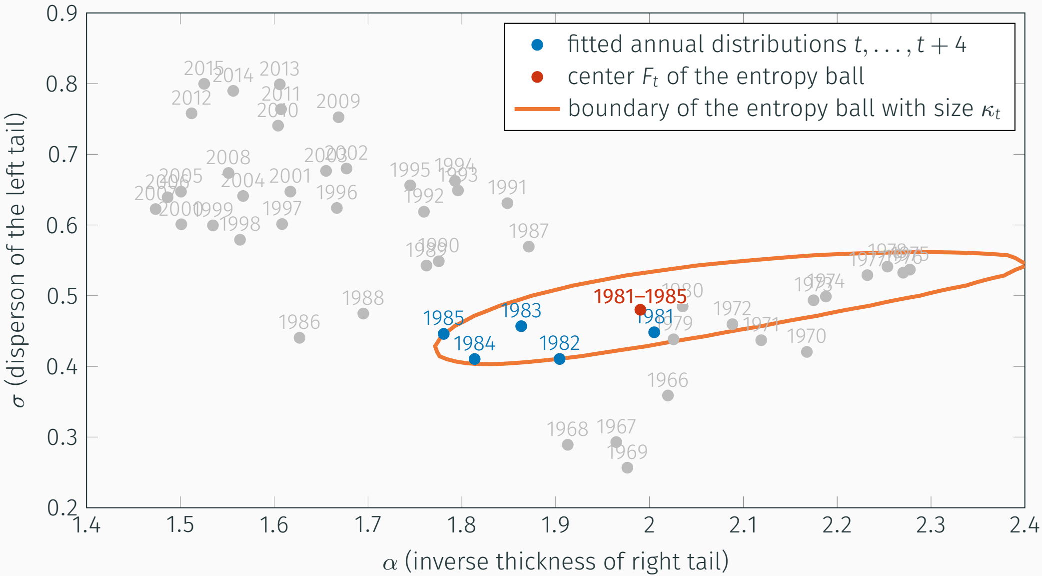
Quantifying uncertainty in income distributions



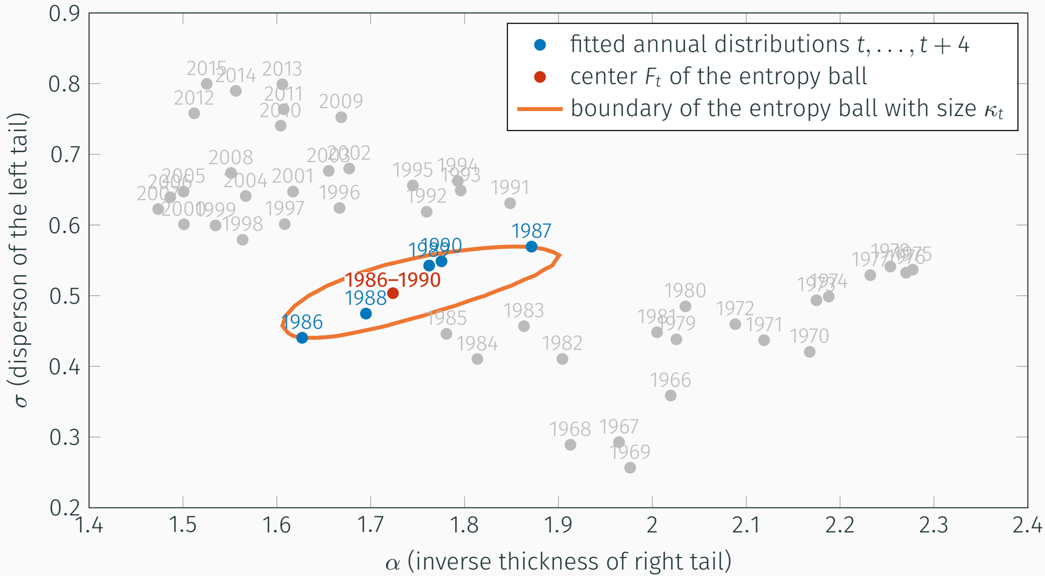

fitted annual distributions $t, \ldots, t+4$
center F_t of the entropy ball
boundary of the entropy ball with size κ_t
σ (disperson of the left tail)
α (inverse thickness of right tail)
1966
1966–1970
1967
1968
1969
1970
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4

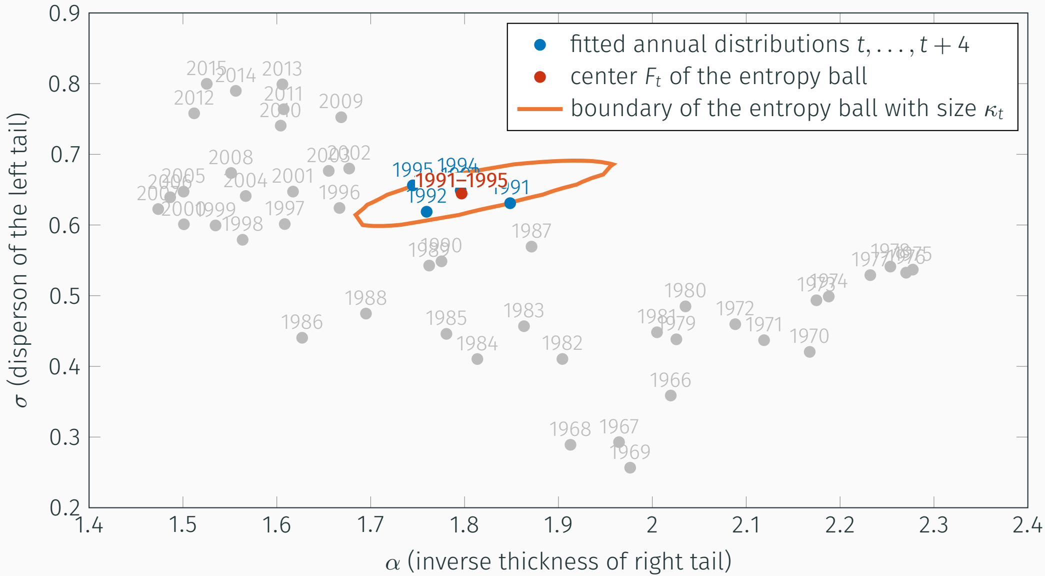

fitted annual distributions $t, \ldots, t+4$
center F_t of the entropy ball
boundary of the entropy ball with size κ_t
1971–1975
1971
1972
1973
1974
1975
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4
σ (disperson of the left tail)
α (inverse thickness of right tail)


fitted annual distributions $t, \ldots, t+4$
center F_t of the entropy ball
boundary of the entropy ball with size κ_t
σ (disperson of the left tail)
α (inverse thickness of right tail)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4
1976–1980
1980
1979
1978
1976
2015
2014
2013
2012
2011
2009
2008
2004
2001
1996
1997
1998
1995
1992
1991
1987
1988
1986
1985
1984
1983
1982
1972
1971
1966
1968
1967
1969

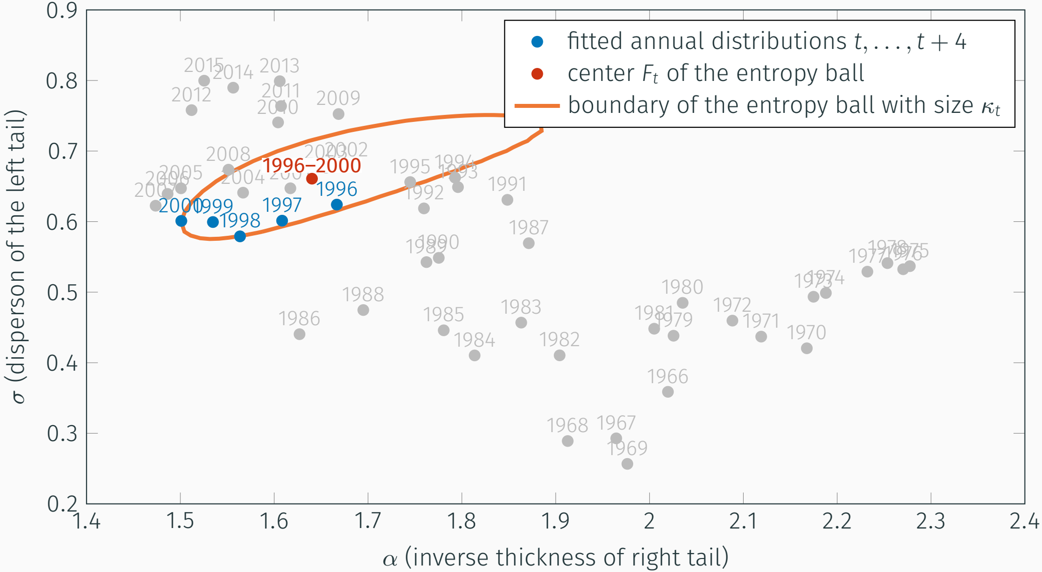

fitted annual distributions $t, \ldots, t+4$
center F_t of the entropy ball
boundary of the entropy ball with size κ_t
1981–1985
1985
1984
1983
1982
1981
2015
2014
2013
2012
2011
2009
2008
2005
2004
2001
1996
1997
1998
1995
1994
1992
1991
1987
1988
1986
1966
1968
1967
1969
1972
1971
1970
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4
σ (disperson of the left tail)
α (inverse thickness of right tail)

Quantifying uncertainty in income distributions

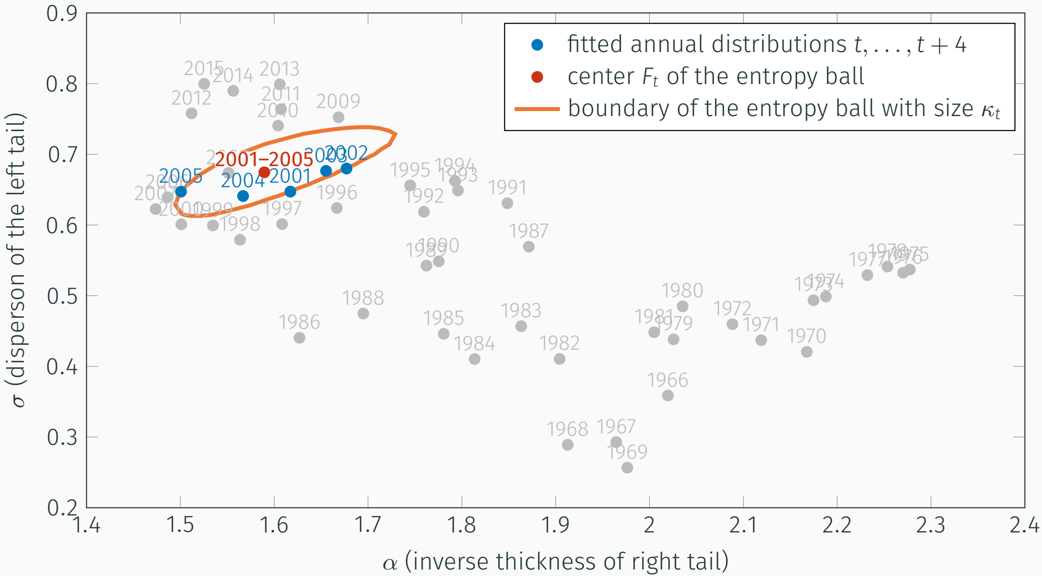


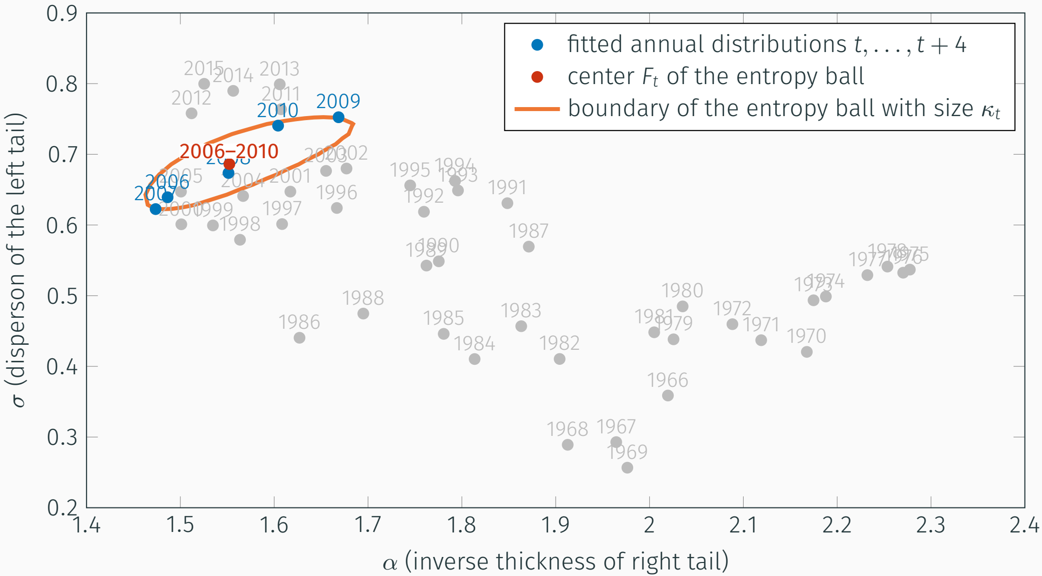

fitted annual distributions $t, \dots, t+4$
center F_t of the entropy ball
boundary of the entropy ball with size κ_t
1991–1995
σ (disperson of the left tail)
α (inverse thickness of right tail)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4

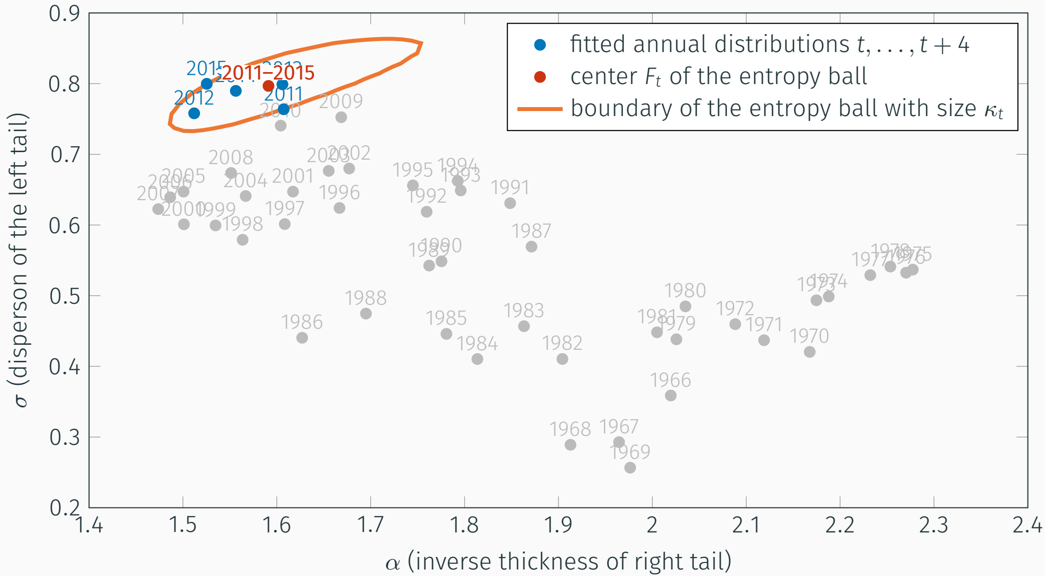
Quantifying uncertainty in income distributions

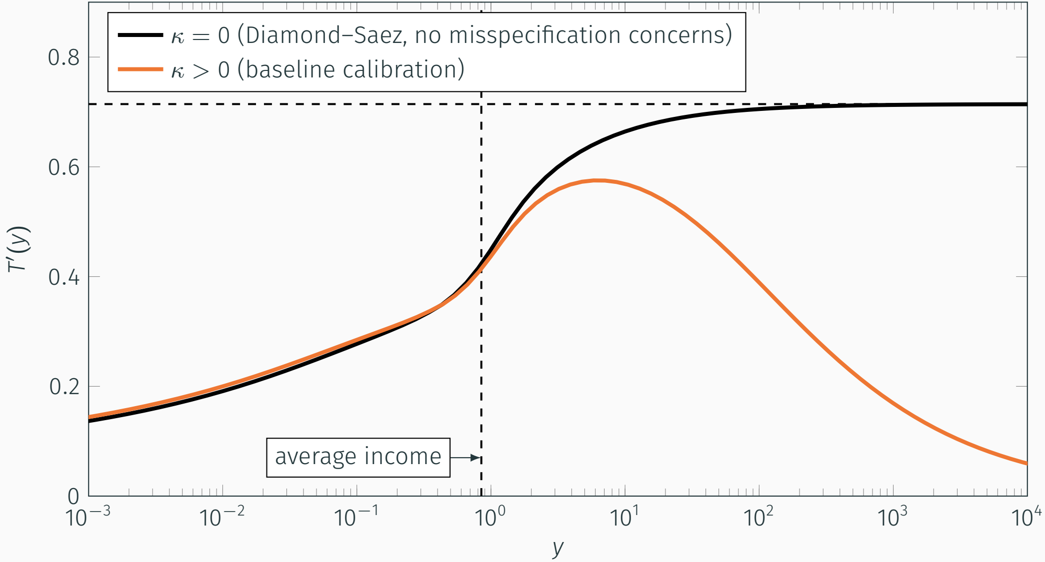


Quantifying uncertainty in income distributions

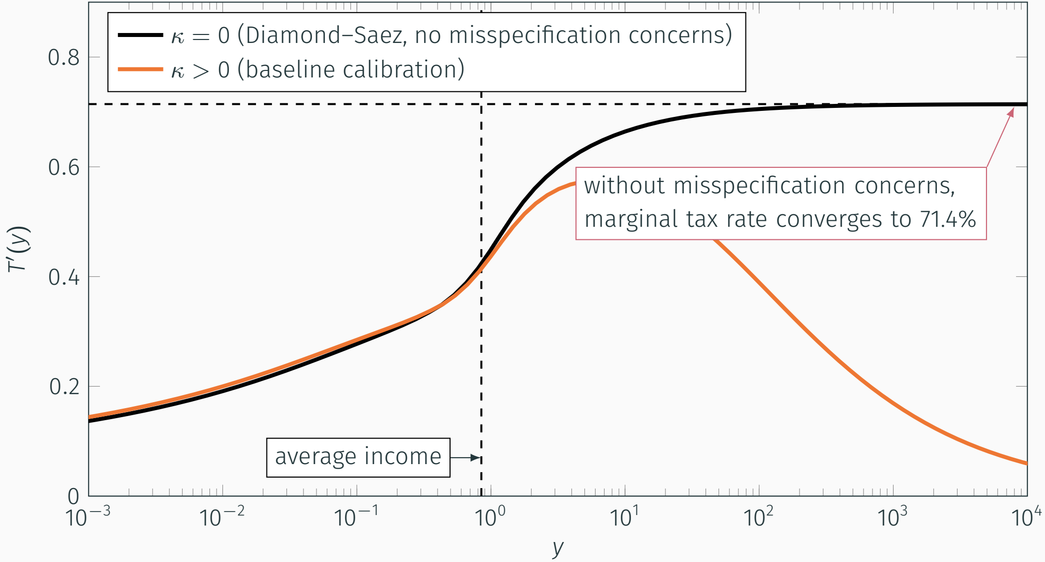



fitted annual distributions $t, \ldots, t+4$
center F_t of the entropy ball
boundary of the entropy ball with size κ_t
σ (disperson of the left tail)
α (inverse thickness of right tail)
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4
2006–2010
2009
2010
2006
2007
2015
2014
2013
2012
2011
2005
2004
2001
2002
2000
1999
1998
1997
1996
1995
1994
1993
1992
1991
1987
1990
1989
1988
1986
1985
1984
1983
1982
1981
1980
1979
1972
1971
1970
1974
1973
1977
1976
1975
1966
1968
1967
1969

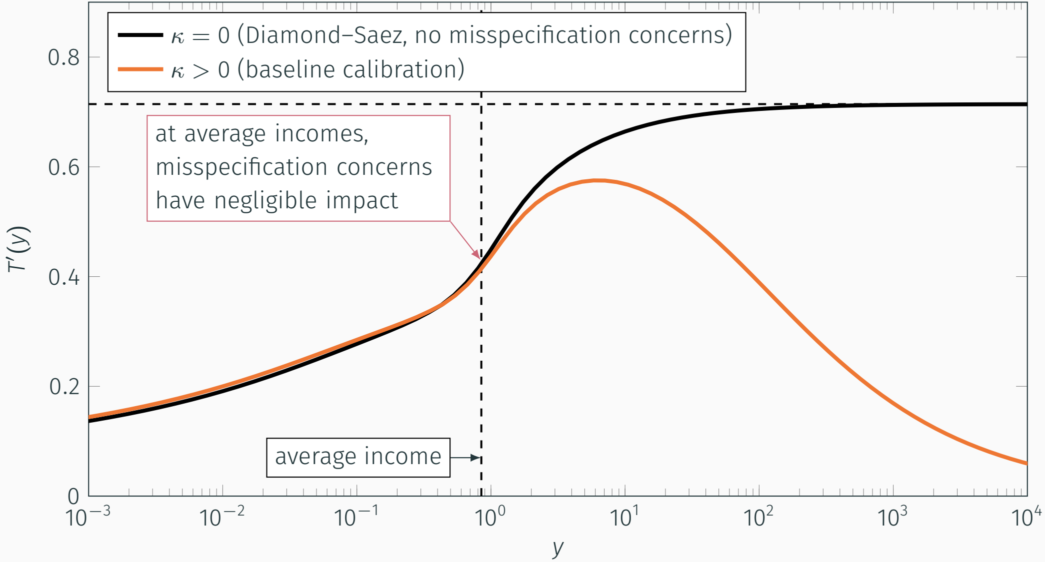

fitted annual distributions $t, \ldots, t+4$
center F_t of the entropy ball
boundary of the entropy ball with size κ_t
2011–2015
σ (disperson of the left tail)
α (inverse thickness of right tail)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4

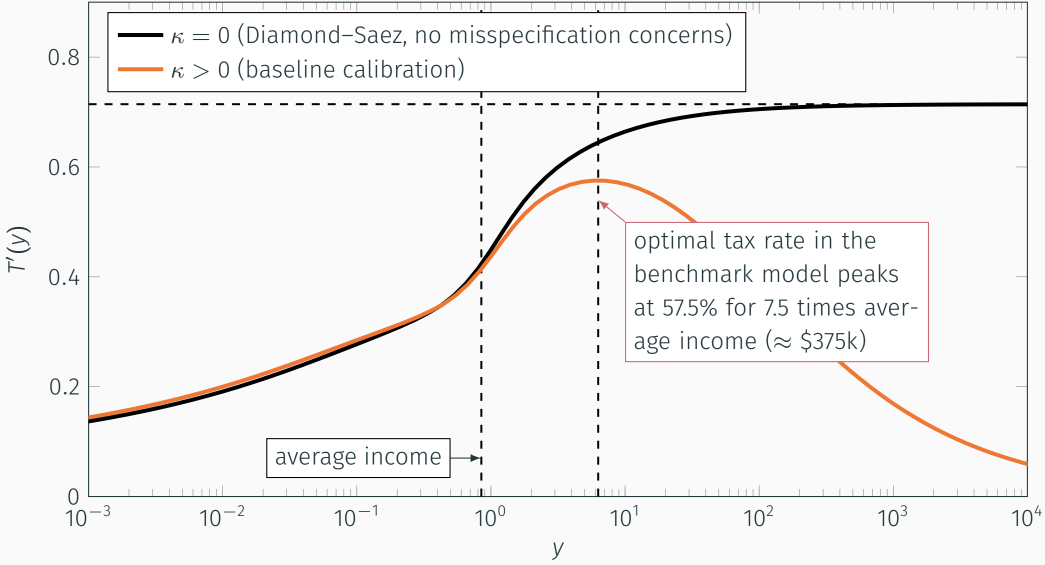

$\kappa = 0$ (Diamond–Saez, no misspecification concerns)
$\kappa > 0$ (baseline calibration)
0.8
0.6
0.4
0.2
0
$T'(y)$
average income
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
y

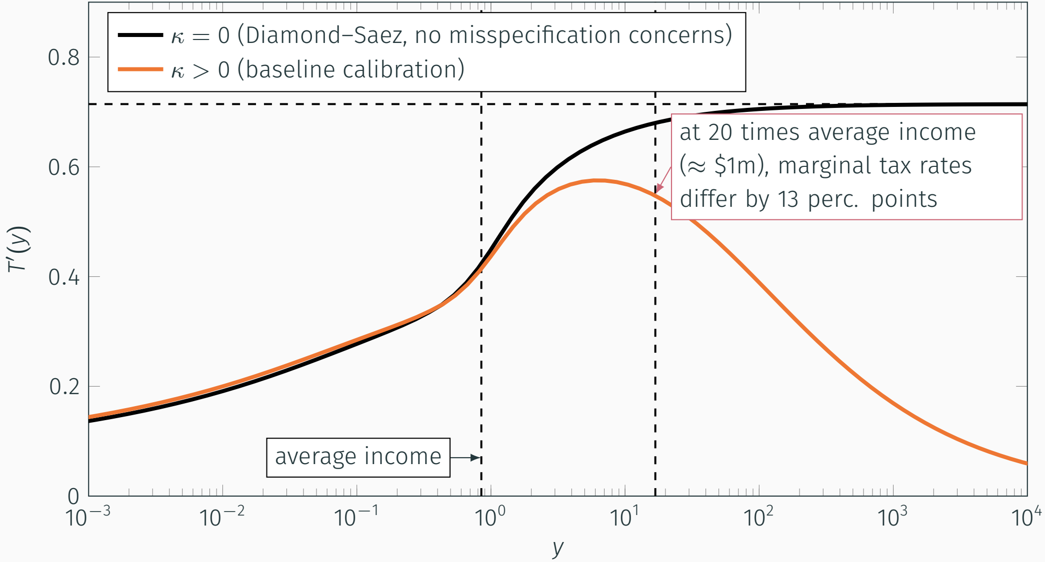
Optimal marginal tax schedules

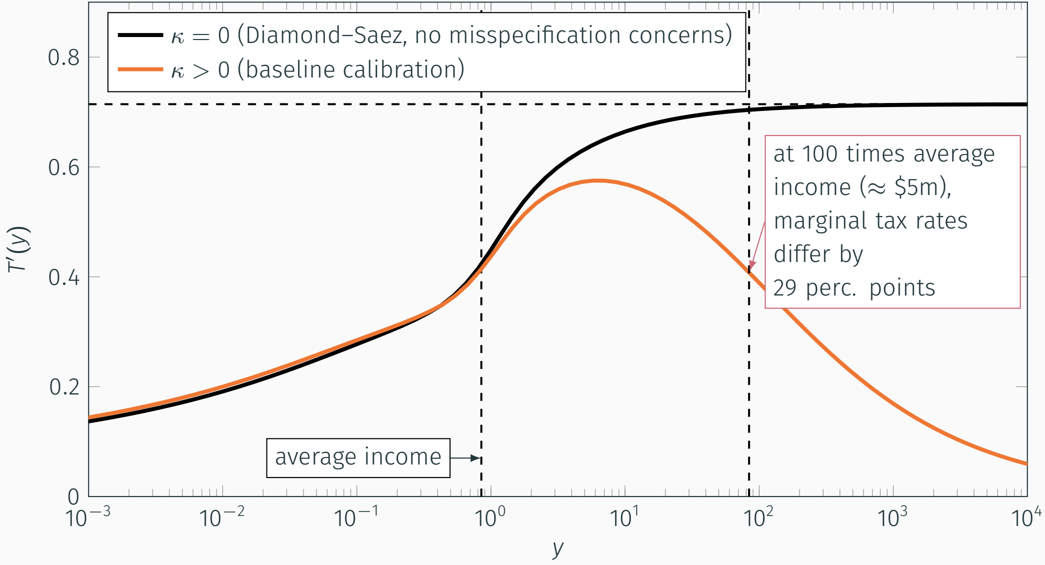


Optimal marginal tax schedules

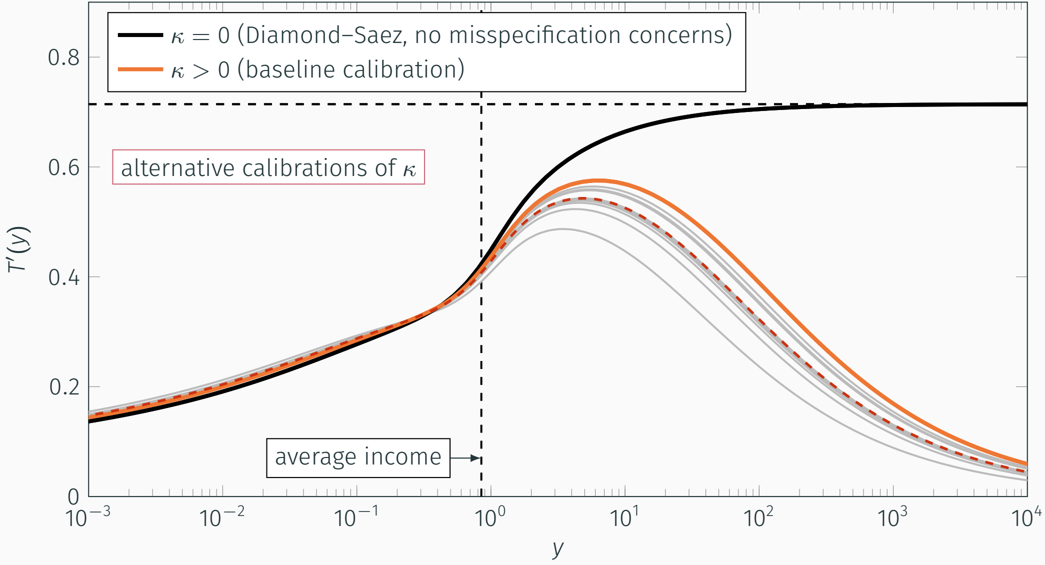


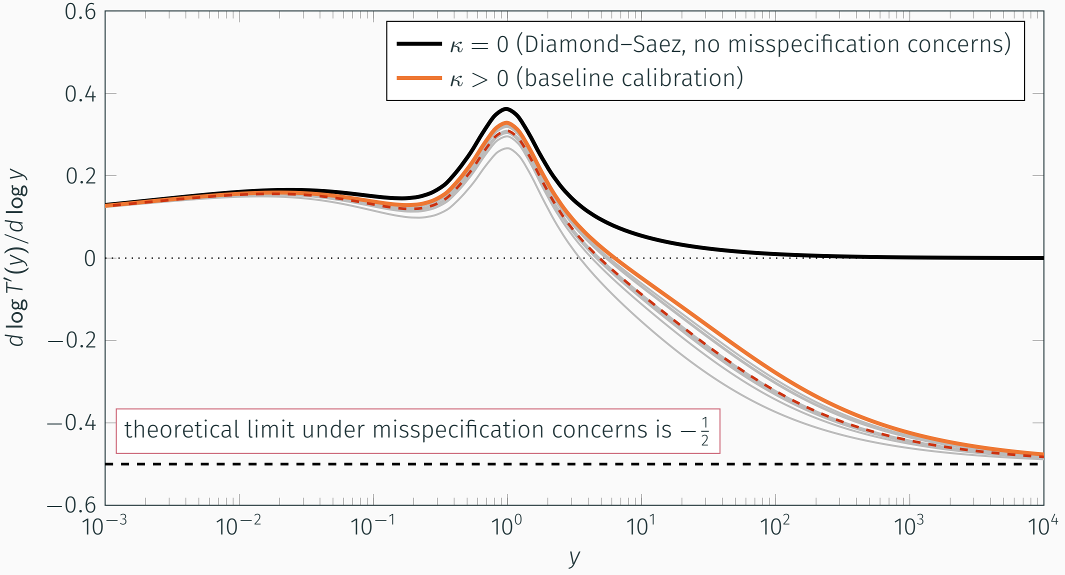

$\kappa = 0$ (Diamond–Saez, no misspecification concerns)
$\kappa > 0$ (baseline calibration)
0.8
0.6
0.4
0.2
0
$T'(y)$
optimal tax rate in the benchmark model peaks at 57.5% for 7.5 times average income (≈ $375k)
average income
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
y


κ = 0 (Diamond–Saez, no misspecification concerns)
κ > 0 (baseline calibration)
at 20 times average income (≈ $1m), marginal tax rates differ by 13 perc. points
average income
T′(y)
0.8
0.6
0.4
0.2
0
10^−3
10^−2
10^−1
10^0
10^1
10^2
10^3
10^4
y


$\kappa = 0$ (Diamond–Saez, no misspecification concerns)
$\kappa > 0$ (baseline calibration)
0.8
0.6
0.4
0.2
0
$T'(y)$
at 100 times average income (≈ $5m), marginal tax rates differ by 29 perc. points
average income
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
y

Optimal marginal tax schedules




$\kappa = 0$ (Diamond–Saez, no misspecification concerns)
$\kappa > 0$ (baseline calibration)
theoretical limit under misspecification concerns is $-\frac{1}{2}$
$d \log T'(y)/d \log y$
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
y

The worst-case density is characterized by the distortion

$$m(z) = \bar{m} \exp\left(-\frac{1}{\theta}\left[\mathcal{U}(z) + \mu T(y(z))\right]\right)$$

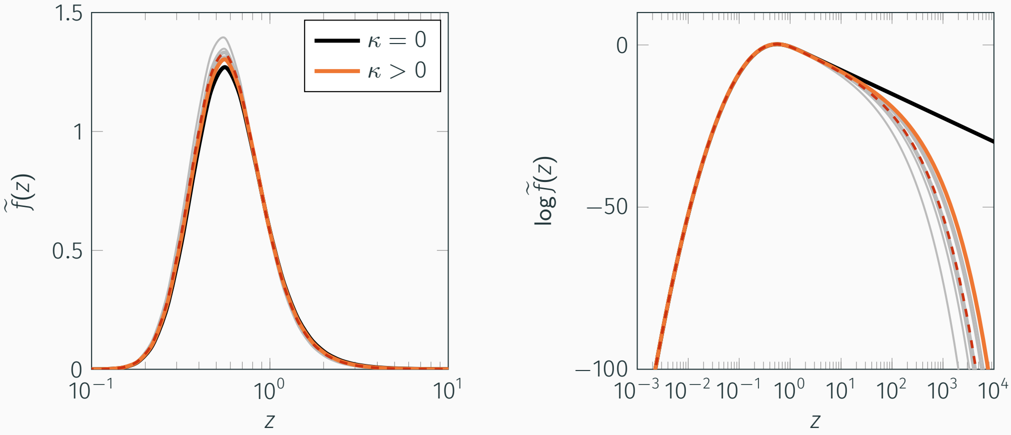
Left tail of the type distribution

- without redistribution, we would have $\lim_{z\to 0} \mathcal{U}(z) = -\infty$, and $\lim_{z\to 0} m(z) = \infty$
- redistributive transfers bound $\mathcal{U}(z)$ from below, and so $m(z)$ is bounded above

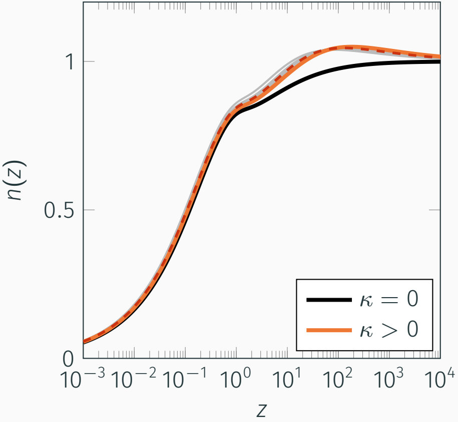
Right tail of the type distribution

- dominated by budgetary concerns
- since $\lim_{z\to\infty} T(y(z)) = \infty$, we also have $\lim_{z\to\infty} m(z) = 0$

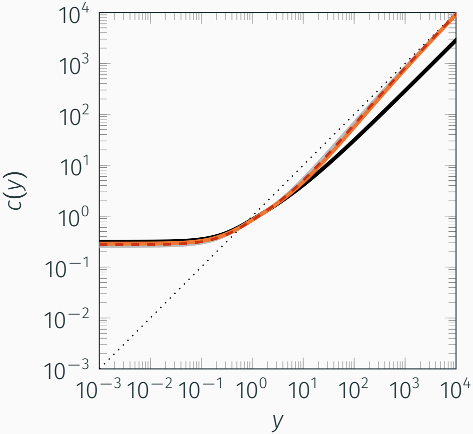
Worst-case distributions



- worst-case distributions $\widetilde{f}(z)$ for alternative levels of misspecification concerns given by θ
- $\kappa = 0$ corresponds to the rational benchmark for which $\widetilde{f}(z) = f(z)$



n(z)
1
0.5
0
10−3 10−2 10−1 10^0 10^1 10^2 10^3 10^4
z
κ = 0
κ > 0



c(y)
10^4
10^3
10^2
10^1
10^0
10−1
10−2
10−3
10−3 10−2 10−1 10^0 10^1 10^2 10^3 10^4
y

Moments under benchmark and worst-case distributions

moments \ κ	0	$\kappa_{baseline}$	κ_{median}
$\mathbb{E}[z]$	1.000	1.000	1.000
$\widetilde{\mathbb{E}}[z]$	1.000	0.944	0.914
$\mathbb{E}[y]$	0.823	0.841	0.850
$\widetilde{\mathbb{E}}[y]$	0.823	0.787	0.768
μ	1.215	1.270	1.303
T_0	-0.315	-0.289	-0.276
$\max_y T'(y)$ (%)	71.4	57.5	54.3
$\arg\max_y T'(y)$	∞	6.305	4.856
$\mathbb{E}[T]$	0.000	0.027	0.039
$\widetilde{\mathbb{E}}[T]$	0.000	0.000	0.000

Effects of an increase in uncertainty ($\nearrow \kappa$)

Moments under benchmark and worst-case distributions

moments \ κ	0	$\kappa_{baseline}$	κ_{median}
$\mathbb{E}[z]$	1.000	1.000	1.000
$\widetilde{\mathbb{E}}[z]$	1.000	0.944	0.914
$\mathbb{E}[y]$	0.823	0.841	0.850
$\widetilde{\mathbb{E}}[y]$	0.823	0.787	0.768
μ	1.215	1.270	1.303
T_0	-0.315	-0.289	-0.276
$\max_y T'(y)$ (%)	71.4	57.5	54.3
$\arg\max_y T'(y)$	∞	6.305	4.856
$\mathbb{E}[T]$	0.000	0.027	0.039
$\widetilde{\mathbb{E}}[T]$	0.000	0.000	0.000

Effects of an increase in uncertainty ($\nearrow \kappa$)

- more pessimistic subjective distribution of productivity $\left(\searrow \widetilde{\mathbb{E}}[z]\right)$

Moments under benchmark and worst-case distributions

moments \ κ	0	$\kappa_{baseline}$	κ_{median}
$\mathbb{E}[z]$	1.000	1.000	1.000
$\widetilde{\mathbb{E}}[z]$	1.000	0.944	0.914
$\mathbb{E}[y]$	0.823	0.841	0.850
$\widetilde{\mathbb{E}}[y]$	0.823	0.787	0.768
μ	1.215	1.270	1.303
T_0	−0.315	−0.289	−0.276
$\max_y T'(y)$ (%)	71.4	57.5	54.3
$\arg\max_y T'(y)$	∞	6.305	4.856
$\mathbb{E}[T]$	0.000	0.027	0.039
$\widetilde{\mathbb{E}}[T]$	0.000	0.000	0.000

Effects of an increase in uncertainty ($\nearrow \kappa$)

- more pessimistic subjective distribution of productivity $\left(\searrow \widetilde{\mathbb{E}}[z]\right)$
- lower taxes prop up labor supply ($\nearrow \mathbb{E}[y]$)

moments \ κ	0	$\kappa_{baseline}$	κ_{median}
$\mathbb{E}[z]$	1.000	1.000	1.000
$\widetilde{\mathbb{E}}[z]$	1.000	0.944	0.914
$\mathbb{E}[y]$	0.823	0.841	0.850
$\widetilde{\mathbb{E}}[y]$	0.823	0.787	0.768
μ	1.215	1.270	1.303
T_0	-0.315	-0.289	-0.276
$\max_y T'(y)$ (%)	71.4	57.5	54.3
$\arg\max_y T'(y)$	∞	6.305	4.856
$\mathbb{E}[T]$	0.000	0.027	0.039
$\widetilde{\mathbb{E}}[T]$	0.000	0.000	0.000

Effects of an increase in uncertainty ($\nearrow \kappa$)

- more pessimistic subjective distribution of productivity $\left(\searrow \widetilde{\mathbb{E}}[z]\right)$
- lower taxes prop up labor supply ($\nearrow \mathbb{E}[y]$)
- marginal social value of public funds increases, lump-sum transfer T_0 somewhat decreases

Moments under benchmark and worst-case distributions

moments \ κ	0	$\kappa_{baseline}$	κ_{median}
$\mathbb{E}[z]$	1.000	1.000	1.000
$\widetilde{\mathbb{E}}[z]$	1.000	0.944	0.914
$\mathbb{E}[y]$	0.823	0.841	0.850
$\widetilde{\mathbb{E}}[y]$	0.823	0.787	0.768
μ	1.215	1.270	1.303
T_0	−0.315	−0.289	−0.276
$\max_y T'(y)$ (%)	71.4	57.5	54.3
$\arg\max_y T'(y)$	∞	6.305	4.856
$\mathbb{E}[T]$	0.000	0.027	0.039
$\widetilde{\mathbb{E}}[T]$	0.000	0.000	0.000

Effects of an increase in uncertainty ($\nearrow \kappa$)

- more pessimistic subjective distribution of productivity $\left(\searrow \widetilde{\mathbb{E}}[z]\right)$
- lower taxes prop up labor supply ($\nearrow \mathbb{E}[y]$)
- marginal social value of public funds increases, lump-sum transfer T_0 somewhat decreases
- progressivity of the tax system declines

moments $\setminus\ \kappa$	0	$\kappa_{baseline}$	κ_{median}
$\mathbb{E}[z]$	1.000	1.000	1.000
$\widetilde{\mathbb{E}}[z]$	1.000	0.944	0.914
$\mathbb{E}[y]$	0.823	0.841	0.850
$\widetilde{\mathbb{E}}[y]$	0.823	0.787	0.768
μ	1.215	1.270	1.303
T_0	−0.315	−0.289	−0.276
$\max_y T'(y)$ (%)	71.4	57.5	54.3
$\arg\max_y T'(y)$	∞	6.305	4.856
$\mathbb{E}[T]$	0.000	0.027	0.039
$\widetilde{\mathbb{E}}[T]$	0.000	0.000	0.000

Effects of an increase in uncertainty ($\nearrow \kappa$)

- more pessimistic subjective distribution of productivity $\left(\searrow \widetilde{\mathbb{E}}[z]\right)$
- lower taxes prop up labor supply ($\nearrow \mathbb{E}[y]$)
- marginal social value of public funds increases, lump-sum transfer T_0 somewhat decreases
- progressivity of the tax system declines
- under the benchmark distribution, the tax scheme generates additional resources $\mathbb{E}[T]$

Revisit the model under the power divergence penalty

- $\eta = -1$: reverse Kullback–Leibler divergence
- deviations of the distribution in the right tail penalized more heavily
- worst-case is asymptotically Pareto with a thinner tail paramer $\widetilde{\alpha}$

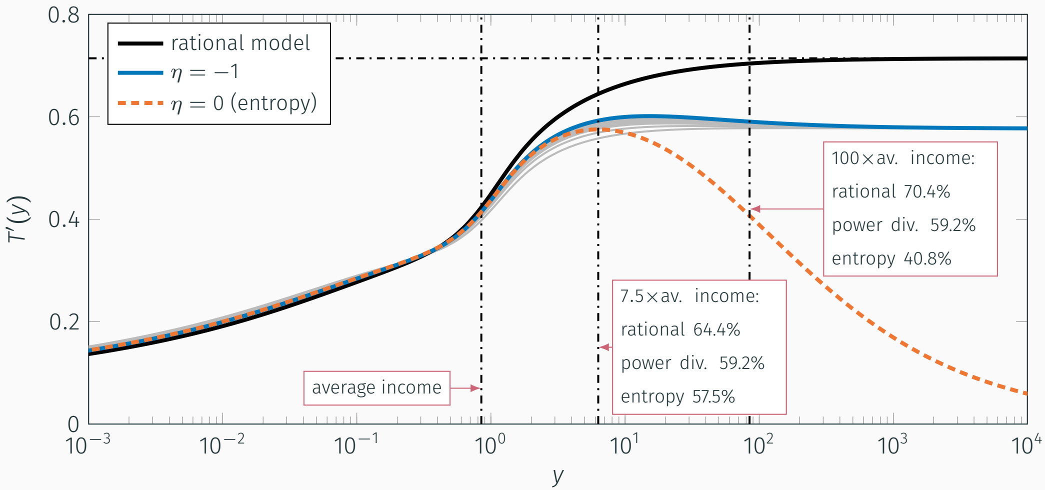
Optimal taxes still exhibit lower progressivity

- top marginal taxes no longer converge to zero

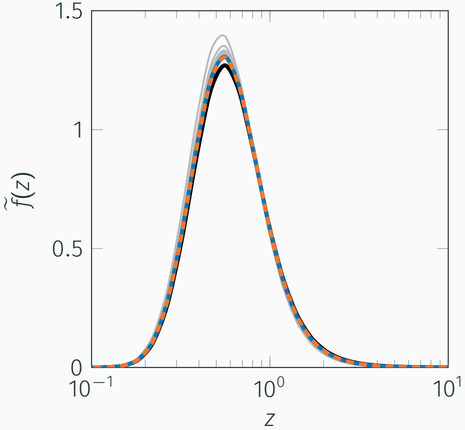
$$\lim_{y \to \infty} T'(y) = \frac{(\gamma + \rho)(1 + \gamma)}{\widetilde{\alpha}(\gamma + \rho) + \gamma(1 + \gamma)} \qquad \widetilde{\alpha} = \alpha - \frac{1 + \gamma}{\rho + \gamma} \frac{1}{\eta}$$

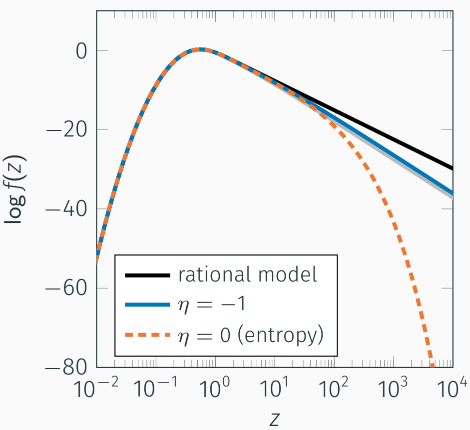
Distance between $\widetilde{\alpha}$ and $\alpha = 2.2$ is calibrated to the range of historical values in our sample and Vries and Toda (2022).

Optimal marginal tax schedule with shape restrictions



Worst case distributions with shape restrictions





Multidimensional type distribution

Uncertain productivity and elasticity of taxable income distribution

Large disagreement over elasticity of taxable income (ETI) estimates

- estimates vary across studied populations, econometric approaches, etc.

Allow for uncertainty over a multidimensional type distribution

- joint distribution of productivity and ETI

Approach

- incorporate multidimensional uncertainty using a modified penalty function
- calibrate using an auxiliary estimate from a California state tax reform
- analyze robust optimal tax policy in this setting

Modified penalty

Joint distribution $f(\gamma, z)$ over the inverse ETI γ and productivity z

Penalty that factorizes the joint distribution

$$\mathcal{P}\left(f, \widetilde{f} | \theta_\gamma, \theta_z\right) = \theta_\gamma \int \underbrace{\mathcal{E}\left(f_{\gamma|z}\left(\cdot|z\right), \widetilde{f}_{\gamma|z}\left(\cdot|z\right)\right)}_{\text{conditional ETI}} \widetilde{f}_z\left(z\right) dz + \theta_z \underbrace{\mathcal{E}\left(f_z, \widetilde{f}_z\right)}_{\text{productivity}}$$

Properties

- $\theta_\gamma = \theta_z = \theta$ implies $\mathcal{P}\left(f, \widetilde{f} | \theta_\gamma, \theta_z\right) = \theta \mathcal{E}\left(f, \widetilde{f}\right)$
- $\theta_\gamma \to \infty$ recovers a version of the one-dimensional uncertainty over labor productivity
- $\theta_z \to \infty$ preserves uncertainty over conditional ETI distributions

Planner solves the multidimensional problem

$$\max_{T} \min_{\substack{m>0 \\ \mathbb{E}[m]=1}} \mathbb{E}\left[m\psi\mathcal{U}\right] + V\left(\mathbb{E}\left[mT\left(\mathcal{Y}\right)\right]\right) + \mathcal{P}\left(f, \widetilde{f} | \theta_{\gamma}, \theta_{z}\right)$$

Planner solves the multidimensional problem

$$\max_{T} \min_{\substack{m>0 \\ \mathbb{E}[m]=1}} \mathbb{E}\left[m\psi\mathcal{U}\right] + V\left(\mathbb{E}\left[mT\left(\mathcal{Y}\right)\right]\right) + \mathcal{P}\left(f, \widetilde{f} | \theta_{\gamma}, \theta_{z}\right)$$

We focus on the distortions in the conditional distribution of ETI

$$m_{\gamma|z}\left(\gamma|z\right) = \frac{\widetilde{f}_{\gamma|z}\left(\gamma|z\right)}{f_{\gamma|z}\left(\gamma|z\right)} = \bar{m}_{\gamma|z}\left(z\right) \exp\left(-\frac{1}{\theta_{\gamma}}\left[\psi\left(\gamma,z\right)\mathcal{U}\left(\gamma,z\right) + \mu T\left(\mathcal{Y}\left(\gamma,z\right)\right)\right]\right)$$

GHH preferences

$$\frac{1}{1-\rho}\left(c - \frac{n^{1+\gamma}}{1+\gamma}\right)^{1-\rho}$$

Consider a tax reform from status quo tax function T^0 to new optimal tax function T.

- calibrate preferences so that under T^0 there is no distortion in conditional distribution of ETI

GHH preferences

$$\frac{1}{1-\rho}\left(c-\Psi\left(\gamma,z\right)\frac{n^{1+\gamma}}{1+\gamma}\right)^{1-\rho}+\Delta\left(\gamma,z\right)$$

Consider a tax reform from status quo tax function T^0 to new optimal tax function T.

- calibrate preferences so that under T^0 there is no distortion in conditional distribution of ETI
- introduce shifters $\Delta\left(\gamma,z\right)$ and $\Psi\left(\gamma,z\right)$ to equalize $\mathcal{U}\left(\gamma,z\right)$ and $\mathcal{Y}\left(\gamma,z\right)$ across γ given z

Calibration

Benchmark joint distribution $f(\gamma, z)$: γ and z are independent

- $\log z$ follows EMG distribution as in the 1D case
- γ^{-1} is distributed as in the pilot survey conducted in Lockwood et al. (2021) Details of $f_{\gamma|z}$

$\mathcal{P}\left(f, \widetilde{f} | \theta_\gamma, \theta_z\right)$: Focus only on concerns about ETI γ^{-1}

- $\theta_z \to \infty$
- θ_γ so that worst case distribution implies high-earning individuals have elasticities in the range found by Rauh and Shyu (2024)

Restrict $T(y)$ to be in a class of cubic spline functions Details of the restricted class

- Set status quo tax function T^0 to zero or to U.S. level

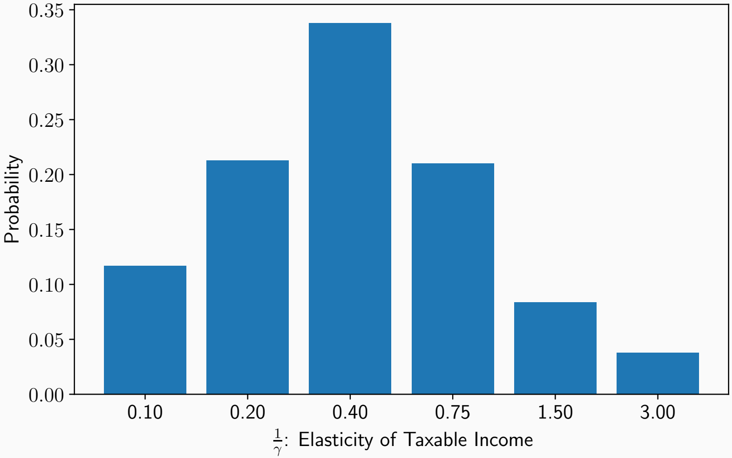
Utility curvature set to $\rho = 1$ as in Lockwood et al. (2021)

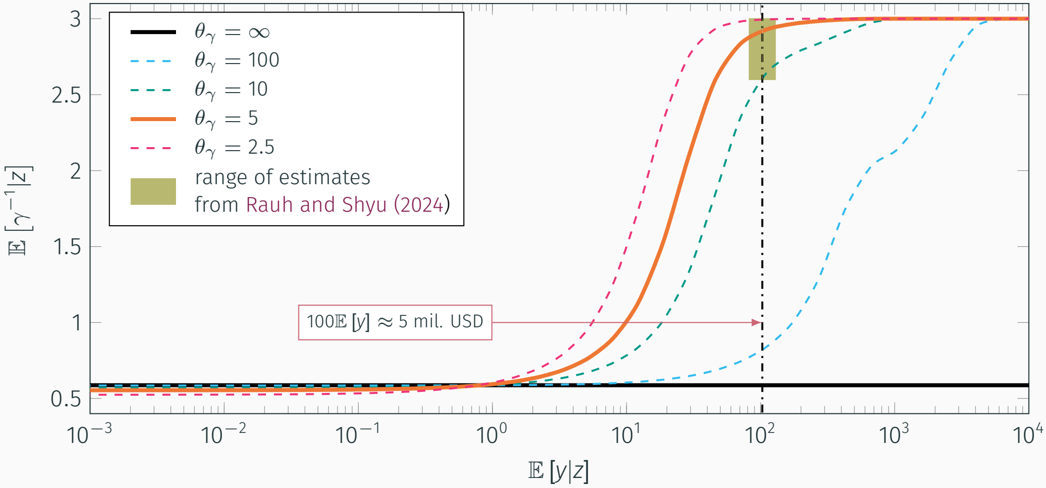
- All other parameters are same as before

Benchmark distribution of ETI

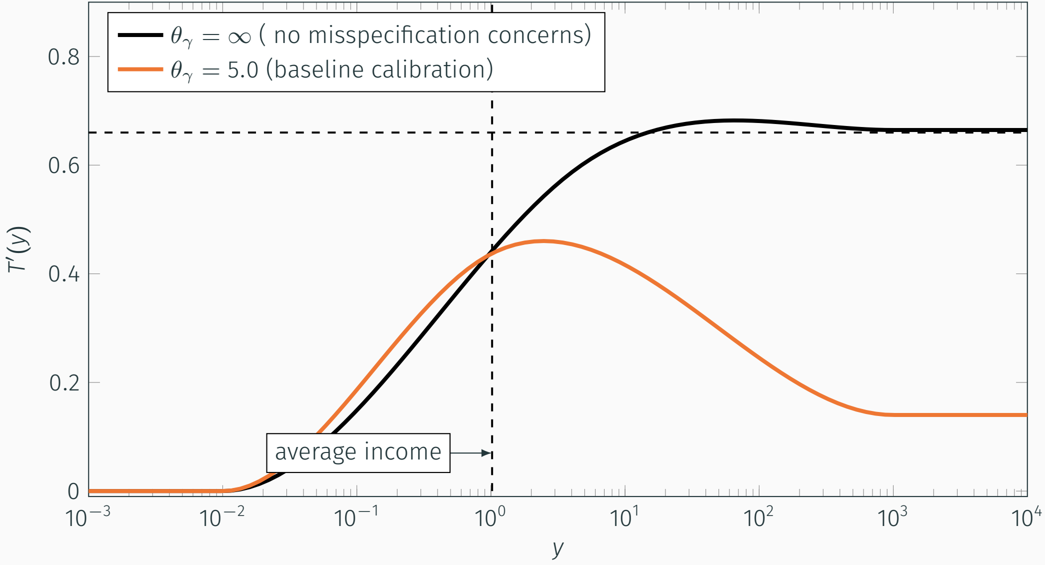
Lockwood et al. (2021) ran a survey to elicit beliefs on ETI among public economists.

Our benchmark distribution of ETI f_γ follows this survey evidence.

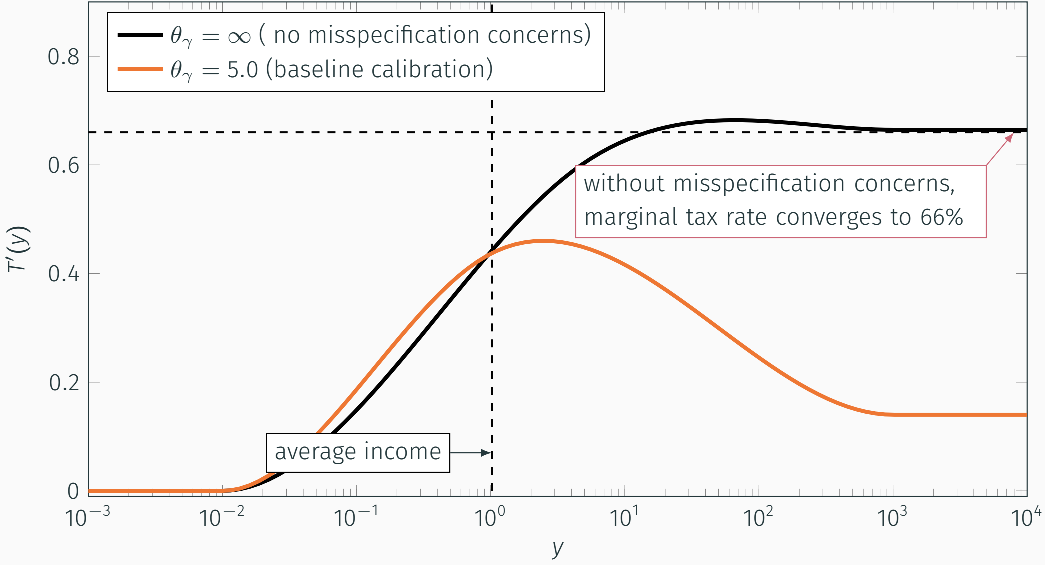



$\theta_\gamma = \infty$
$\theta_\gamma = 100$
$\theta_\gamma = 10$
$\theta_\gamma = 5$
$\theta_\gamma = 2.5$
range of estimates from Rauh and Shyu (2024)
$100\mathbb{E}[y] \approx 5$ mil. USD
$\mathbb{E}[\gamma^{-1}|z]$
3
2.5
2
1.5
1
0.5
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
$\mathbb{E}[y|z]$

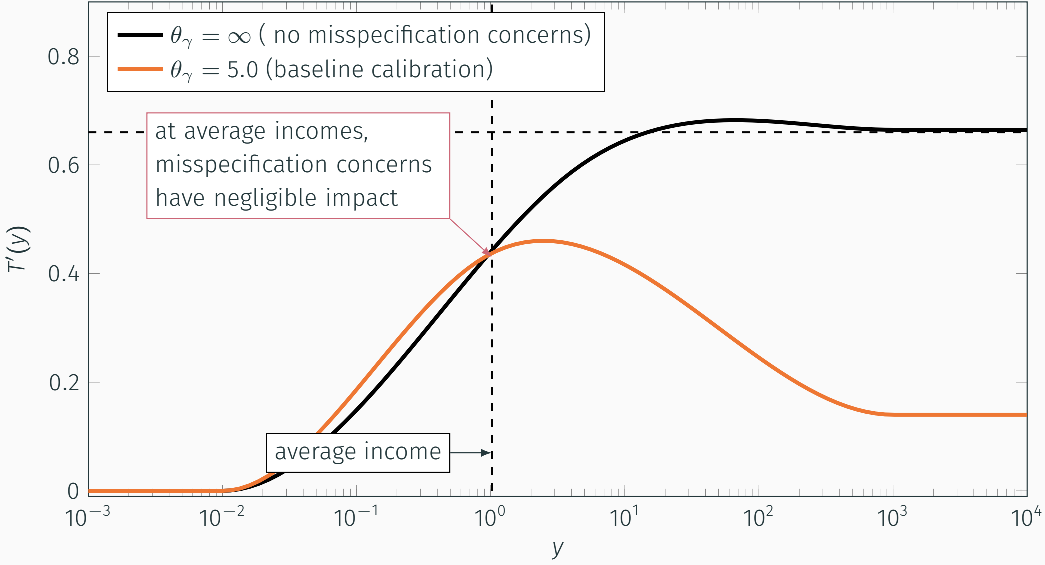
Robust planner lower top taxes

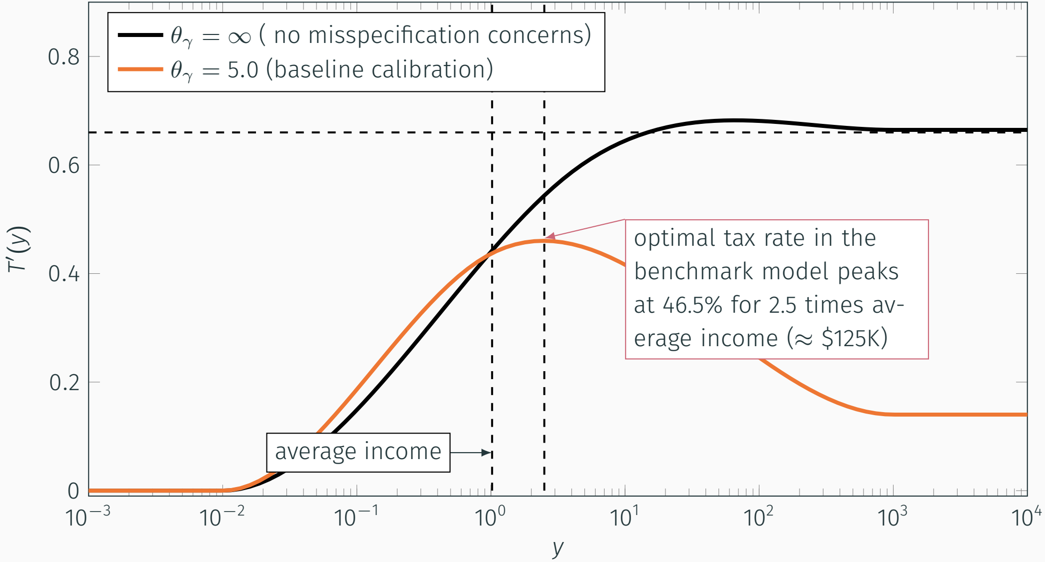


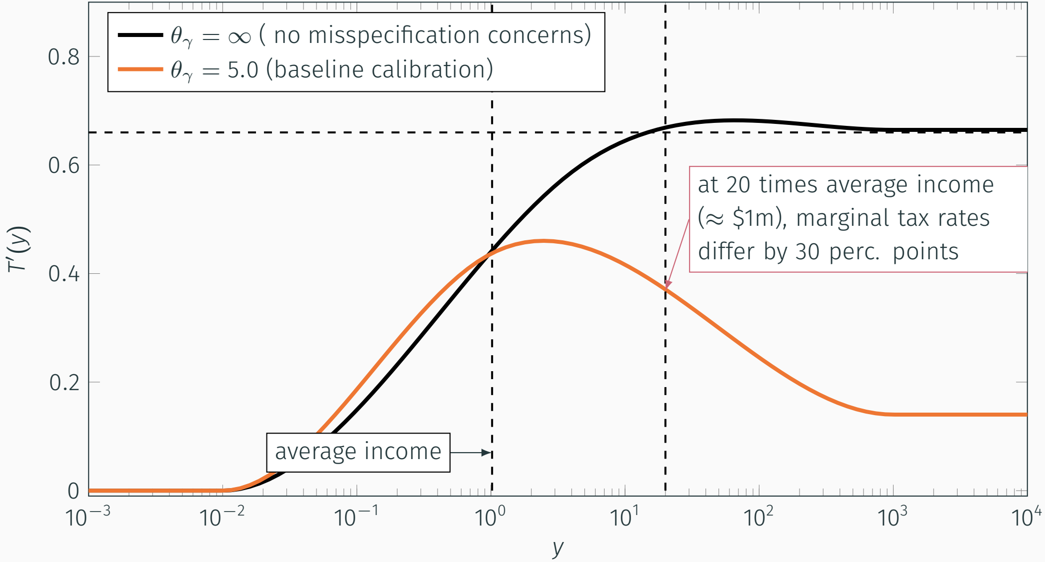
Robust planner lower top taxes

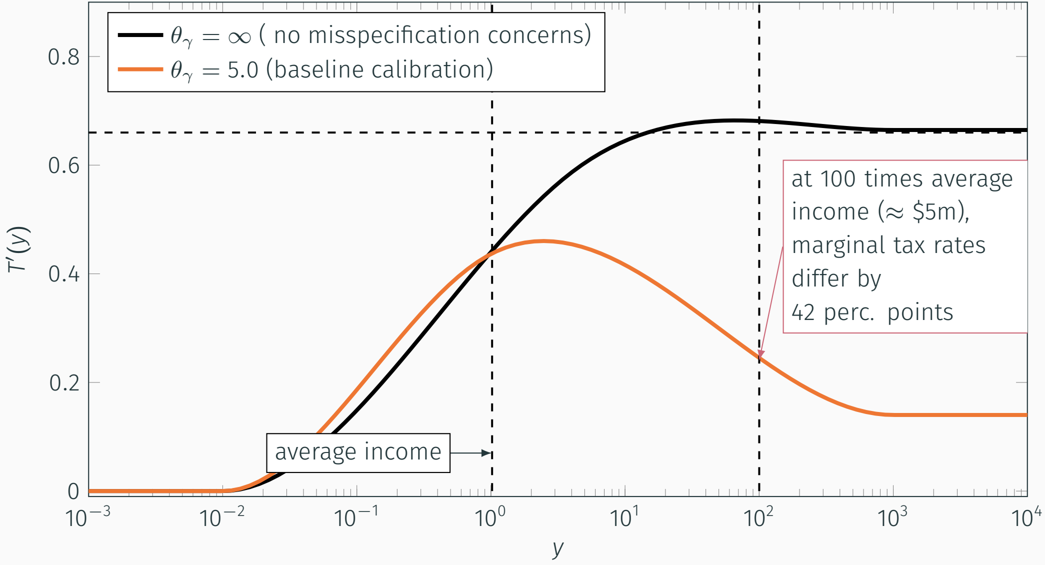


Robust planner lower top taxes

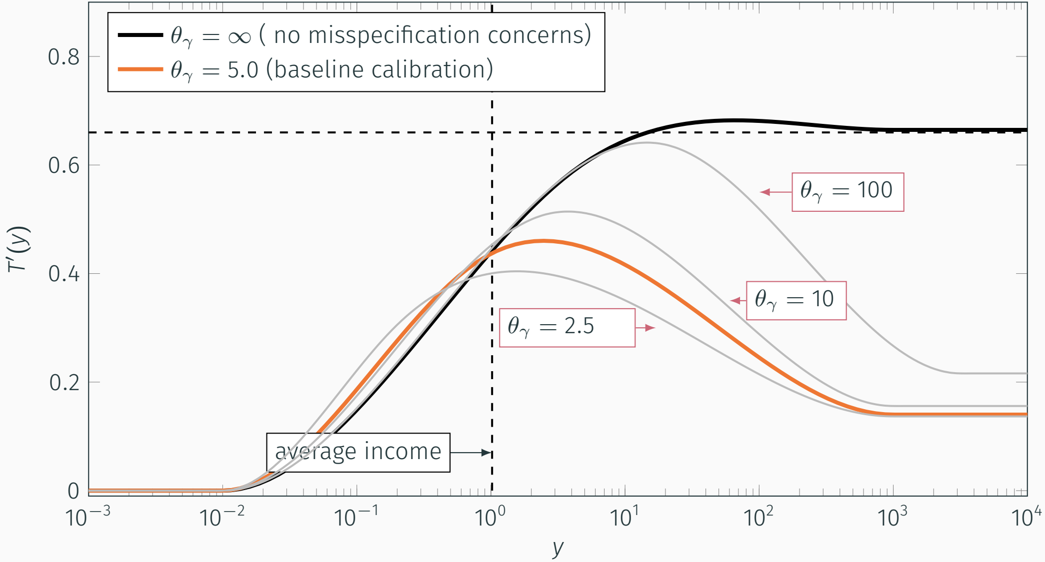


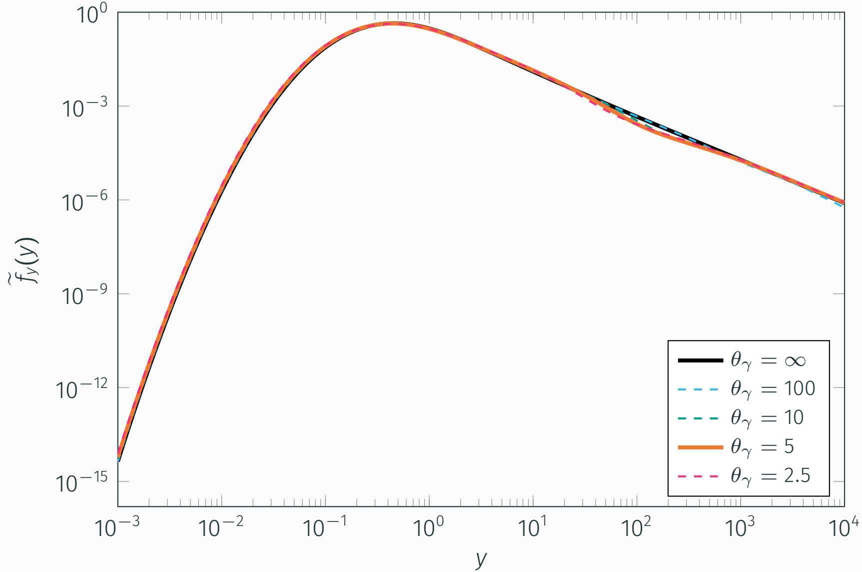

$\theta_\gamma = \infty$ (no misspecification concerns)
$\theta_\gamma = 5.0$ (baseline calibration)
0.8
0.6
0.4
0.2
0
$T'(y)$
optimal tax rate in the benchmark model peaks at 46.5% for 2.5 times average income (≈ $125K)
average income
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
y


$\theta_\gamma = \infty$ (no misspecification concerns)
$\theta_\gamma = 5.0$ (baseline calibration)
at 20 times average income (≈ $1m), marginal tax rates differ by 30 perc. points
average income
$T'(y)$
y
0.8
0.6
0.4
0.2
0
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}



$\theta_\gamma = \infty$ (no misspecification concerns)
$\theta_\gamma = 5.0$ (baseline calibration)
0.8
0.6
0.4
0.2
0
$T'(y)$
at 100 times average income (≈ $5m), marginal tax rates differ by 42 perc. points
average income
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
y


$\theta_\gamma = \infty$ (no misspecification concerns)
$\theta_\gamma = 5.0$ (baseline calibration)
0.8
0.6
0.4
0.2
0
$T'(y)$
$\theta_\gamma = 100$
$\theta_\gamma = 10$
$\theta_\gamma = 2.5$
average income
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
y



10^{0}
10^{-3}
10^{-6}
10^{-9}
10^{-12}
10^{-15}
$\widetilde{f}_y(y)$
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
y
$\theta_\gamma = \infty$
$\theta_\gamma = 100$
$\theta_\gamma = 10$
$\theta_\gamma = 5$
$\theta_\gamma = 2.5$

Main finding: Uncertainty about labor supply elasticity (ETI) $\implies$ lower top taxes

- Baseline calibration: Δ top tax rate ≈ -52 percentage points

Government worries about ability to raise sufficient revenues

- one-dimensional case: not enough highly productive workers
- two-dimensional case: not enough inelastic workers among the highly productive ones

Endogenously arising correlation between elasticities and skills under the worst-case model

- Benchmark: $\gamma^{-1} \perp z$ and $\mathbb{E}\left[\gamma^{-1}\right] \approx 0.587$
- Worst case: γ^{-1} and z positively correlated and $\widetilde{\mathbb{E}}\left[\gamma^{-1}|z\right] > \mathbb{E}\left[\gamma^{-1}\right]$ for high earners

Income distribution mainly driven by skills and minimally distorted.

moments \ θ_γ	∞	100	10	5	2.5
$\widetilde{\mathbb{E}}[z]$	1.000	1.000	1.000	1.000	1.000
$\widetilde{\mathbb{E}}[1/\gamma]$	0.587	0.587	0.590	0.594	0.601
$\mathbb{E}[1/\gamma \mid z \geq \bar{z}]$	0.587	0.587	0.587	0.587	0.587
$\widetilde{\mathbb{E}}[1/\gamma \mid z \geq \bar{z}]$	0.587	1.119	2.746	2.956	2.997
$\mathbb{E}[y]$	1.029	1.067	1.121	1.147	1.176
$\widetilde{\mathbb{E}}[y]$	1.029	1.015	1.001	0.998	0.995
μ	1.374	1.444	1.554	1.613	1.684
$T_0/\widetilde{\mathbb{E}}[y]$	−0.412	−0.387	−0.356	−0.339	−0.321
$\mathbb{E}[T]/\mathbb{E}[y]$	0.000	0.017	0.033	0.038	0.042
$\widetilde{\mathbb{E}}[T]/\widetilde{\mathbb{E}}[y]$	0.000	0.000	0.000	0.000	0.000

Threshold $\bar{z}$ is the minimum z such that $\mathbb{E}[y|z] \geq 100\mathbb{E}[y]$ in the benchmark distribution under zero status quo marginal taxes.

Conclusion

Conclusion

Acknowledging distributional uncertainty points toward lower progressivity.

- especially at the top, where budgetary concerns (per household) are most severe
- the left tail is well insured, leading to only modest concerns, unless overall uncertainty is substantial
- insights robust to variation in underlying distributions and preferences

Magnitude of misspecification concerns can be disciplined using

- administrative data: time-series variability in income distributions
- data on incomes and elasticity: reported elasticities of high-earning individuals

Extensions and follow-up work

If the benchmark distribution is ex-post correct, the optimal policy generates a surplus.

- dynamic debt management model

Other applications with substantial uncertainty about type distribution.

- wealth taxation

Bibliography

Cressie, Noel and Timothy R. C. Read (1984) "Multinomial Goodness-of-Fit Tests," *Journal of the Royal Statistical Society, Series B (Methodological)*, 46 (3), 440–464.

Diamond, Peter A. (1998) "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 88 (1), 83–95.

Diamond, Peter A. and Emmanuel Saez (2011) "The Case for a Progressive Tax: From Basic Research to Policy Recommendations," *Journal of Economic Perspectives*, 25 (4), 165–190.

Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski (2016) "Redistribution and Social Insurance," *American Economic Review*, 106 (2), 359–86, 10.1257/aer.20111550.

Hansen, Lars Peter and Thomas J. Sargent (2001a) "Acknowledging Misspecification in Macroeconomic Theory," *Review of Economic Dynamics*, 4 (3), 519–535.

——— (2001b) "Robust Control and Model Uncertainty," *American Economic Review*, 91 (2), 60–66.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante (2017) "Optimal Tax Progressivity: An Analytical Framework," *Quarterly Journal of Economics*, 132 (4), 1693–1754.

Heathcote, Jonathan and Hitoshi Tsujiyama (2021) "Optimal Income Taxation: Mirrlees Meets Ramsey," *Journal of Political Economy*, 129 (11), 3141–3184.

Lockwood, Benjamin B., Afras Sial, and Matthew Weinzierl (2021) "Designing, Not Checking, for Policy Robustness: An Example with Optimal Taxation," *Tax Policy and the Economy*, 35, 1–54, 10.1086/713492.

Mirrlees, James A. (1971) "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38 (2), 175–208.

Ramsey, Frank P. (1927) "A Contribution to the Theory of Taxation," *Economic Journal*, 37 (145), 47–61.

Rauh, Joshua D. and Ryan Shyu (2024) "Behavioral Responses to State Income Taxation of High Earners: Evidence from California," *American Economic Journal: Economic Policy*, 16 (1), 34–86.

Saez, Emmanuel (2001) "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68 (1), 205–229.

Vries, Tjeerd and Alexis Akira Toda (2022) "Capital and Labor Income Pareto Exponents Across Time and Space," *Review of Income and Wealth*, 68 (4), 1058–1078.

Additional slides

Application of the minimax theorem

We focus on the discrete-type formulation with separable isoelastic household utility

$$\max_{\{c_i, y_i\}_{i=1}^I \in \mathbb{X}} \min_{\{m_i\}_{i=1}^I \in \mathbb{Y}} \sum_{i=1}^I \pi_i m_i \left(u(c_i) - v\left(\frac{y_i}{z_i}\right)\right) + \bar{\nu} \sum_{i=1}^I \pi_i m_i (y_i - c_i) + \theta \sum_{i=1}^I \pi_i m_i \log m_i,$$

subject to the incentive compatibility constraints

$$u(c_i) - v\left(\frac{y_i}{z_i}\right) \geq u(c_j) - v\left(\frac{y_j}{z_i}\right), \qquad i, j \in \{1, \cdots, I\},$$

and Radon–Nikodým derivative constraints

$$\sum_{i=1}^I \pi_i m_i = 1, \qquad m_i \geq 0, \qquad i \in \{1, \ldots, I\}.$$

The minimax theorem requires

1. concave-convex objective function
2. compact and convex subsets of $\mathbb{X}$ and $\mathbb{Y}$

Application of the minimax theorem – convex IC constraints

Rewrite the problem as a function of maximizing controls u_i and v_i.

$$\max_{\{u_i,v_i\}_{i=1}^{I}\in\mathbb{X}} \ \min_{\{m_i\}_{i=1}^{I}\in\mathbb{Y}} \sum_{i=1}^{I} \pi_i m_i \left(u_i - v_i + \bar{\nu} z_i v^{-1}(v_i) - \bar{\nu} u^{-1}(u_i)\right) + \theta \sum_{i=1}^{I} \pi_i m_i \log m_i$$

subject to linear constraints

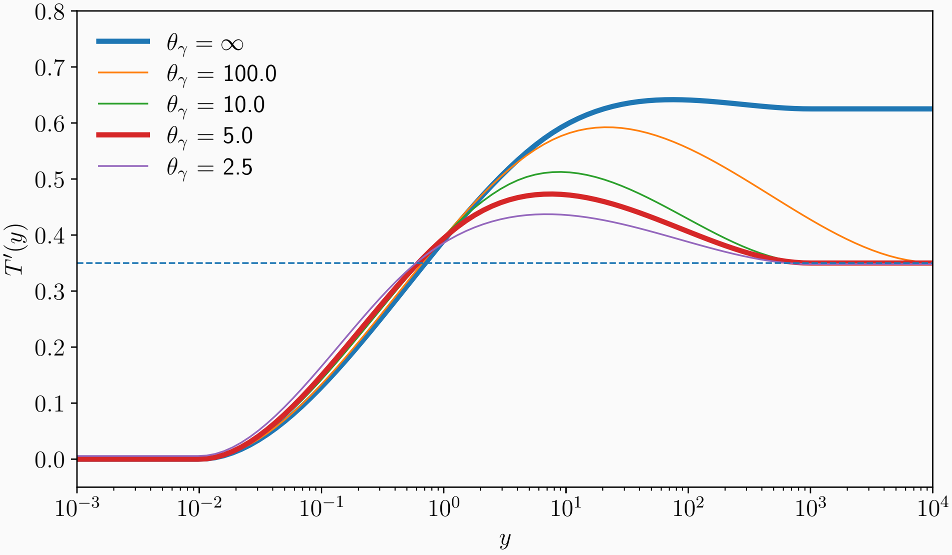
$$u_i - v_i \geq u_j - \left(\frac{z_j}{z_i}\right)^{1+\gamma} v_j, \qquad i,j \in \{1,\cdots,I\},$$

and

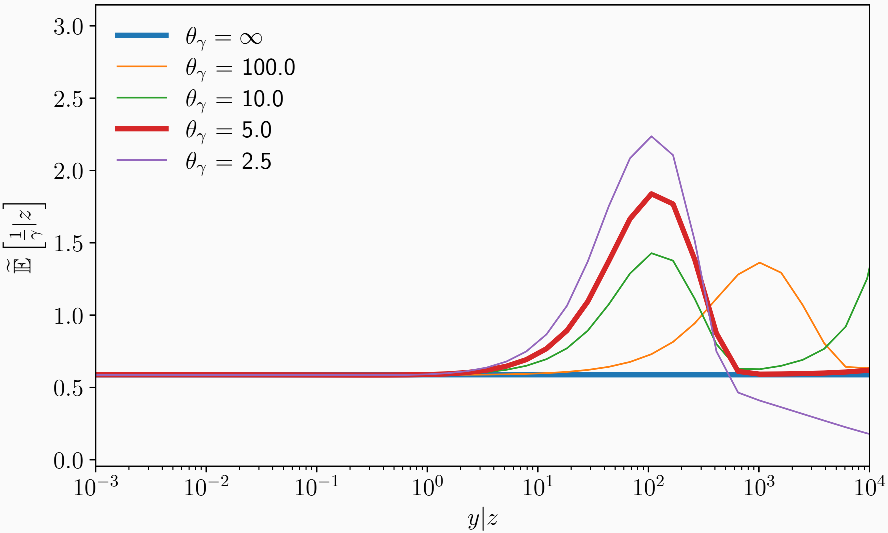
$$\sum_{i=1}^{I} \pi_i m_i = 1, \qquad m_i \geq 0, \qquad i \in \{1,\ldots,I\}.$$

The incentive compatibility constraints are convex in u_i and v_i with isoelastic utility.

Results: Optimal Taxes, T^0 = US (35%)



Results: Worst case distribution with different θ_γ



Results with $T^0 = \text{US}$: Intuition

Main finding: High uncertainty about elasticity $\rightarrow$ optimal T converges to T^0 at the top.

- choosing status quo $T^0 = US$ decreases top tax rate by about 25 p.p. relative to no uncertainty

Government worries about potential adverse impact of the tax reform among high earners.

- Increase tax rate: Workers may be more responsive than anticipated $\rightarrow$ larger distortive effects
- Decrease the tax rate: Workers may be less responsive than anticipated $\rightarrow$ revenue loss

This creates the tendency for the government to preserve the status quo.

- Top earners do not change labor supply as long as T' is not changed

Bias towards the status quo: When in doubt, do not change the tax rate.

Let $\mathcal{C}(\gamma, z|T), \mathcal{N}(\gamma, z|T)$ be optimal choices for

$$\mathcal{U}(\gamma, z|T) = \max_{c,n,y:\ c \leq y - T(y)} \frac{1}{1-\rho}\left(c - \Psi(\gamma, z)\frac{n^{1+\gamma}}{1+\gamma}\right)^{1-\rho} + \Delta(\gamma, z)$$

Reverse engineer Ψ and Δ

- so that $T = T^0$ is sufficient for optimal choices and indirect utilities to be independent of curvature on labor supply given productivity

$\implies$ incentives to distort conditional distributions $\gamma|z$ vanish when $T \to T^0$

Illustrate using $T^0 = 0$ (general case in paper)

Proposition 1.2

Let $\Psi(\gamma, z) = \bar{\psi}^{\frac{\gamma}{\bar{\gamma}}} z^{1-\frac{\gamma}{\bar{\gamma}}}$ *and* $\Delta(\gamma, z) = \frac{\bar{\gamma}}{1+\bar{\gamma}} \frac{z^{1+\frac{1}{\bar{\gamma}}}}{\bar{\psi}^{\frac{1}{\bar{\gamma}}}} - \left(1 - \frac{1}{1+\gamma}\left(\frac{z}{\bar{\psi}}\right)^{\left(-1+\frac{\gamma}{\bar{\gamma}}\right)\frac{\gamma}{\bar{\gamma}}}\right) \frac{z^{1+\frac{\gamma}{\bar{\gamma}^2}}}{\bar{\psi}^{\frac{\gamma}{\bar{\gamma}^2}}}$ *for some constants* $\bar{\psi}, \bar{\gamma}$.

If $T^0 = 0$*, then for all* γ', γ'', z *we have*

$$\mathcal{C}(\gamma', z|T^0) = \mathcal{C}(\gamma'', z|T^0) \quad \mathcal{N}(\gamma', z|T^0) = \mathcal{N}(\gamma'', z|T^0) \quad \mathcal{U}(\gamma', z|T^0) = \mathcal{U}(\gamma'', z|T^0).$$

Restricted class of $T(y)$

In the 2D set up, we restrict marginal tax rate $T'(y) = \sum_j c_j \phi^j(\ln y)$ where ϕ^j are cubic polynomials

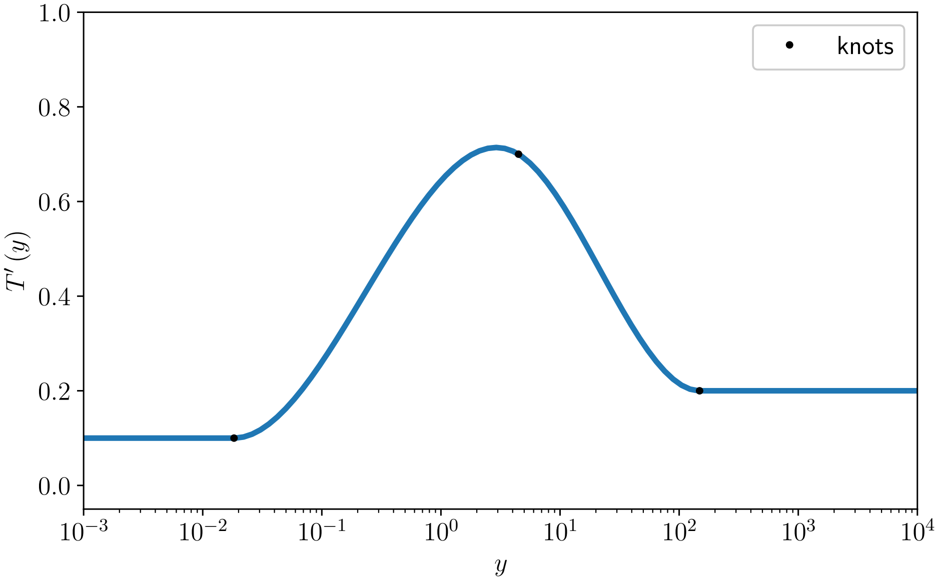
Parametrize $T'(y)$ using N knots $\{(\ln y_i, \tau_i)\}_{i=1,\cdots,N}$:

- Impose $T'(y)$ to be constant for $(0, y_1]$ and $[y_N, \infty)$
- Require $T'(y)$ is smooth at $y = y_i \ \ (i = 2, \cdots, N-1)$ and differentiable at $y = y_1$ and y_N

$$T'(y) = \begin{cases} \tau_1 & (y < y_1) \\ \text{CubicSpline}\left(\ln y; \{(\ln y_i, \tau_i)\}_{i=1,\cdots,N}\right) & (y_1 \leq y \leq y_N) \\ \tau_N & (y_N < y) \, . \end{cases}$$

Optimal T is obtained maximizing welfare given this class of functions and budget balance

Example of $T'(y)$ with Cubic Spline



How to pick N?

Large N introduces numerical instability

- welfare is not guaranteed to well-behaved with respect to underlying parameters

Small N introduces welfare losses

- insufficient flexibility might limit the welfare gains from optimal taxes

Find the smallest N so that welfare gains are “sufficiently close” to the Mirrlees solution

- define “sufficiently close” using a consumption-equivalent welfare gains threshold
- implement in cases where the full Mirrlees solution is feasible

Benchmarking Cubic Splines in 1D

Well-known that multidimensional screening problems are difficult to characterize

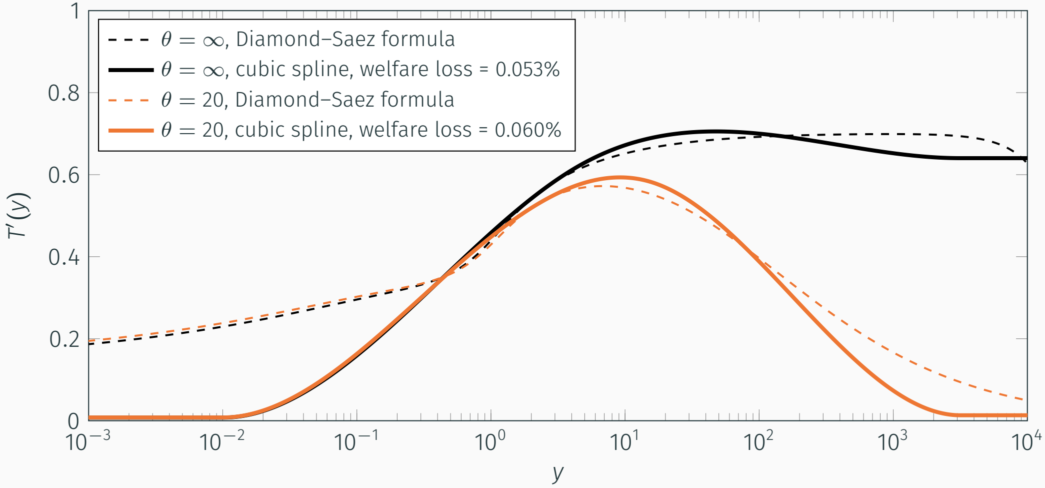
- e.g., first-order approach is not guaranteed to work

Use various parametrizations of 1D setup as a laboratory

- figure out the appropriate N for the multidimensional case
- test the cubic spline method

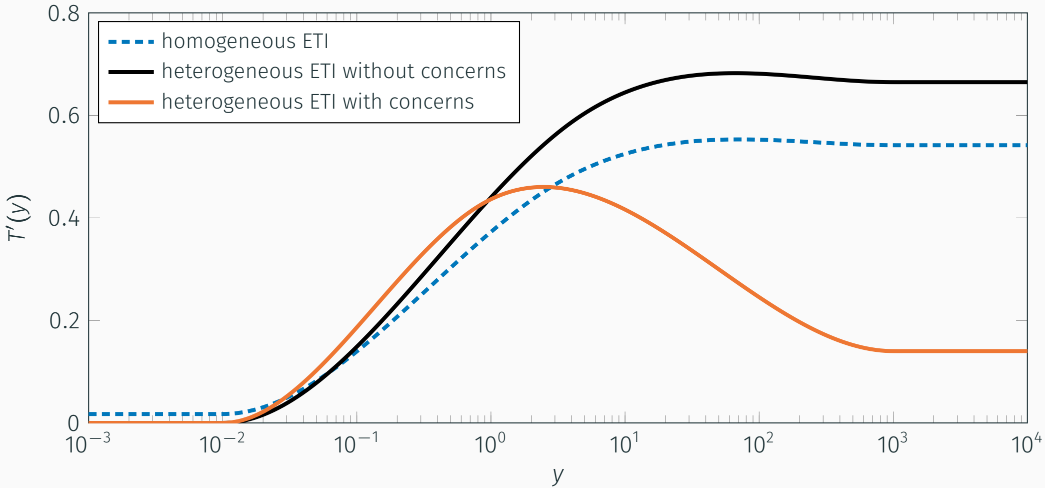
Finding: cubic splines with $N = 3$ do a good job in capturing shape and welfare gains

Cubic Spline approximates Diamond-Saez solution in 1D



- Welfare loss relative to the Diamond–Saez solution is approximately 0.05% of consumption

Comparison with Lockwood Sial, and Weinzierl (2021)



- Lockwood et al. (2021): Heterogeneous ETI increases the top marginal tax rate (Blue → Black)
- Ours: Concerns over the ETI distribution reduce the top marginal tax rate (Black → Orange)