Robust Bounds on Optimal Tax Progressivity

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

December 2024

[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)

This paper

Optimal tax design acknowledging uncertainty about distribution of individual characteristics

- build on decision theory under ambiguity 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

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

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

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

subject to

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

- $\psi(z)$ 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

A robust utilitarian government solves the max-min problem

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

subject to

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

- the minimax theorem allows switching the order of optimization
- inner problem can be approached using tools from the Mirrleesian literature

Theoretical analysis

The (inner) optimal tax problem can be cast as a mechanism design problem (Mirrlees (1971))

- revelation principle allows to focus on direct mechanisms
- workers provide a report z' of their type z
- government offers a menu of allocations $(c(z'), y(z'))$ that incentivizes truthtelling, $z' = z$
- implied tax function $T(y(z)) = y(z) - c(z)$

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

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

Ex-post Bayesian interpretation

Fixing $m(z)$ (fixing a distribution $\widetilde{F}(z)$), the problem is as in Mirrlees (1971), now under $\widetilde{F}(z)$.

- ex-post Bayesian interpretation of $\widetilde{F}(z)$

Incentive-compatibility constraints are type-by-type, do not depend on the distribution.

- misspecification concerns do not alter incentive compatibility

Optimal allocation and the minimizing 'worst-case' distribution determined jointly.

Worst-case distribution

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

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

- 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

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

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.

$$
\begin{aligned}
\frac{T'(y(z))}{1-T'(y(z))} &= (1+\gamma)\frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)} \\
T(y(z)) &= T(y(\underline{z})) + \int_{y(\underline{z})}^{y(z)} T'(\eta)d\eta \\
m(z) &= \bar{m}\exp\left(-\frac{\mu}{\theta}T(y(z))\right)
\end{aligned}
$$

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

$$\begin{aligned}
\frac{T'(y(z))}{1-T'(y(z))} &= (1+\gamma)\frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)} \\
T(y(z)) &= T(y(\underline{z})) + \int_{y(\underline{z})}^{y(z)} T'(\eta)d\eta \\
m(z) &= \bar{m}\exp\left(-\frac{\mu}{\theta}T(y(z))\right)
\end{aligned}$$

- when $T'(y(z))$ decays more slowly, $T(y(z))$ grows faster

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

$$\begin{aligned}
\frac{T'(y(z))}{1-T'(y(z))} &= (1+\gamma)\frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)} \\
T(y(z)) &= T(y(\underline{z})) + \int_{y(\underline{z})}^{y(z)} T'(\eta)d\eta \\
m(z) &= \bar{m}\exp\left(-\frac{\mu}{\theta}T(y(z))\right)
\end{aligned}$$

- when $T'(y(z))$ decays more slowly, $T(y(z))$ grows faster
- the distortion $m(z)$ thins out the density at the top faster

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

$$\begin{aligned}
\frac{T'(y(z))}{1-T'(y(z))} &= (1+\gamma)\frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)} \\
T(y(z)) &= T(y(\underline{z})) + \int_{y(\underline{z})}^{y(z)} T'(\eta)d\eta \\
m(z) &= \bar{m}\exp\left(-\frac{\mu}{\theta}T(y(z))\right)
\end{aligned}$$

- when $T'(y(z))$ decays more slowly, $T(y(z))$ grows faster
- the distortion $m(z)$ thins out the density at the top faster
- the optimal tax formula then implies a faster decay rate of $T'(y(z))$

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

$$\begin{aligned} \frac{T'(y(z))}{1-T'(y(z))} &= (1+\gamma)\frac{1-\widetilde{F}(z)}{z\widetilde{f}(z)} \\ T(y(z)) &= T(y(\underline{z})) + \int_{y(\underline{z})}^{y(z)} T'(\eta)d\eta \\ m(z) &= \bar{m}\exp\left(-\frac{\mu}{\theta}T(y(z))\right) \end{aligned}$$

- when $T'(y(z))$ decays more slowly, $T(y(z))$ grows faster
- the distortion $m(z)$ thins out the density at the top faster
- the optimal tax formula then implies a faster decay rate of $T'(y(z))$

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

Assume that the benchmark type distribution $F(z)$ is Pareto with shape parameter α.

- combining equations that characterize the fixed point argument and differentiating yields

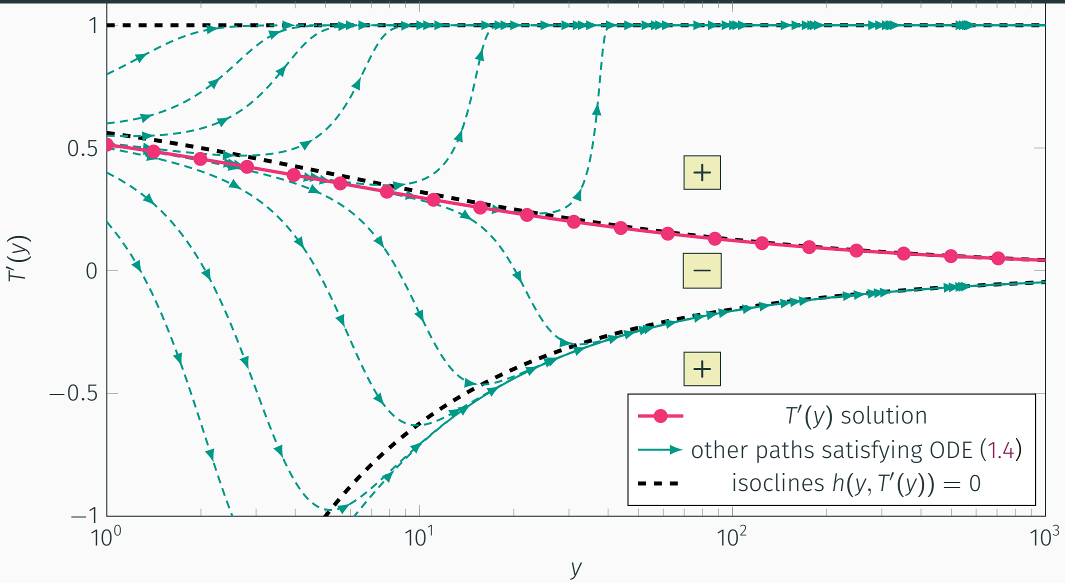
$$-\frac{T''(y)\,y}{1-T'(y)} = -\left[2-\frac{1+\gamma+\alpha}{1+\gamma}T'(y)\right]^{-1}\left[\frac{\mu}{\theta}\left[T'(y)\right]^{2}y-\gamma+\gamma\frac{1+\gamma+\alpha}{1+\gamma}T'(y)\right] \tag{1.3}$$

We thus obtain the differential equation

$$T''(y) = h\left(y, T'(y)\right). \tag{1.4}$$

- unique strictly positive solution that satisfies the transversality condition

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



1
0.5
0
−0.5
−1
$T'(y)$
10^0
10^1
10^2
10^3
y
+
−
+
$T'(y)$ solution
other paths satisfying ODE (1.4)
isoclines $h(y, T'(y)) = 0$

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

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

Results carry over to a general class of power divergence functions of Cressie and Read (1984).

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

For example,

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

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

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

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

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)

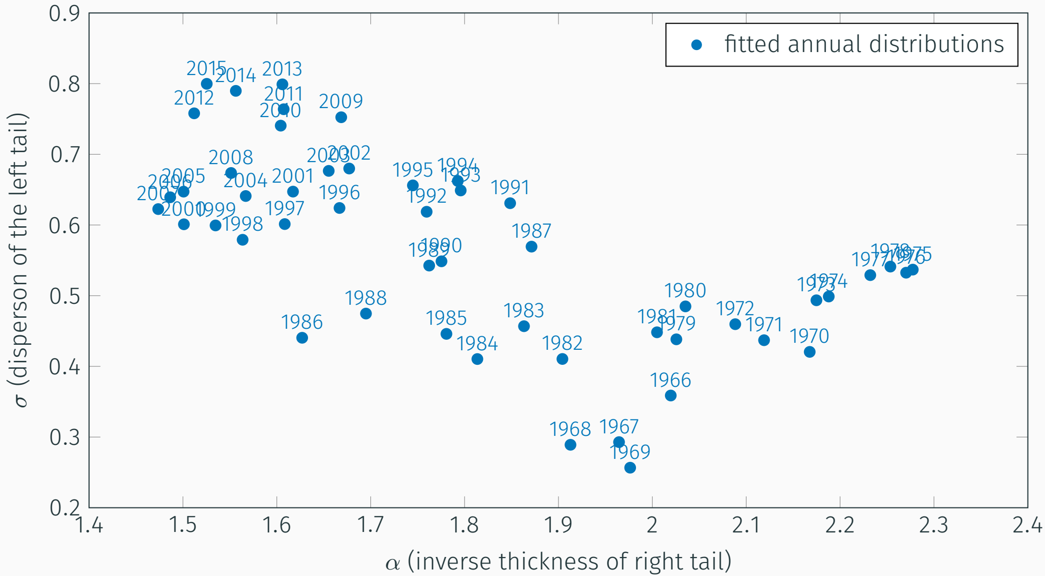
Entropy bound κ

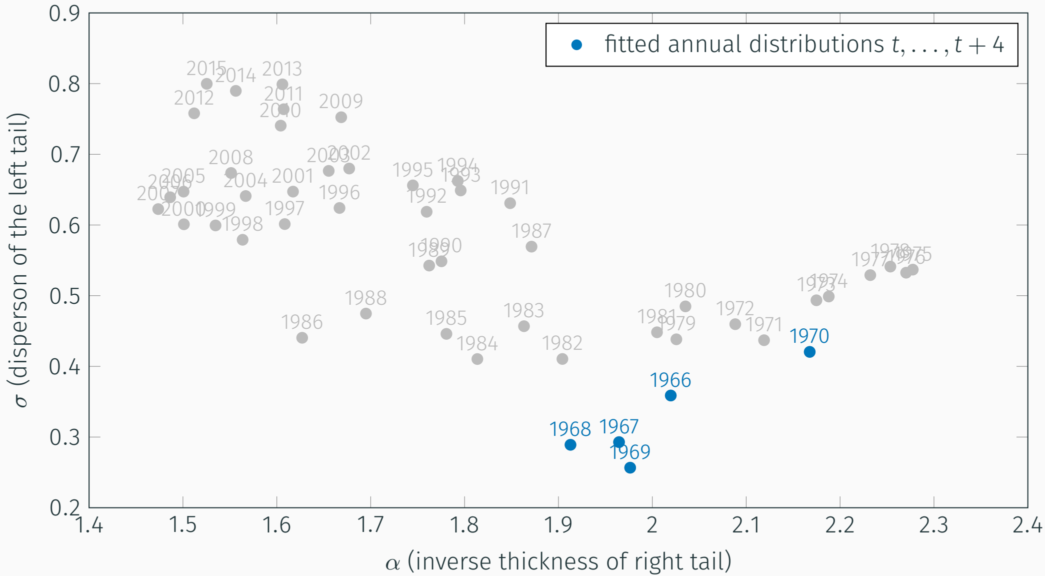
- use time-series variation in observed income distributions (World Income Database)

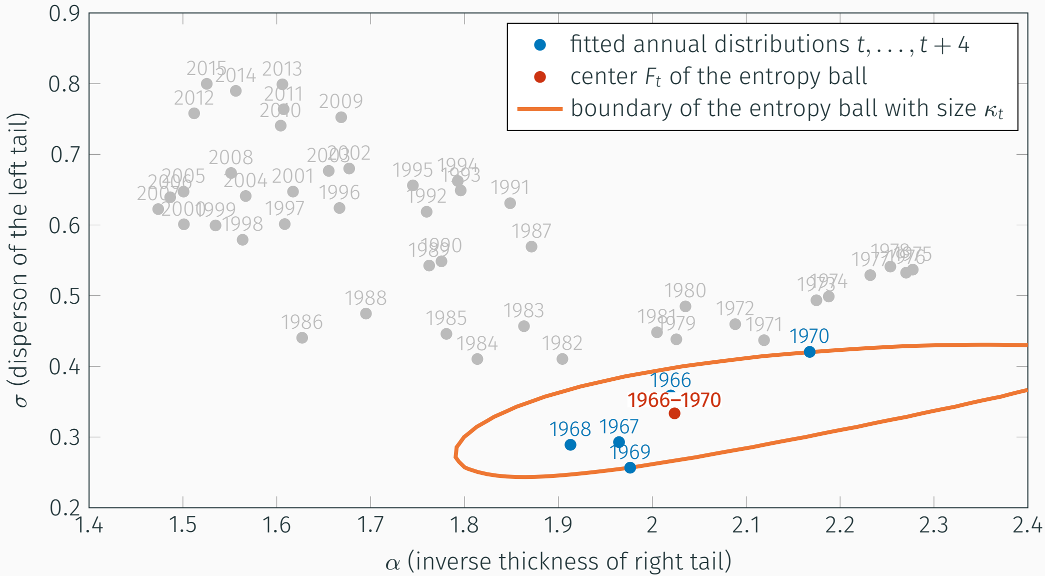
1. For each year t, we fit the EMG distribution to obtain parameters $(\mu_t, \sigma_t, \alpha_t)$.
2. For each 5-year window $\{t, \ldots, t+4\}$, we construct $\mathcal{F}(F_t, \kappa_t)$ as the set that
   - includes all fitted EGM distributions from years $\{t, \ldots, t+4\}$
   - has the smallest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

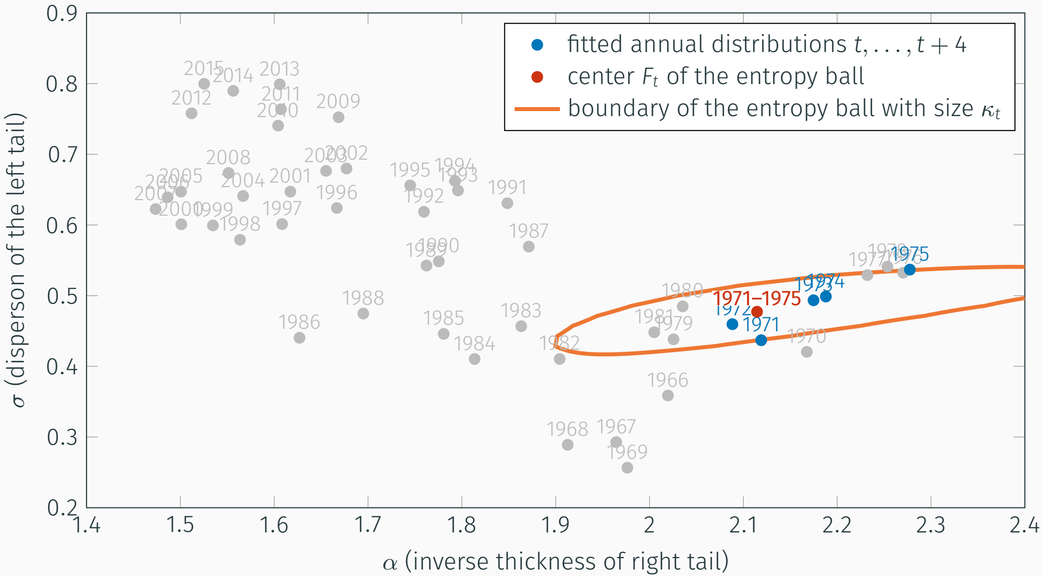
The set $\mathcal{F}(F_t, \kappa_t)$ is rich:

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

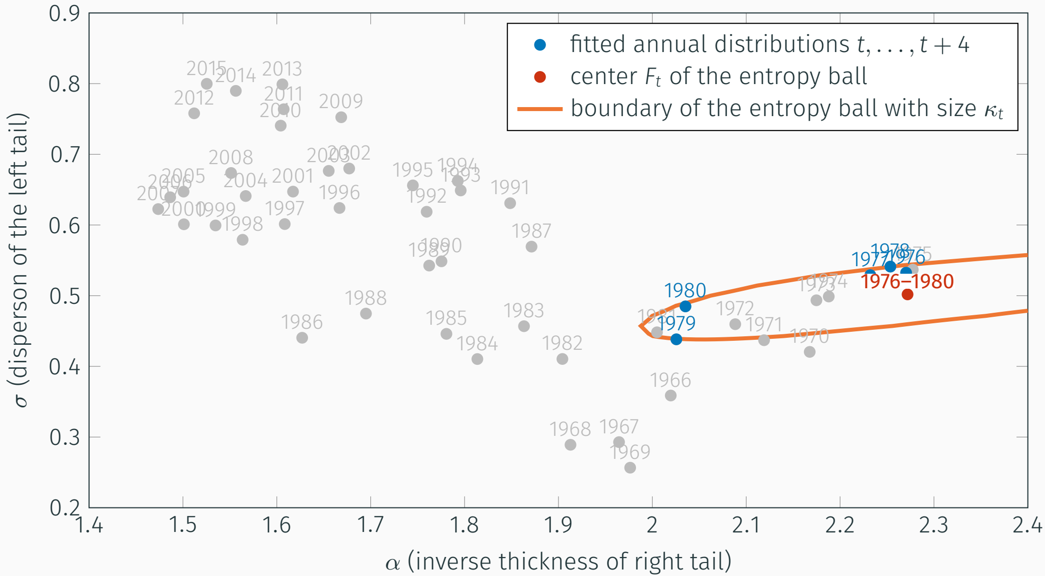

fitted annual distributions
σ (disperson of the left tail)
α (inverse thickness of right tail)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4
2015
2014
2013
2012
2011
2010
2009
2008
2003
2002
2005
2006
2007
2004
2001
2000
1999
1998
1997
1996
1995
1994
1993
1992
1991
1987
1989
1990
1988
1986
1985
1984
1983
1982
1981
1980
1979
1978
1977
1976
1975
1974
1973
1972
1971
1970
1966
1968
1967
1969

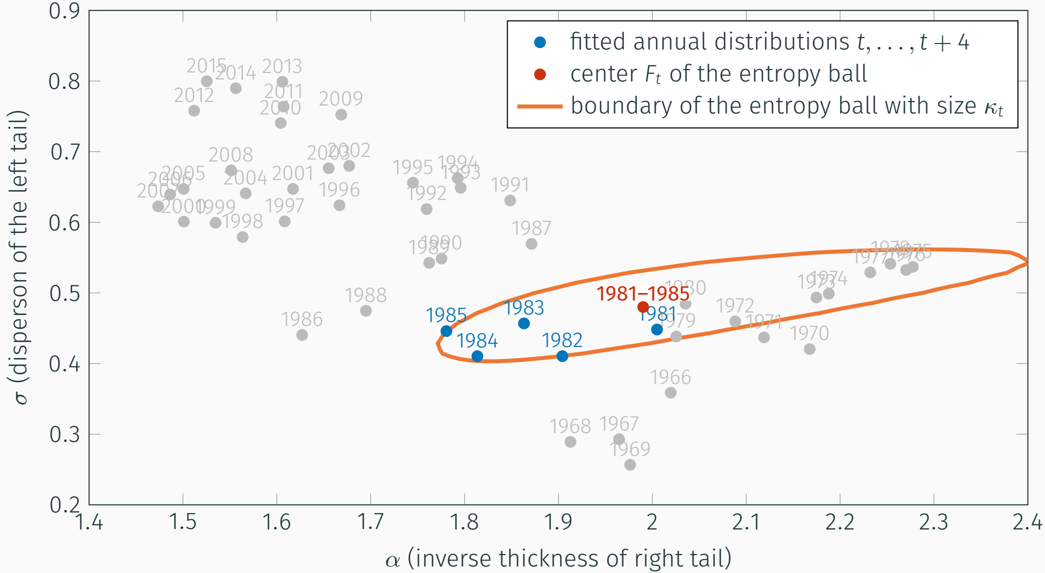

fitted annual distributions $t, \ldots, t+4$
σ (disperson of the left tail)
α (inverse thickness of right tail)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4
2015
2014
2013
2012
2011
2009
2008
2005
2004
2001
2002
1996
1995
1994
1992
1991
1998
1997
1987
1988
1986
1985
1984
1983
1982
1980
1972
1971
1970
1966
1968
1967
1969

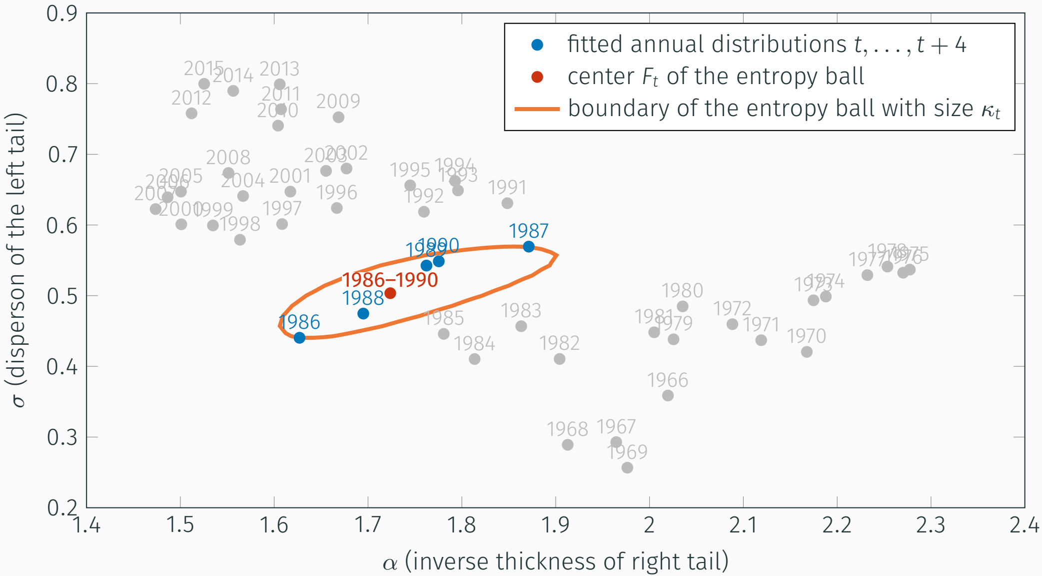

fitted annual distributions $t, \ldots, t+4$
center F_t of the entropy ball
boundary of the entropy ball with size κ_t
σ (disperson of the left tail)
α (inverse thickness of right tail)
1966
1966–1970
1967
1968
1969
1970



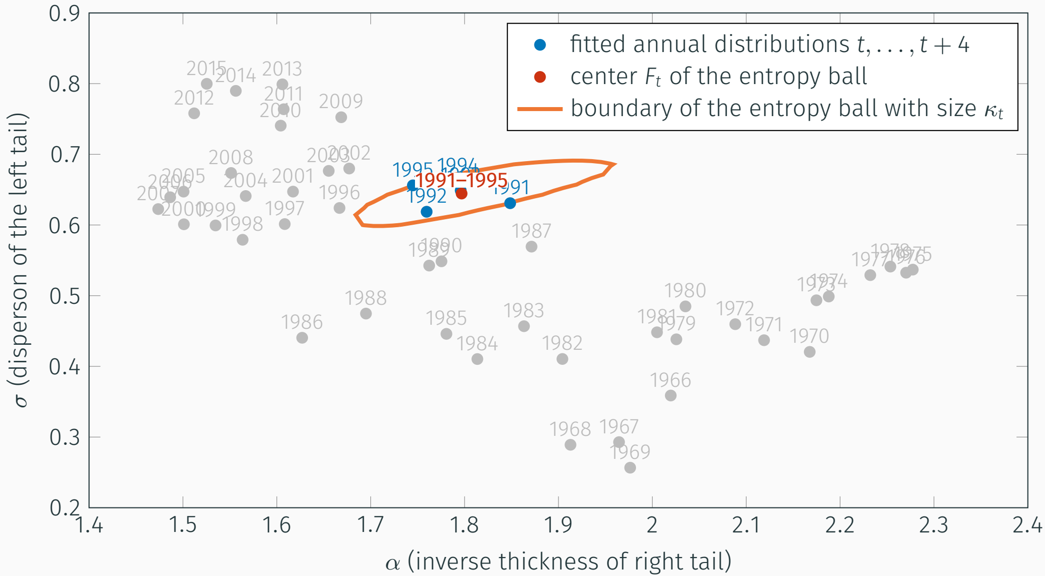
fitted annual distributions $t, \ldots, t+4$
center F_t of the entropy ball
boundary of the entropy ball with size κ_t
1971–1975
σ (disperson of the left tail)
α (inverse thickness of right tail)


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
2013
2009
2008
2001
1996
1997
1998
1995
1992
1991
1987
1988
1986
1985
1984
1983
1982
1966
1968
1967
1969
1972
1971
2015
2012
2004

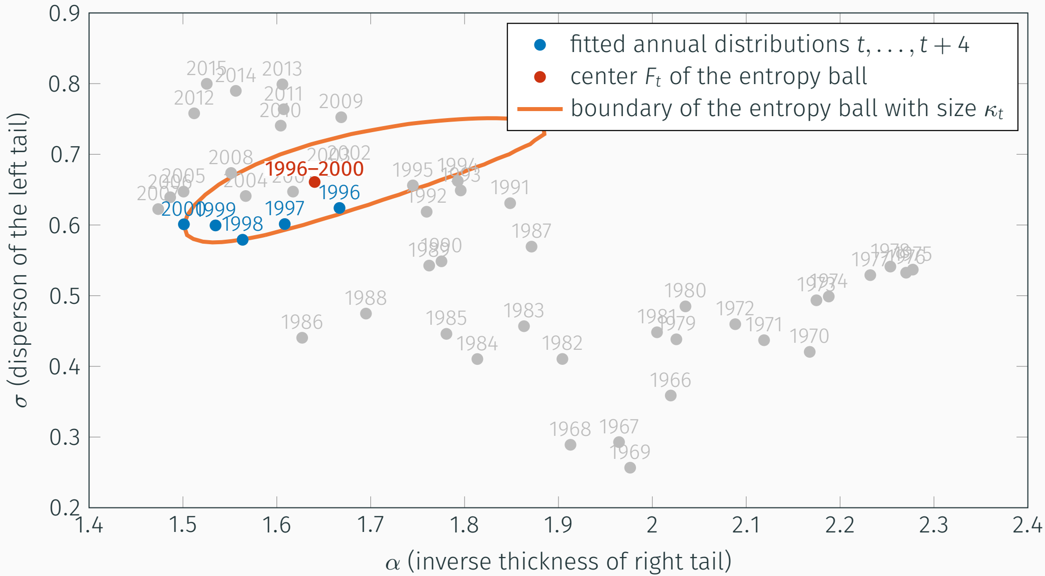

fitted annual distributions $t, \dots, t+4$
center F_t of the entropy ball
boundary of the entropy ball with size κ_t
1981–1985
1985
1984
1983
1982
1981
σ (disperson of the left tail)
α (inverse thickness of right tail)
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4
2015
2014
2013
2012
2011
2009
2008
2005
2004
2001
2002
1995
1994
1991
1996
1992
1998
1997
1987
1988
1986
1972
1971
1970
1966
1968
1967
1969

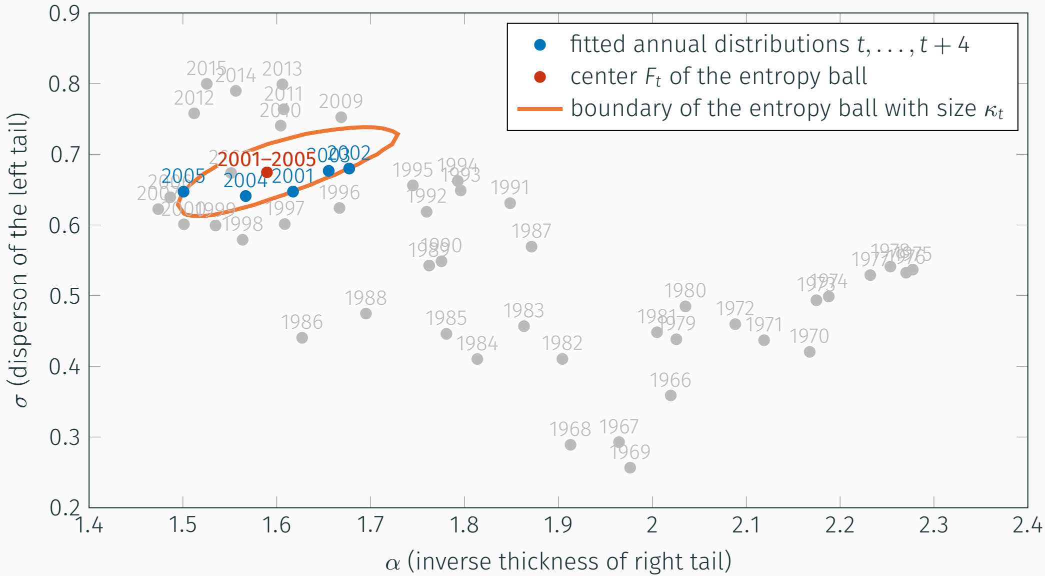


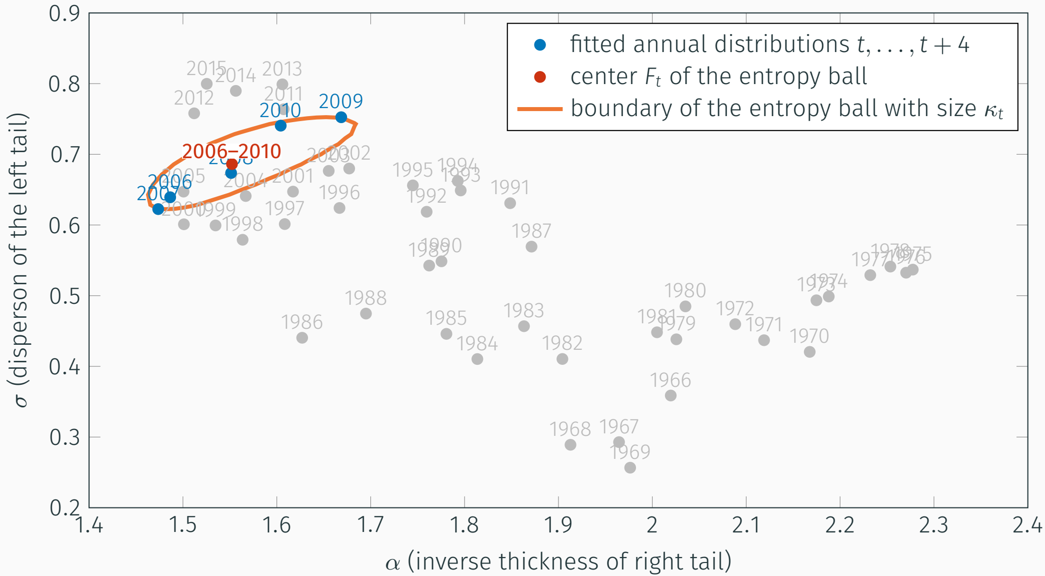
fitted annual distributions $t, \ldots, t+4$
center F_t of the entropy ball
boundary of the entropy ball with size κ_t
σ (disperson of the left tail)
α (inverse thickness of right tail)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4
1986–1990
1986
1987
1988
1989
1990
2015
2014
2013
2012
2011
2010
2009
2008
2005
2006
2007
2004
2001
2003
2002
2000
1999
1998
1997
1996
1995
1994
1993
1992
1991
1985
1984
1983
1982
1981
1980
1979
1972
1971
1970
1973
1974
1977
1978
1976
1975
1966
1968
1967
1969

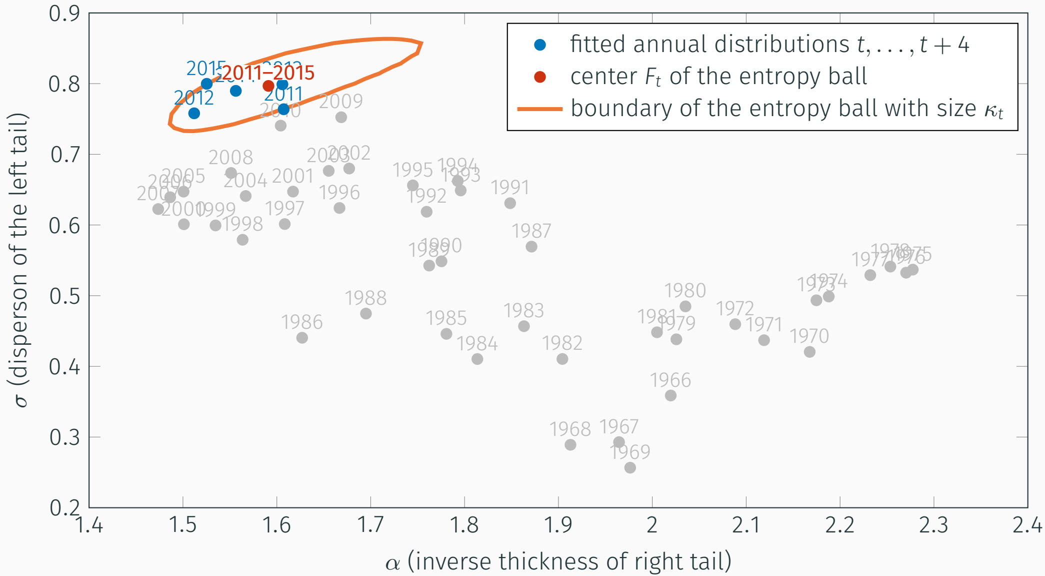


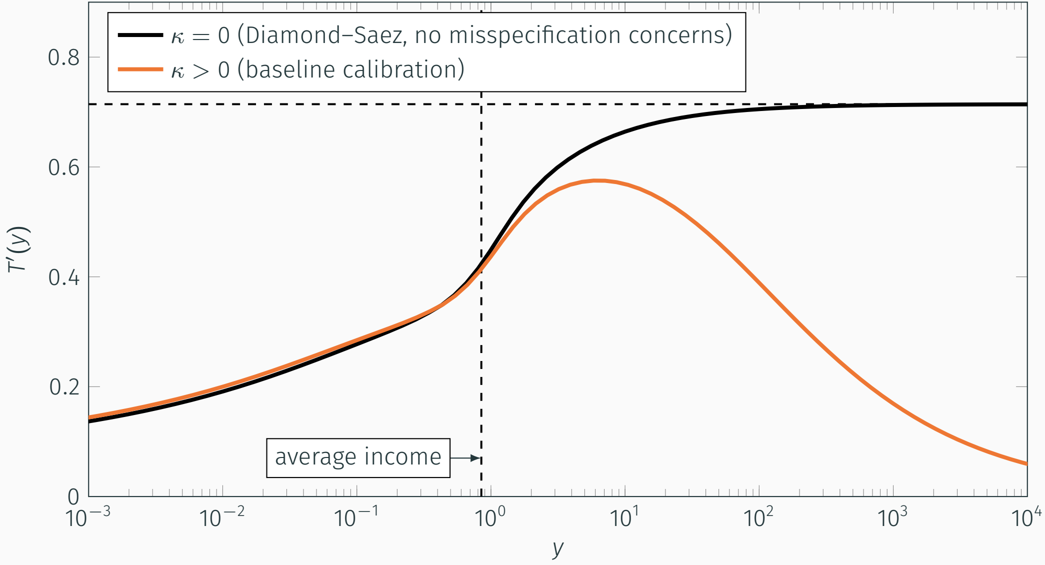
fitted annual distributions $t, \ldots, t+4$
center F_t of the entropy ball
boundary of the entropy ball with size κ_t
1991–1995
1995
1994
1992
1991
σ (disperson of the left tail)
α (inverse thickness of right tail)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4

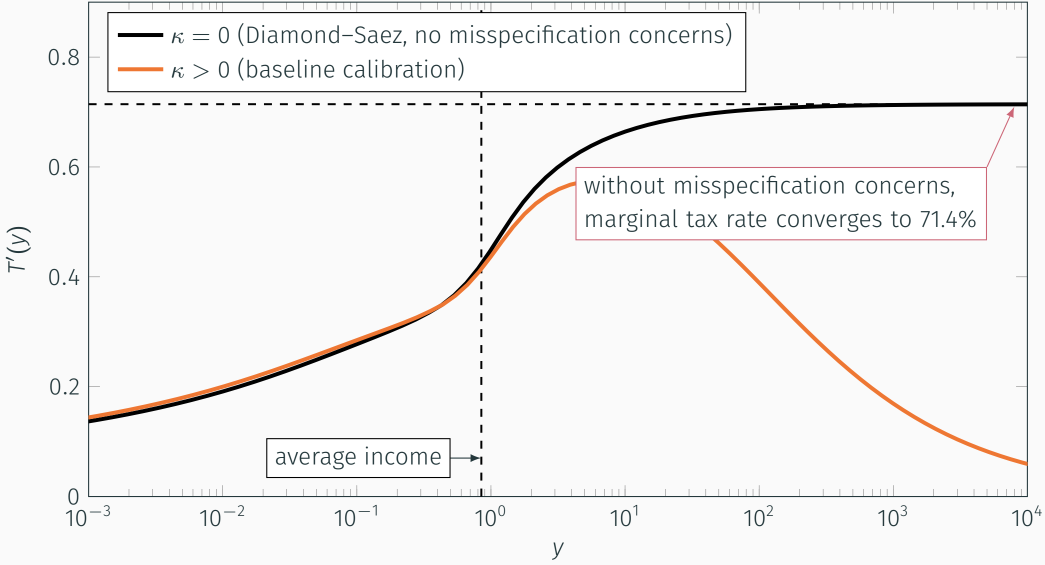

fitted annual distributions $t, \ldots, t+4$
center F_t of the entropy ball
boundary of the entropy ball with size κ_t
σ (disperson of the left tail)
α (inverse thickness of right tail)
1996–2000
1996
1997
1998
1999
2000
2015
2014
2013
2012
2011
2010
2009
2008
2005
2004
2002
1995
1994
1993
1992
1991
1990
1989
1988
1987
1986
1985
1984
1983
1982
1981
1980
1979
1978
1977
1976
1975
1974
1973
1972
1971
1970
1969
1968
1967
1966
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


fitted annual distributions $t, \dots, t+4$
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)
2001–2005
2005
2004
2001
2002
2015
2014
2013
2012
2011
2009
1996
1997
1998
1995
1994
1992
1991
1987
1988
1986
1985
1984
1983
1982
1980
1979
1972
1971
1970
1966
1968
1967
1969

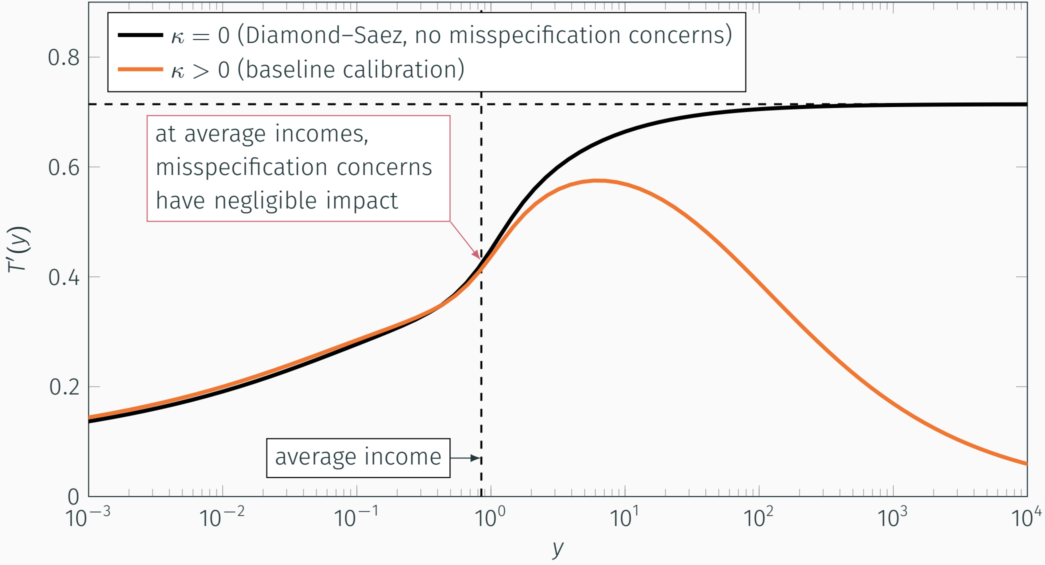

fitted annual distributions $t, \ldots, t+4$
center F_t of the entropy ball
boundary of the entropy ball with size κ_t
σ (disperson of the left tail)
α (inverse thickness of right tail)
2006–2010
2009
2010
2006
2007

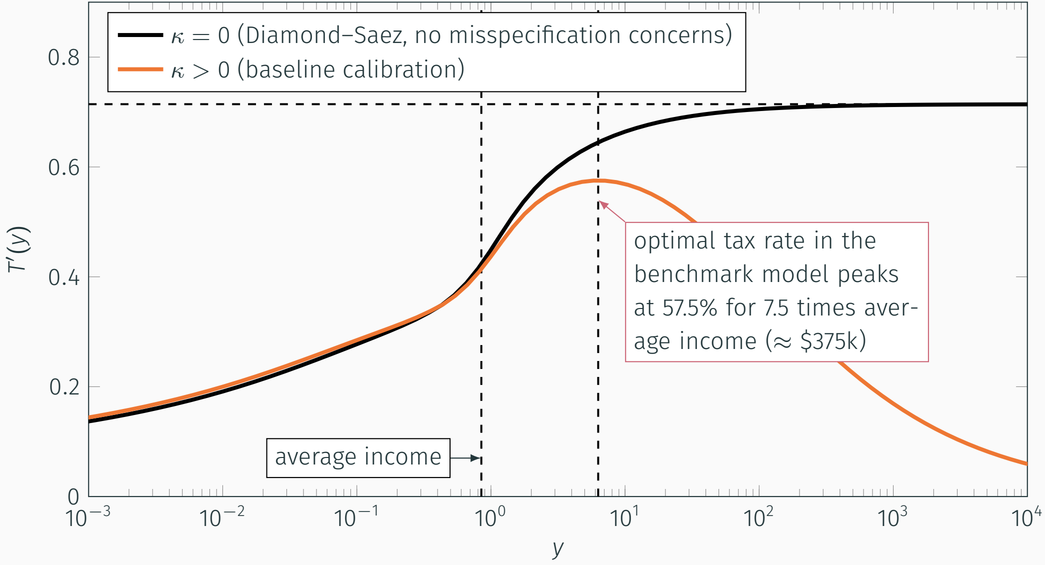

fitted annual distributions $t, \ldots, t+4$
center F_t of the entropy ball
boundary of the entropy ball with size κ_t
2011–2015
σ (disperson of the left tail)
α (inverse thickness of right tail)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4


$\kappa = 0$ (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


$\kappa = 0$ (Diamond–Saez, no misspecification concerns)
$\kappa > 0$ (baseline calibration)
without misspecification concerns, marginal tax rate converges to 71.4%
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

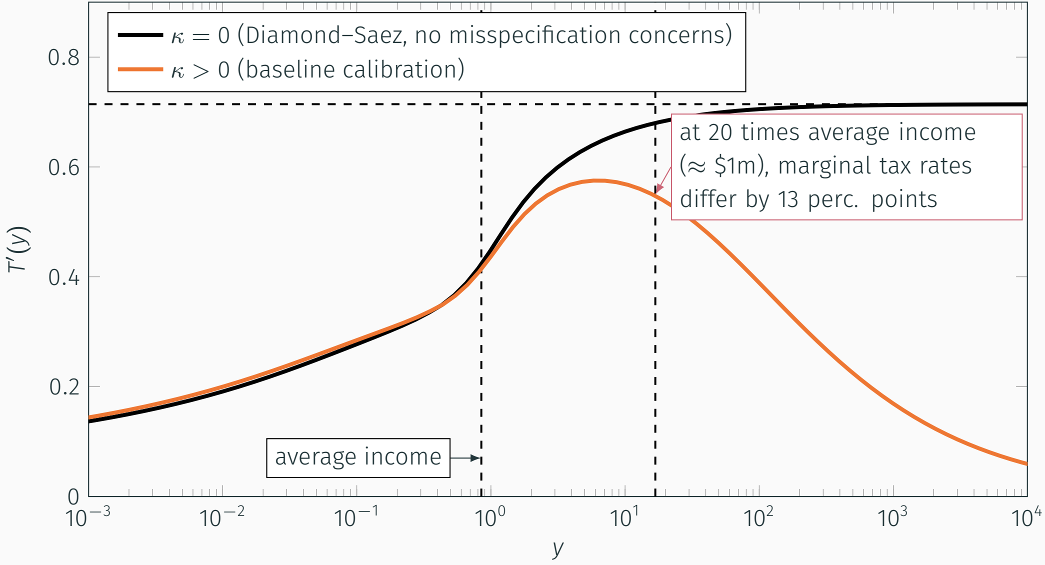
Optimal marginal tax schedules

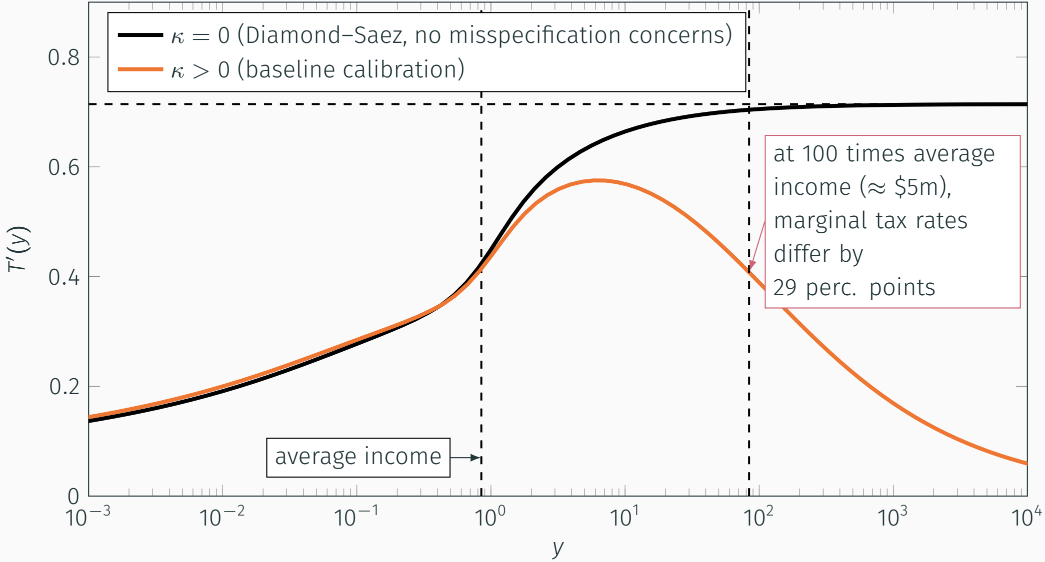


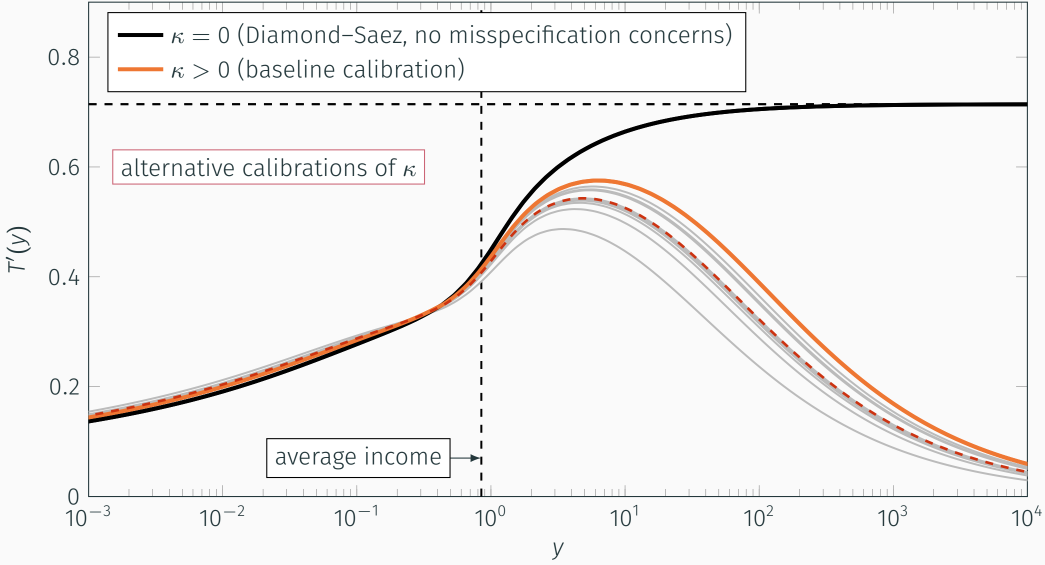


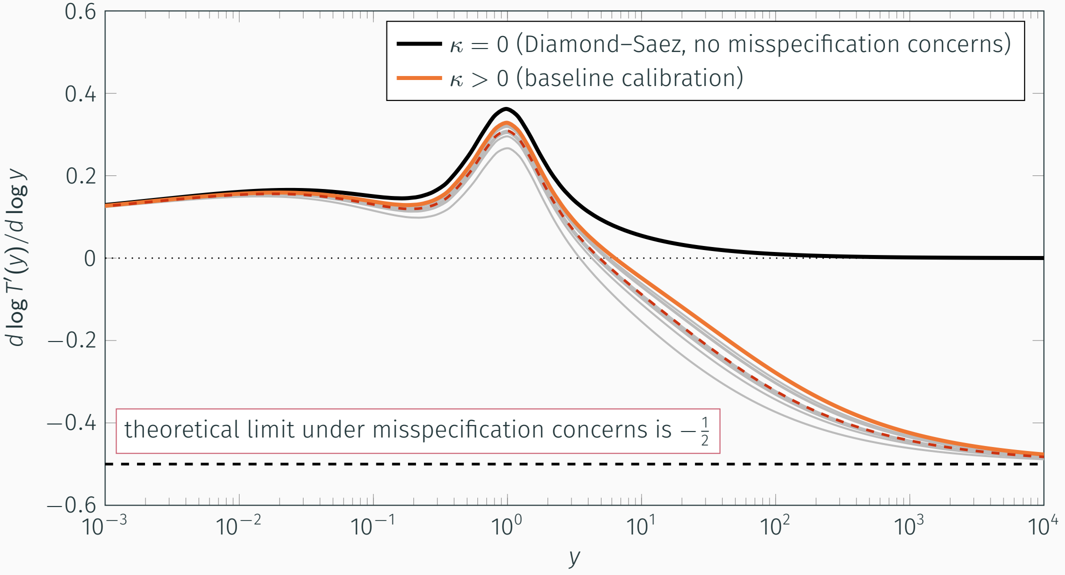
$\kappa = 0$ (Diamond–Saez, no misspecification concerns)
$\kappa > 0$ (baseline calibration)
0.8
0.6
0.4
0.2
0
$T'(y)$
optimal tax rate in the benchmark model peaks at 57.5% for 7.5 times average income (≈ $375k)
average income
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
y

Optimal marginal tax schedules




$\kappa = 0$ (Diamond–Saez, no misspecification concerns)
$\kappa > 0$ (baseline calibration)
at 100 times average income (≈ $5m), marginal tax rates differ by 29 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}


$\kappa = 0$ (Diamond–Saez, no misspecification concerns)
$\kappa > 0$ (baseline calibration)
alternative calibrations of κ
average income
$T'(y)$
0
0.2
0.4
0.6
0.8
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)
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

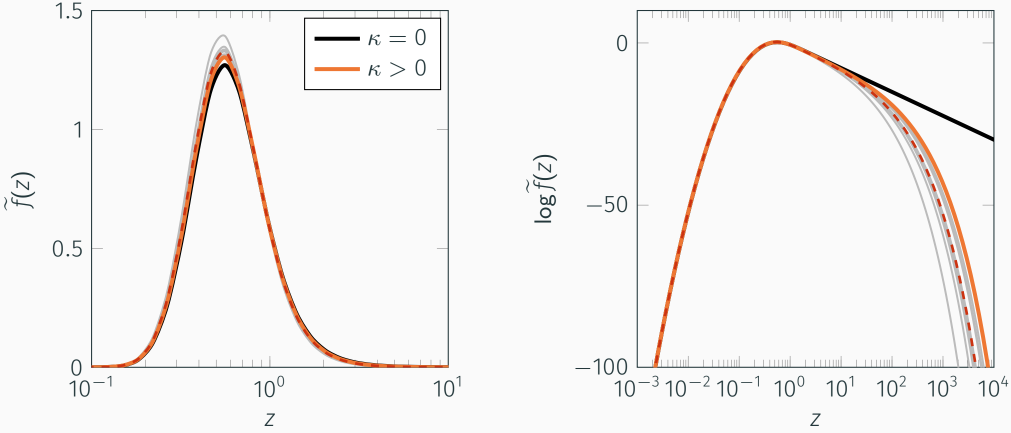
$$m(z) = \bar{m} \exp\left(-\frac{1}{\theta}\left[\mathcal{U}(z) + \mu T(y(z))\right]\right)$$

Left tail of the type distribution

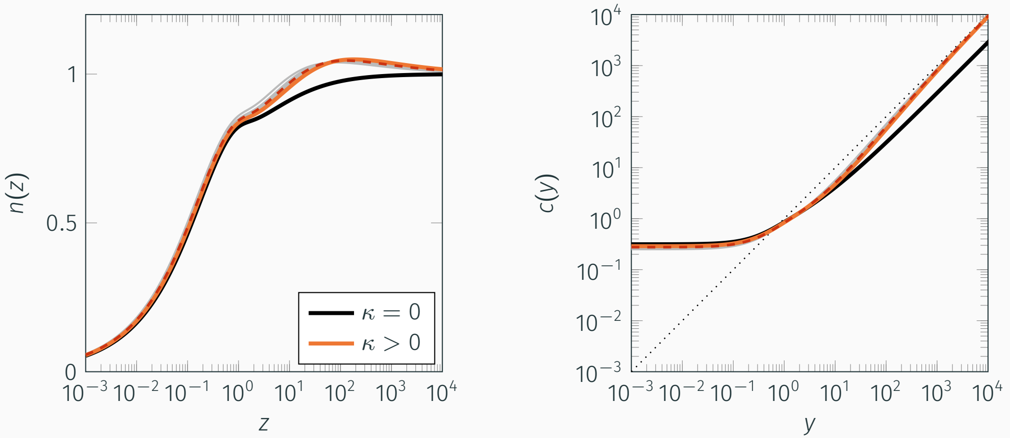
- without redistribution, we would have $\lim_{z \to 0} \mathcal{U}(z) = -\infty$, and $\lim_{z \to 0} m(z) = \infty$
- redistributive transfers bound $\mathcal{U}(z)$ from below, and so $m(z)$ is bounded above

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 $\widetilde{f}(z)$ 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 \ 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 \ 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 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
- under the benchmark distribution, the tax scheme generates additional resources $\mathbb{E}[T]$

Multidimensional type distribution

Uncertain productivity and labor supply elasticity distribution

Large disagreement over labor supply elasticities

- estimates vary across population studied, econometric approach etc.

Allow for uncertainty over a multidimensional type distribution

- joint distribution of productivity and labor supply elasticity

Next:

- incorporate multidimensional uncertainty using a modified penalty function
- calibrate using auxiliary data on Frisch elasticities
- analyze robust optimal tax policy in this setting

Let $f(z,\gamma)$ be the joint distribution of productivity z and the inverse Frisch elasticity γ

- factor the joint distributions $f(\gamma,z) = f^z(z) f^{\gamma|z}(\gamma|z)$ and $\tilde{f}(\gamma,z) = \tilde{f}^z(z)\tilde{f}^{\gamma|z}(\gamma|z)$

Proposed penalty: Let θ_γ, θ_z be two scalars

$$\text{Penalty}\left(f,\tilde{f}|\theta_\gamma,\theta_z\right) \equiv \theta_\gamma \int \underbrace{\mathcal{E}\left(f^{\gamma|z}(\cdot \mid z), \tilde{f}^{\gamma|z}(\cdot \mid z)\right)}_{\text{Entropy of cond. dist of } \gamma} m(z) f^z(z)\, dz$$

$$+ \theta_z \underbrace{\mathcal{E}\left(f^z, \tilde{f}^z\right)}_{\text{Entropy of marg. of } z}$$

Properties

- $\theta_\gamma = \theta_z = \theta$ implies $\text{Penalty}\left(f,\tilde{f}\right) = \theta\mathcal{E}\left(f,\tilde{f}\right)$
- $\theta_\gamma \to \infty$ recovers a version of the one-dimensional uncertainty about labor productivity
- $\theta_z \to \infty$ implies $m(z) = 1$

The min problem is reformulated as

$$\min_{m} \int \mathcal{U}(z,\gamma;T) m(\gamma,z) f(\gamma,z)\, d(\gamma,z) + V(G) + \text{Penalty}\left(f,\tilde{f} | \theta_\gamma,\theta_z\right)$$
$$\text{s.t.} \int m(z) f^{z}(z)\, dz = 1$$
$$\int m(\gamma \mid z) f^{\gamma|z}(\gamma \mid z)\, d\gamma = 1$$

where $\mathcal{U}(z,\gamma;T)$ is the indirect utility for type $s=(\gamma,z)$ given a proposed tax function T. The

robust planner chooses T, G subject to budget balance as before

Consider separable CES preferences from before: $\frac{c^{1-\rho}}{1-\rho} - \psi \frac{n^{1+\gamma}}{1+\gamma}$

Parameter γ affects both the level and elasticity of labor supply

Want to isolate uncertainty about the response of hours to a change in taxes

- consider a tax reform ΔT as a change from some status quo tax function $T^0 \to T$
- for small tax reforms we want small consequences of uncertainty about response to tax reforms

$$\Delta T \approx 0 \to m(\gamma|z) \approx 1$$

Hours vs Elasticity

Consider separable CES preferences from before: $\frac{c^{1-\rho}}{1-\rho} - \psi \frac{n^{1+\gamma}}{1+\gamma}$

Adjust preferences so that

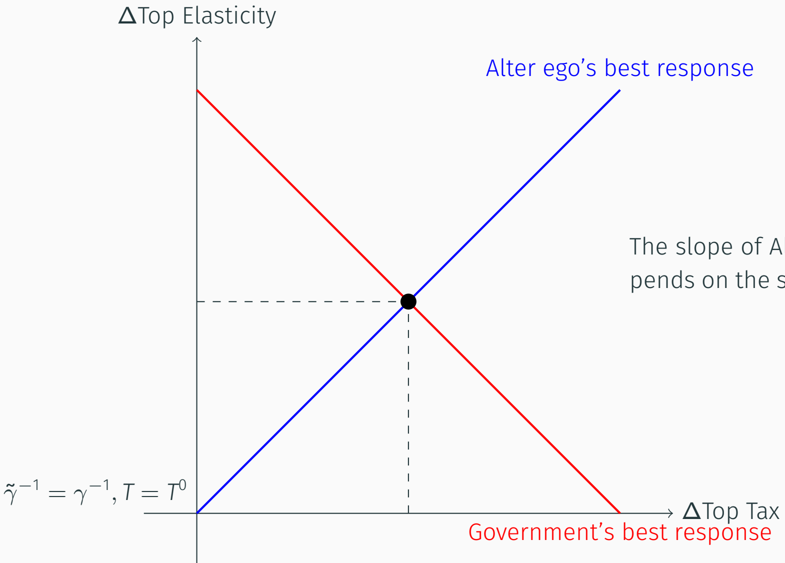
$$\frac{c^{1-\rho}}{1-\rho} - \Psi(\gamma, z) \frac{n^{1+\gamma}}{1+\gamma} - \Delta(\gamma, z)$$

Impose restrictions on $\Psi(\gamma, z)$ and $\Delta(\gamma, z)$ so that

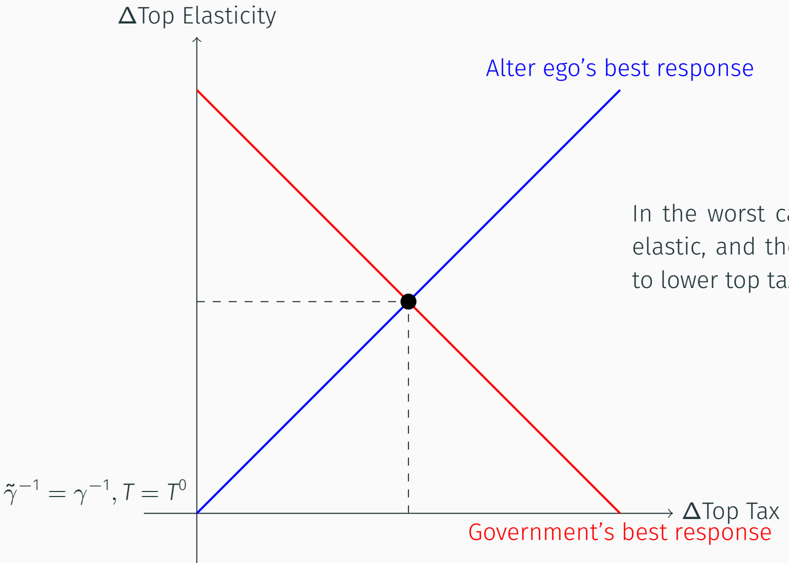
- when $T = T^0$, the allocation and utility levels are independent of γ given z $\Psi(\gamma, z)$ and $\Delta(\gamma, z)$

Implication: $T = T^0$ would give $m(\gamma|z) = 1$

Economic mechanism: Best response functions



The slope of Alter ego's best response depends on the strength of the wealth effect.'



In the worst case, high earners are more elastic, and the government's response is to lower top tax rates.

Baseline $f(\gamma, z)$: Covers elasticities reported in the micro and macro literature

- $f(\gamma \mid z) \sim$ log normal and independent of z
- $\mathbb{E}\gamma = 2$, and $Q_{75}(\gamma) - Q_{25}(\gamma) = 3 - \frac{1}{3}$

Penalty $\left(f, \tilde{f} | \theta_\gamma, \theta_z\right)$: Focus on only concerns about elasticities

- $\theta_z \to \infty$
- θ_γ so that worst case distribution implies high-earning individuals have elasticities in the range found by Rauh et al.

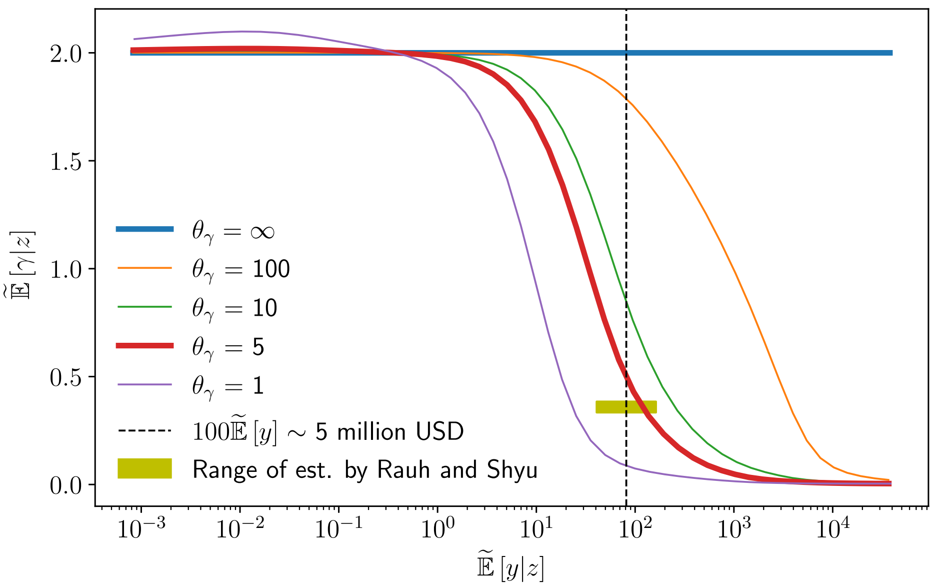
Restrict $T(y)$ to be in a class of functions Details of the restricted class

- Set T^0 to zero (robustness with T^0=U.S.)

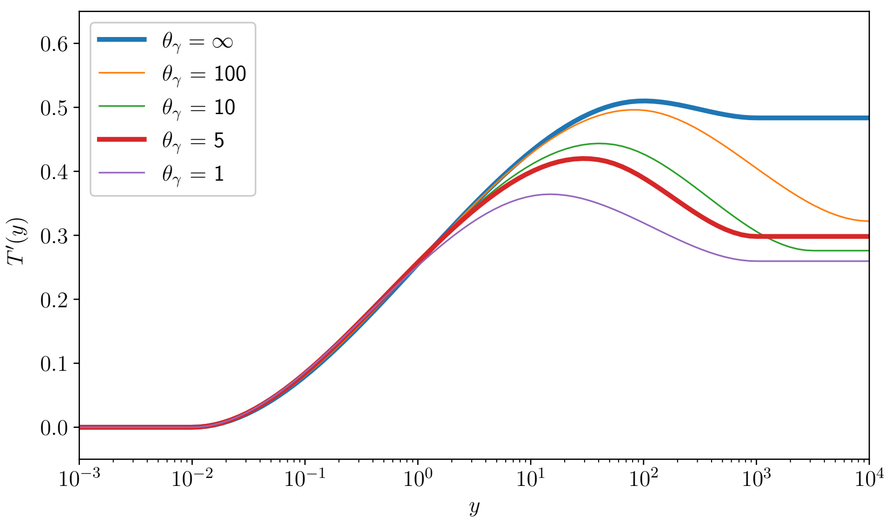
Calibrate the strength of wealth effect using evidence from lottery winnings $\rho = 0.56$

- All other parameters are same as before

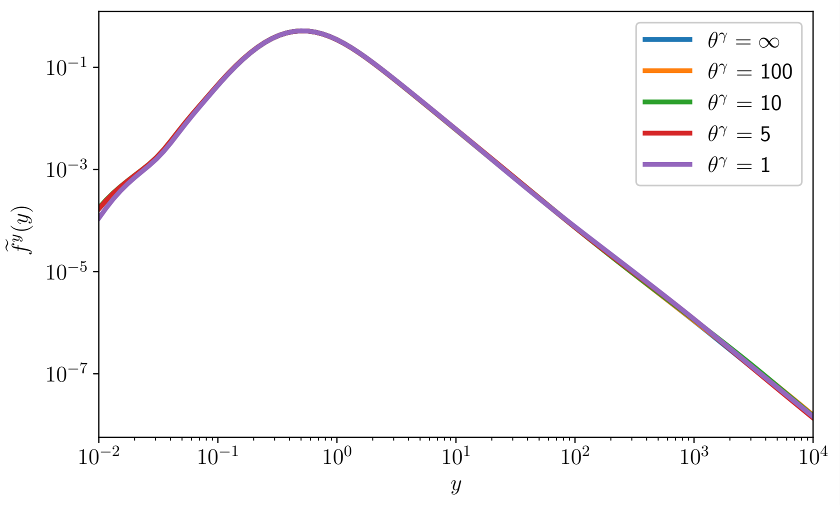
Results: Calibration of θ_γ



Results: Robust Planner lowers Top Taxes



Optimal Tax when T^0 = US



Income distribution when T^0 = US

Results: Intuition

Main finding: Uncertainty about elasticity $\rightarrow$ lower top taxes

- Baseline calibration: Δ top tax rate $\approx$ -15 p.p

Government worries about raising sufficient revenues

- 1D: not enough high productive people
- 2D: not enough inelastic people among high productive

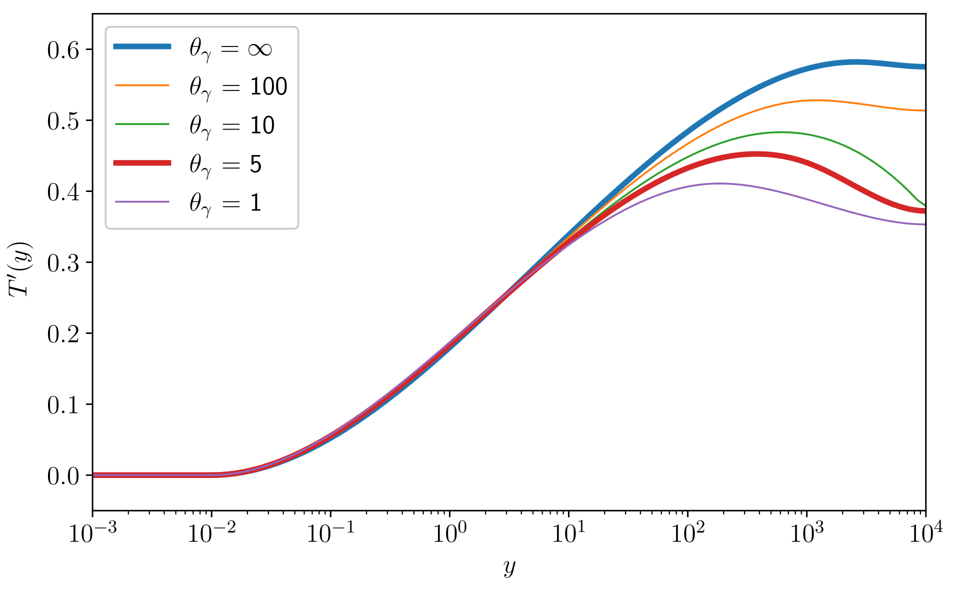
Endogenously correlation between elasticities and skills

- Baseline: $\gamma \perp z$ and $\mathbb{E}\gamma \approx 2$
- Worse case: $\gamma \not\perp z$ and $\tilde{\mathbb{E}}\gamma|z < 2$ for high earners

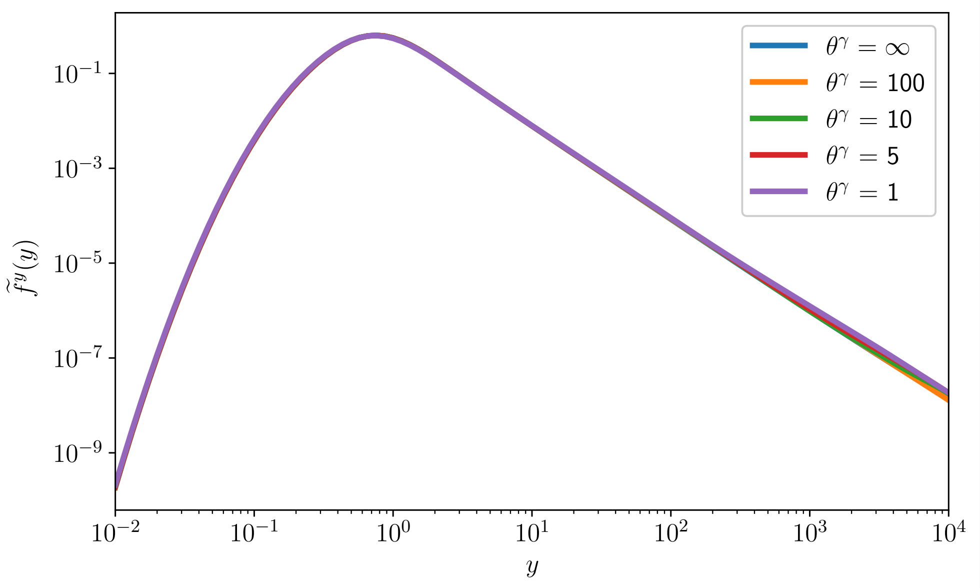
Income distribution mainly driven by skills and not distorted

moments $\backslash\theta_\gamma$	∞	100	10	5	1
$\mathbb{E}[z]$	1.00	1.00	1.00	1.00	1.00
$\widetilde{\mathbb{E}}[z]$	1.00	1.00	1.00	1.00	1.00
$\mathbb{E}[\gamma]$	2.00	2.00	2.00	2.00	2.00
$\widetilde{\mathbb{E}}[\gamma]$	2.00	2.00	1.99	1.99	1.95
$\mathbb{E}[\gamma \mid z \geq \bar{z}]$	2.00	2.00	2.00	2.00	2.00
$\widetilde{\mathbb{E}}[\gamma \mid z \geq \bar{z}]$	2.00	1.65	0.60	0.33	0.06
$\mathbb{E}[y]$	0.81	0.81	0.82	0.82	0.83
$\widetilde{\mathbb{E}}[y]$	0.81	0.81	0.81	0.81	0.82
μ	1.18	1.18	1.18	1.18	1.18
T_0	-0.17	-0.17	-0.17	-0.17	-0.16
$\mathbb{E}[T]$	-0.00	0.00	0.00	0.00	0.00
$\widetilde{\mathbb{E}}[T]$	-0.00	-0.00	-0.00	-0.00	-0.00
$\mathbb{E}[T]/\mathbb{E}[y]$	-0.00	0.00	0.00	0.00	0.00
$\widetilde{\mathbb{E}}[T]/\widetilde{\mathbb{E}}[y]$	-0.00	-0.00	-0.00	-0.00	-0.00

$\bar{z}$ is minimum z such that $\mathbb{E}[y \mid z] \geq 100\mathbb{E}[y]$


$\theta_\gamma = \infty$
$\theta_\gamma = 100$
$\theta_\gamma = 10$
$\theta_\gamma = 5$
$\theta_\gamma = 1$
0.6
0.5
0.4
0.3
0.2
0.1
0.0
$T'(y)$
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
y

Results: Income distribution $T^0 = \text{US}$



Choice for $\Psi(\gamma, z)$ and $\Delta(\gamma, z)$

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

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

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 $\Psi(\gamma, z) = \bar{\psi}^{\frac{\rho+\gamma}{\rho+\bar{\gamma}}} z^{1+\gamma-(1+\bar{\gamma})\frac{\gamma+\rho}{\bar{\gamma}+\rho}}$ and $\Delta(\gamma, z) = \frac{1}{1+\bar{\gamma}}\left(\frac{z^{1+\bar{\gamma}}}{\bar{\psi}}\right)^{\frac{1-\rho}{\bar{\gamma}+\rho}} - \frac{1}{1+\gamma}\left(\frac{z^{1+2\gamma-\bar{\gamma}}}{\bar{\psi}}\right)^{\frac{1-\rho}{\bar{\gamma}+\rho}}$ for some constants $\bar{\psi}, \bar{\gamma}$. If $T^0 = 0$, then for all γ', γ'', z we have

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

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

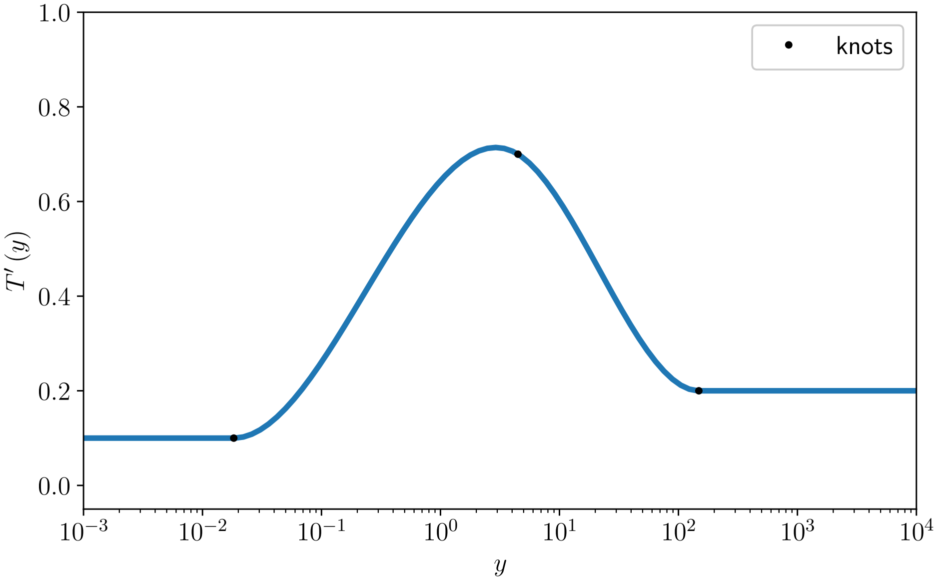
Parametrize $T'(y)$ using N knots $\{(\ln y_i, \tau_i)\}_{i=1,\cdots,N}$:

- 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

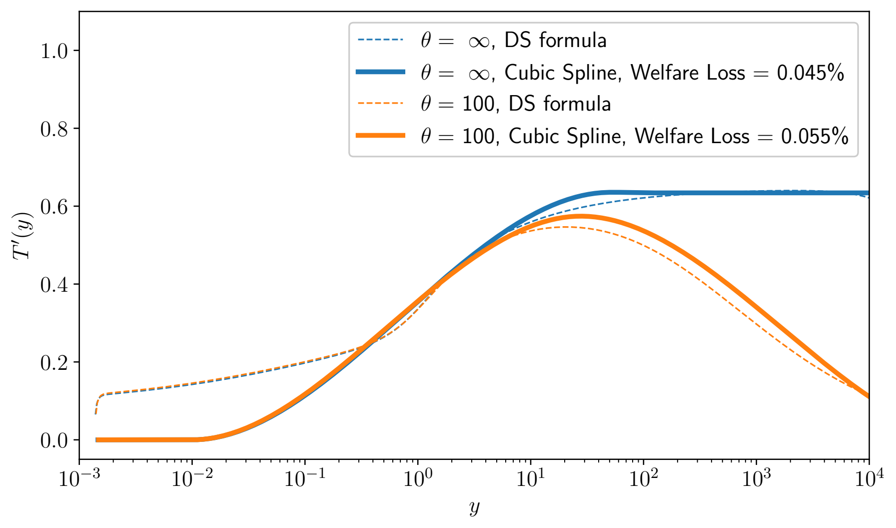
Well-known that multidimensional screening problems are difficult to characterize

- for 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



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

Curvature parameter ρ affects the size of income effect

- Workers' response to tax reform depends on the size of income effect

Use estimates from the lottery study in the US by Golosov, Graber, Mogstad, Novgorodsky (2023)

Under an affine tax $T(y) = \tau y - \text{tr}$, income effect is measured by

$$\frac{dy}{d\text{tr}} = -\frac{1}{\frac{\gamma}{\rho}\left(1 - \tau + \frac{\text{tr}}{y}\right) + 1 - \tau}$$

Golosov et al estimate: $\frac{dy}{d\text{tr}} \approx -0.367$ for a lottery winner

The rest is calibrated to high income earners in the US

- top marginal tax rate: $\tau = 0.40$,
- transfer is small relative to income for high income earners: $\text{tr}/y \approx 0$
- use the baseline value of labor supply elasticity: $\gamma = 2$

$\implies \rho = 0.56$

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

State-dependent misspecification concerns expressed by $\theta(z)$.

- administrative data and surveys are differentially informative about parts of the type distribution

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.

Literature II

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.

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