Robust Bounds on Optimal Tax Progressivity

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

January 2024

[1]University of Minnesota and NBER

[2]New York University and NBER

[3]University of Kent

Large literature on optimal tax design

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

Key predictions depend on hard to measure objects

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

Some information on these objects are available from administrative and survey data.

Optimal tax design acknowledging uncertainty about distribution of individual characteristics

- combine robust control approach of Hansen and Sargent (2001a,b) to model welfare consequences of statistical uncertainty about type distributions with Mirrlees (1971)
- quantify uncertainty using information from administrative and survey data

Key sources of uncertainty

- tails of the productivity distribution with scarce information relative to their welfare implications
- correlation of productivity and labor supply elasticity

Framework

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

- type may index productivity z, labor supply elasticity γ, ...

Households choose labor supply 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 $F(s)$

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

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

subject to the budget constraint

$$c = \underbrace{z(s)\,n}_{y = z(s)n} - T(z(s)\,n)\,.$$

Indirect utility function $\mathcal{U}(s; T)$ and decision rules $\mathcal{C}(s; T)$, $\mathcal{N}(s; T)$, $\mathcal{Y}(s; T)$.

Without uncertainty concerns, the utilitarian government is endowed with welfare function

$$\mathbb{E}\left[\psi\mathcal{U}\left(T\right)\right] = \int \psi\left(s\right)\mathcal{U}(s;T)dF\left(s\right)$$

- $\psi\left(s\right)$ is a Pareto/Negishi weighting function, normalized to $\mathbb{E}\left[\psi\right] = 1$

The government solves

$$\max_{T} \int \psi\left(s\right)\mathcal{U}(s;T)dF\left(s\right)$$

subject to the government budget constraint

$$\int T(\mathcal{Y}\left(s;T\right))dF\left(s\right) = B.$$

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

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

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

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

- $m(s) = \frac{d\widetilde{F}(s)}{dF(s)}$ be 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

A robust utilitarian government solves the max-min problem

$$\max_{T} \min_{\widetilde{F} \in \mathcal{F}} \int \psi(s) \mathcal{U}(s; T) d\widetilde{F}(s)$$

subject to

$$\int T(\mathcal{Y}(s; T)) d\widetilde{F}(s) = B.$$

A robust utilitarian government solves the max-min problem

$$\max_{T} \min_{m:\widetilde{F}\in\mathcal{F}} \int \psi(s)\,\mathcal{U}(s;T)\,m(s)\,dF(s)$$

subject to

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

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

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

subject to

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

Theoretical analysis: Scalar case

We now restrict attention to a scalar type $s = z$.

- typical case studied in much of the literature

The 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

$$\max_{c,y} \min_{\substack{m>0 \\ E[m]=1}} \int \psi(z) U\left(c(z), \frac{y(z)}{z}\right) m(z)\, dF(z) + \theta \int m(z) \log m(z)\, dF(z)$$

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

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

The robust government solves

$$\min_{\substack{m>0 \\ E[m]=1}} \max_{c,y} \int \psi(z) U\left(c(z), \frac{y(z)}{z}\right) m(z)\, dF(z) + \theta \int m(z) \log m(z)\, dF(z)$$

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

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

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 $\widetilde{F}(z)$ ($\min$ and $\max$ can be interchanged)

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

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

- 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 $T(y(z))$

We first focus on the theoretical characterization of top marginal tax rates.

- here, we for simplicity assume quasilinear utility

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

- insights carry over to general separable preferences

We then provide a quantitative evaluation.

- concave utility, type distribution calibrated to data
- discipline the amount of uncertainty the planner faces

Optimal marginal tax schedule is given by the Diamond (1998)–Saez (2001) 'ABC' formula

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

- (A): adverse effect of taxes on labor supply via labor supply elasticity
- (B): desire to redistribute

$$\widetilde{\Psi}(z) = \int^{z} \frac{\psi(\zeta)\widetilde{f}(\zeta)}{\int \psi(\xi)\widetilde{f}(\xi)\,d\xi} d\zeta$$

- (C): tradeoff between labor supply distortion at z and revenue from taxing types above z

Assume planner puts zero welfare weight on top households

- $\psi(z) = 0$ for $z \geq \hat{z}$ where $\hat{z}$ is some threshold

Benchmark distribution in the tail above $\hat{z}$ is Pareto with shape parameter α

- then $(1 - F(z)) / (zf(z)) = \alpha^{-1}$

Without misspecification concerns, the tax formula for $z \geq \hat{z}$ simplifies to

$$\frac{T'(y(z))}{1 - T'(y(z))} = \frac{1 + \gamma}{\alpha}.$$

- with a fat-tailed type distribution, taxes at the top are nonzero and quantitatively possibly large
- intuition: the tax revenue from types above z outweighs the labor supply distortion at z

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

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

$$T(y(z)) = T(y(\underline{z})) + \int_{y(\underline{z})}^{y(z)} T'(\eta)d\eta$$

$$m(z) = \bar{m}\exp\left(-\frac{\mu}{\theta}T(y(z))\right)$$

- when $T'(y(z))$ decays to zero faster, $T(y(z))$ grows more slowly
- the distortion $m(z)$ thins out the density at the top more gradually
- the optimal tax formula then implies a slower 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 $\theta < \infty$. 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

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

- if (1.1) were not true, then $T(y(z))$ could be bounded by linear growth, and

$$\tilde{f}(z) = m(z) f(z) = \bar{m} \exp\left(-\frac{\mu}{\theta} T(y(z))\right) f(z)$$

  has a thin right tail
- (1.2) pins down the shape of $T'(y)$ that solves the fixed point argument

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

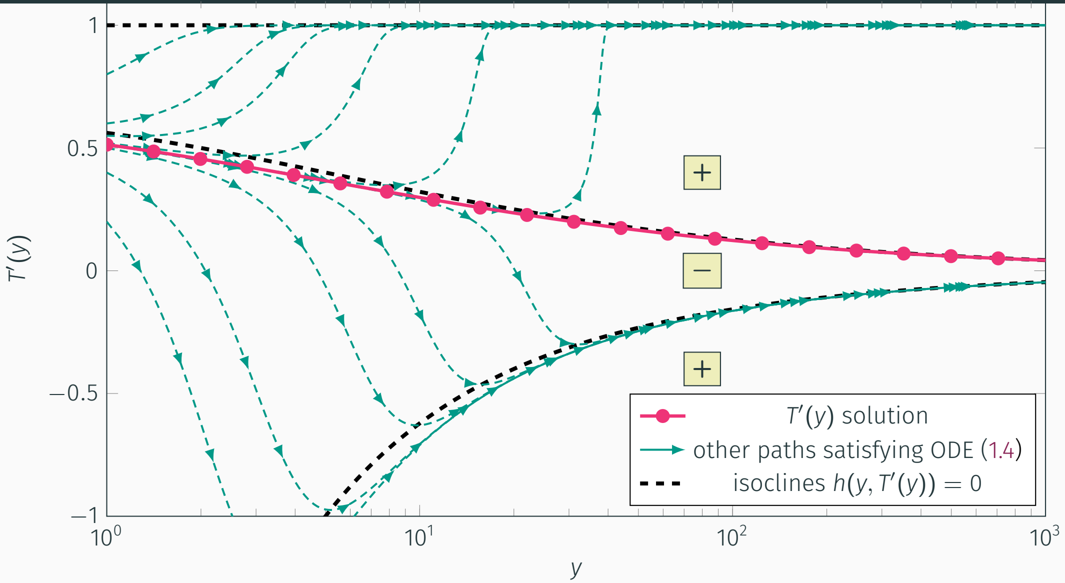
$$-\frac{T''(y)\,y}{1-T'(y)} = -\left[2-\frac{1+\gamma+\alpha}{1+\gamma}T'(y)\right]^{-1}\left[\frac{\mu}{\theta}\left[T'(y)\right]^{2}y-\gamma+\gamma\frac{1+\gamma+\alpha}{1+\gamma}T'(y)\right] \tag{1.3}$$

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
10^0
10^1
10^2
10^3
y
$T'(y)$
+
−
+
$T'(y)$ 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

Calibration

Preferences and technology

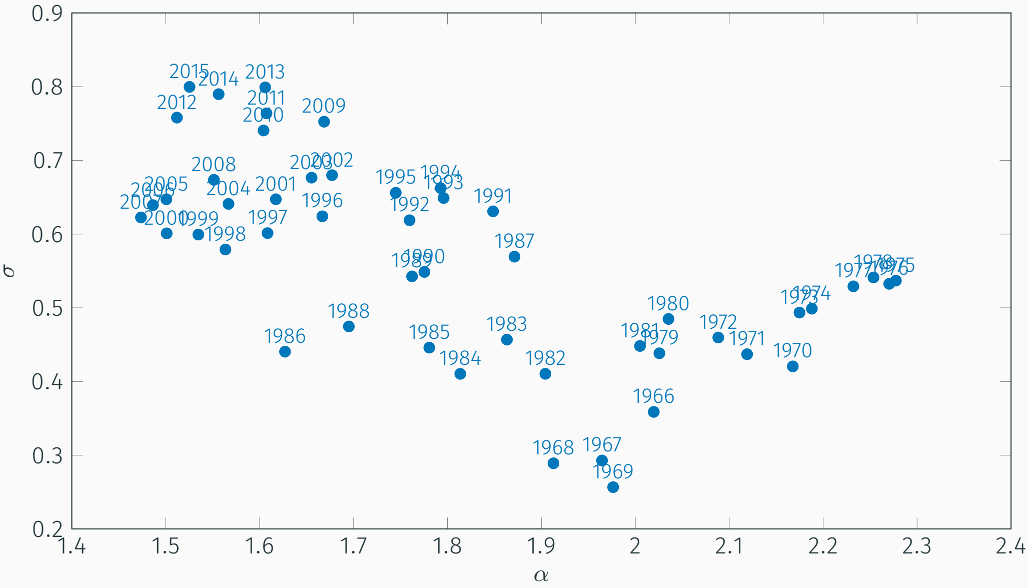
- isoelastic preferences: $U(c, n) = \frac{c^{1-\rho}}{1-\rho} - \chi \frac{n^{1+\gamma}}{1+\gamma}$ with $\rho = 1$, $\chi = 1$, $\gamma = 2$
- government spending $B = 0$

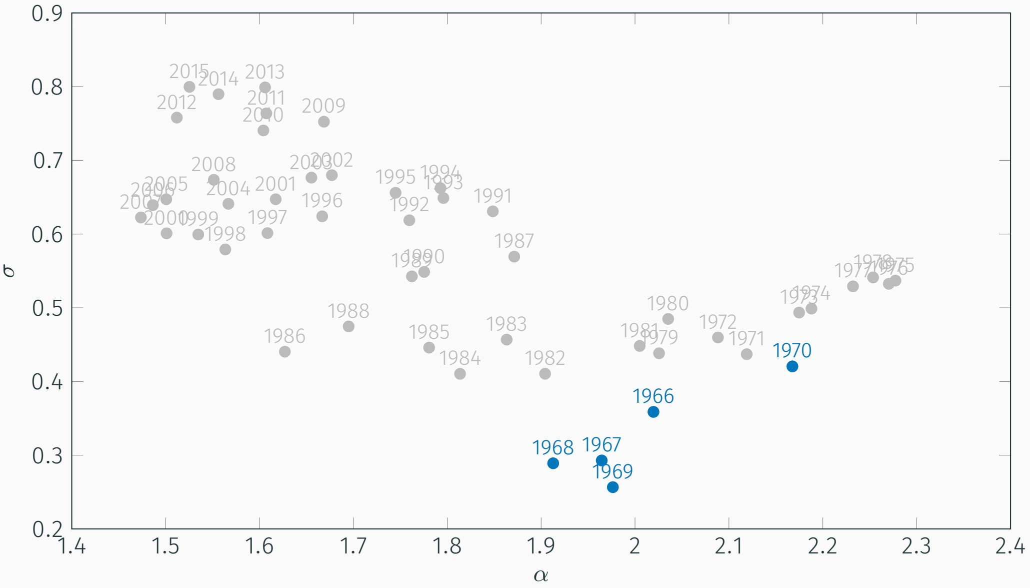
Benchmark distribution F

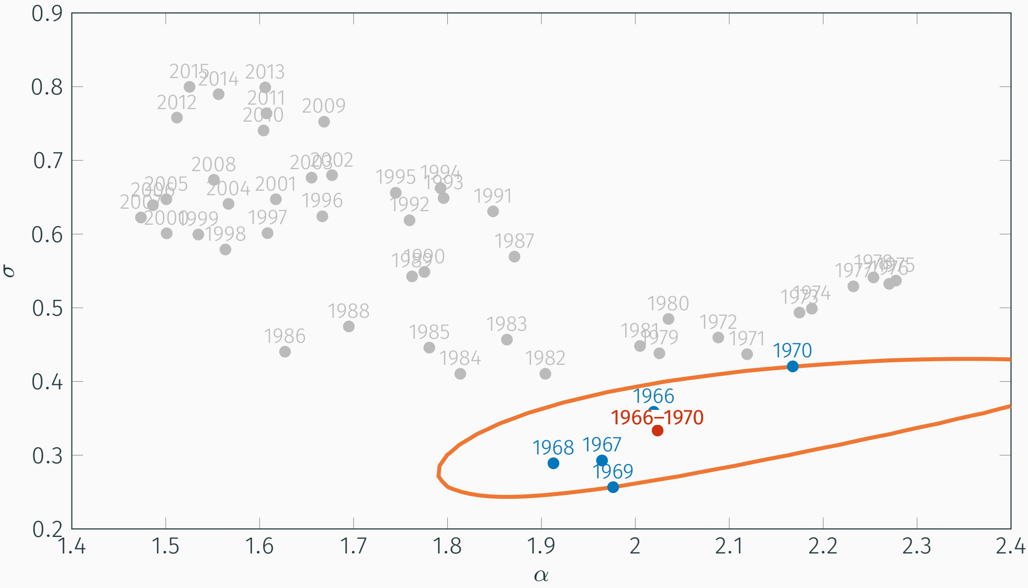
- $\log z$ has exponentially modified Gaussian (EGM) distribution (Heathcote and Tsujiyama (2021))
- left tail of z distribution is lognormal (parameters μ, σ)
- right tail approximately Pareto (parameter α)

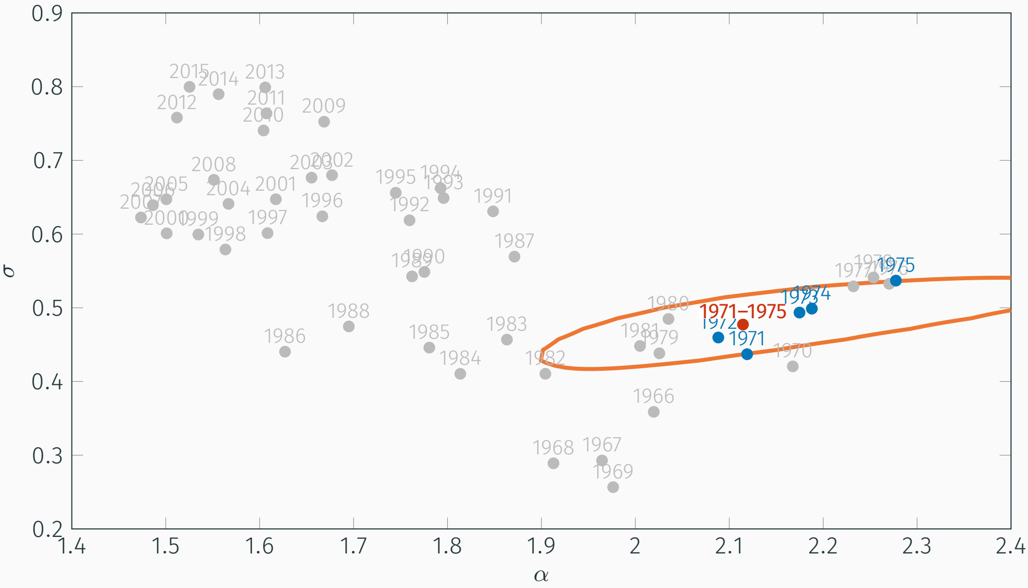
Entropy bound κ

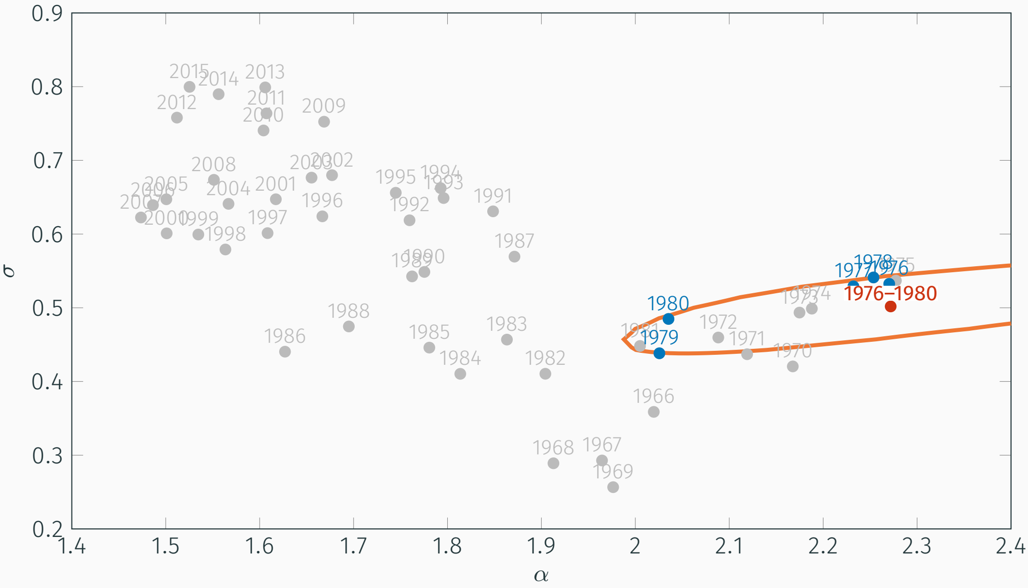
- use time-series variation in observed income distributions (World Income Database)
- alternative: use survey data and detection error probabilites

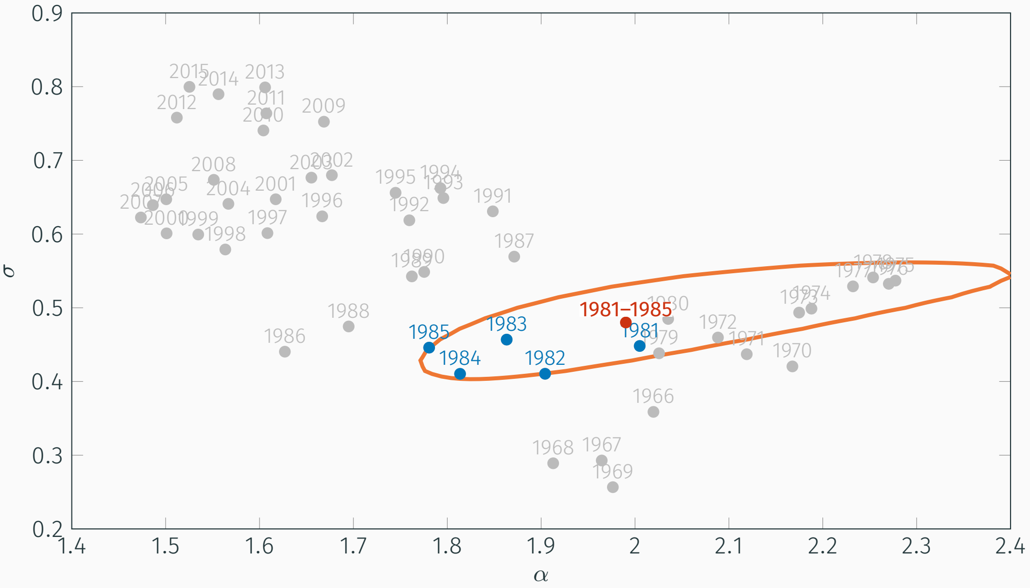

0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
σ
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4
α
2015
2014
2013
2012
2011
2009
2008
2002
2004
2001
1995
1991
1996
1992
1997
1998
1987
1988
1986
1985
1983
1984
1982
1980
1972
1971
1970
1966
1968
1967
1969


0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
σ
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4
α
2015
2014
2013
2012
2011
2010
2009
2008
2005
2004
2001
2002
1995
1994
1993
1996
1992
1991
1999
1998
1997
1987
1990
1988
1986
1985
1983
1984
1982
1980
1981
1979
1972
1971
1974
1970
1966
1968
1967
1969

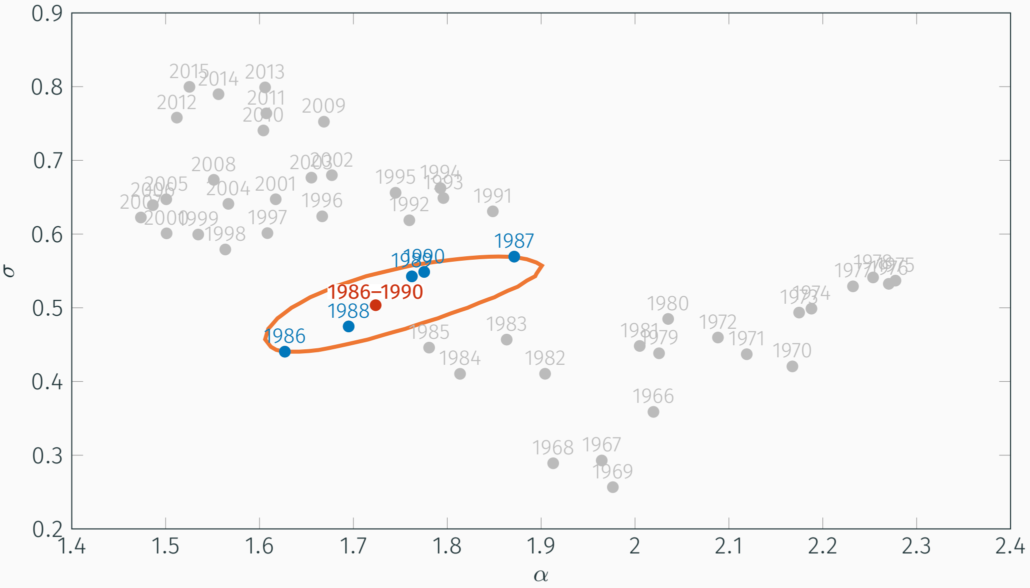

0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
σ
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4
α
2015
2014
2013
2012
2011
2009
2008
2002
2005
2004
2001
1996
1995
1991
1992
1999
1997
1998
1987
1986
1988
1985
1983
1984
1982
1980
1979
1972
1971
1970
1974
1966
1966–1970
1968
1967
1969

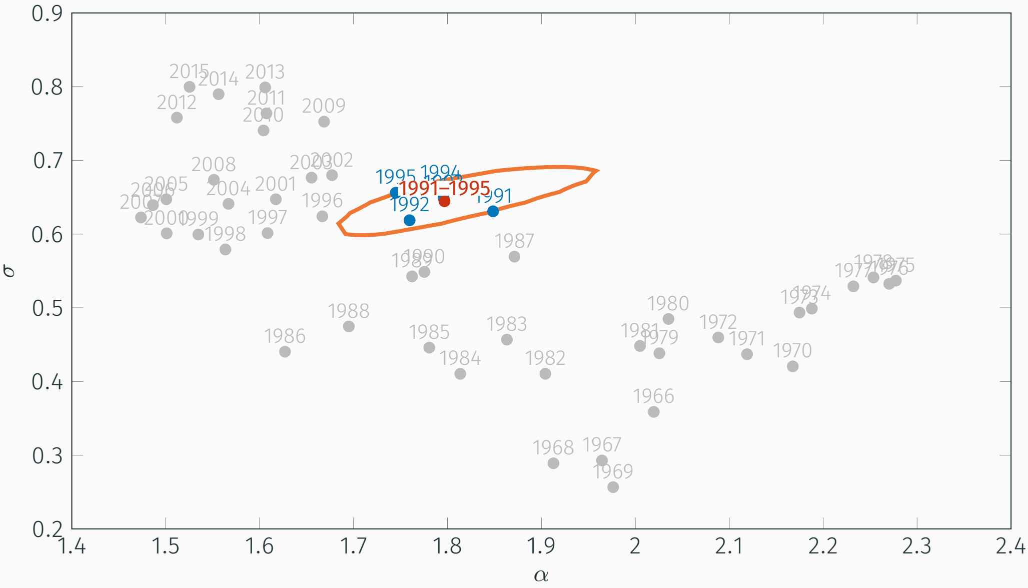

σ
α
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4
1971–1975
1971
1975
2013
2014
2015
2012
2009
2008
2005
2004
2001
1996
1995
1991
1992
1997
1998
1987
1988
1986
1985
1983
1984
1980
1979
1970
1966
1968
1967
1969

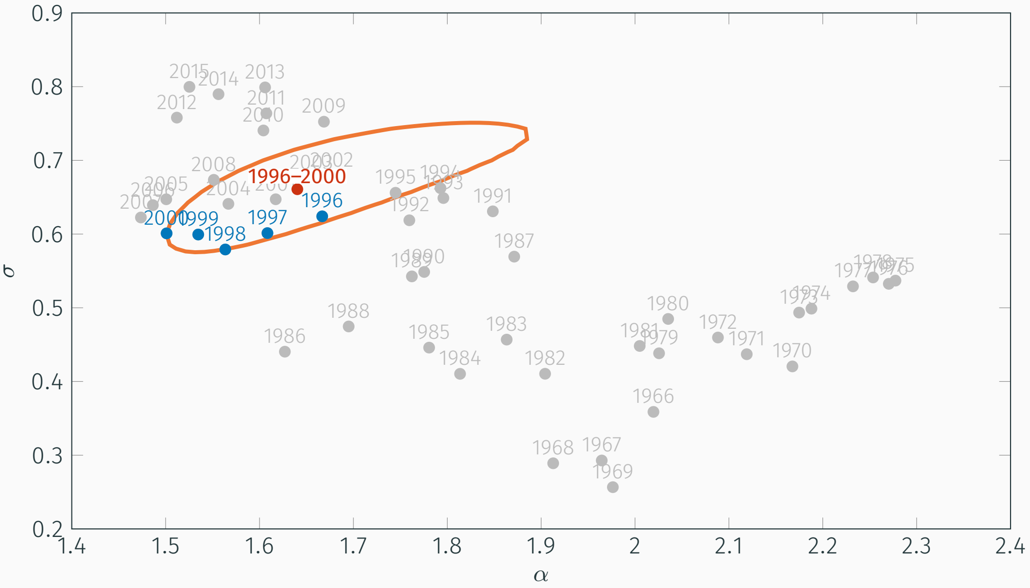

0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
σ
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4
α
1980
1979
1978
1977
1976
1976–1980
2015
2014
2013
2012
2011
2010
2009
2008
2005
2004
2001
2003
2002
1995
1994
1993
1996
1992
1991
1999
1998
1997
1987
1989
1990
1988
1986
1985
1984
1983
1982
1981
1972
1971
1970
1973
1974
1966
1968
1967
1969

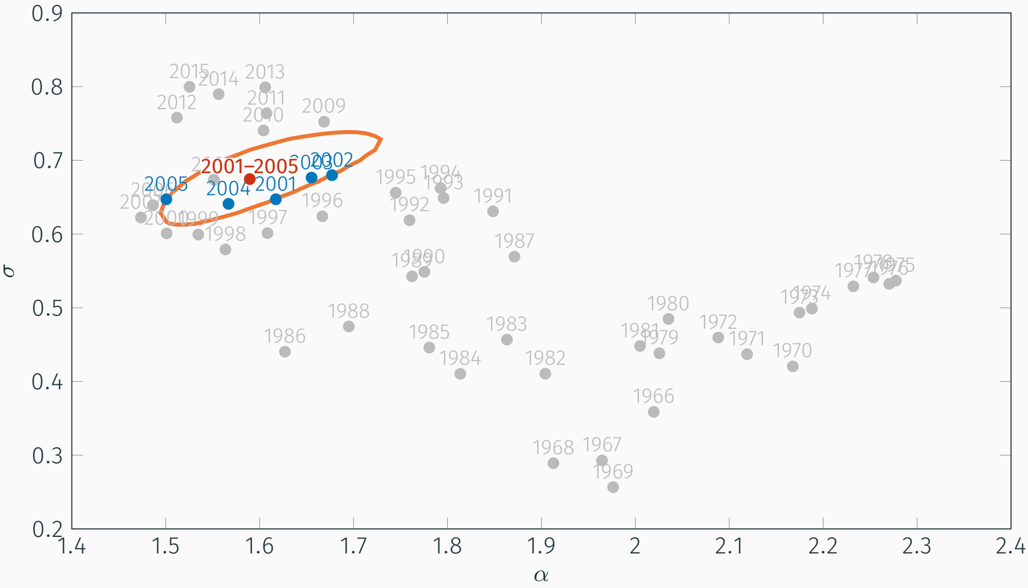


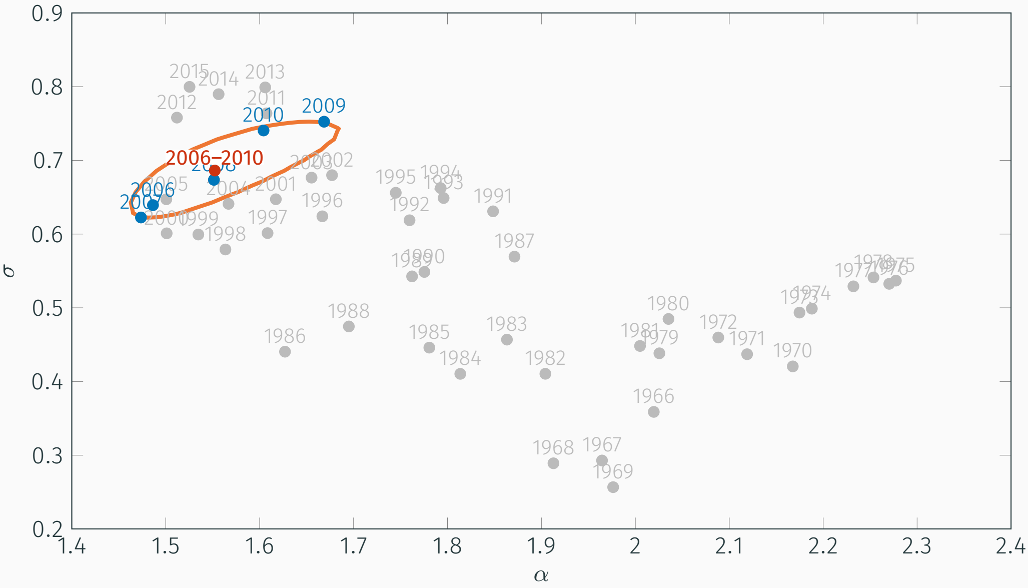
σ
α
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4
2015
2014
2013
2012
2011
2009
2008
2005
2004
2001
2002
1995
1996
1992
1991
1997
1998
1987
1988
1986
1985
1983
1984
1982
1981–1985
1981
1979
1972
1971
1970
1966
1968
1967
1969

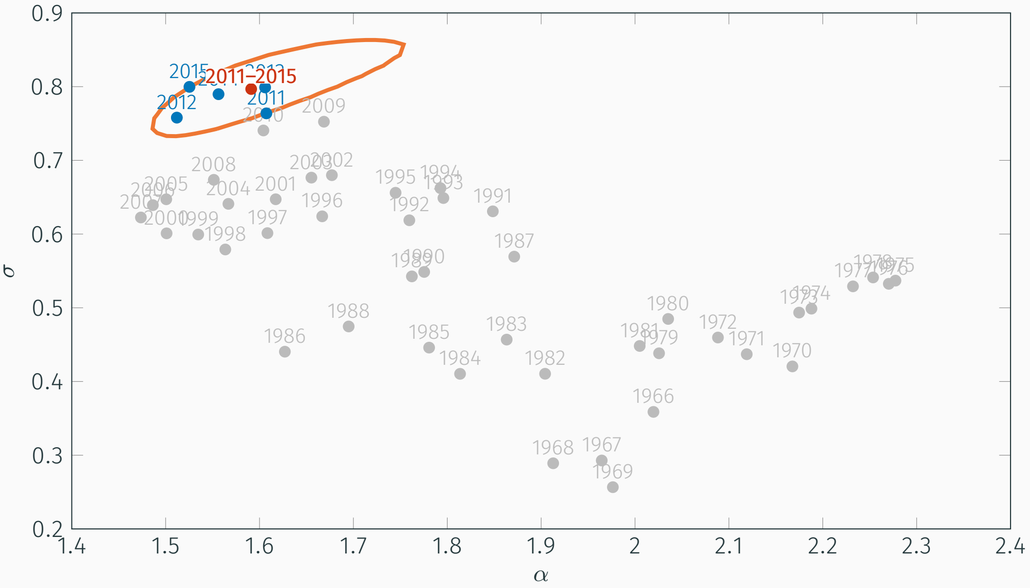

0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4
α
σ
2015
2014
2013
2012
2011
2009
2008
2002
2005
2004
2001
1996
1999
1998
1997
1995
1994
1993
1992
1991
1987
1990
1989
1986–1990
1988
1986
1985
1984
1983
1982
1980
1979
1972
1971
1970
1974
1975
1976
1966
1968
1967
1969


σ
α
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
1991–1995
1995
1994
1992
1991
2015
2014
2013
2012
2011
2010
2009
2008
2005
2004
2001
2002
1996
1999
1998
1997
1987
1990
1989
1988
1986
1985
1984
1983
1982
1981
1980
1979
1972
1971
1970
1974
1973
1977
1976
1975
1966
1968
1967
1969

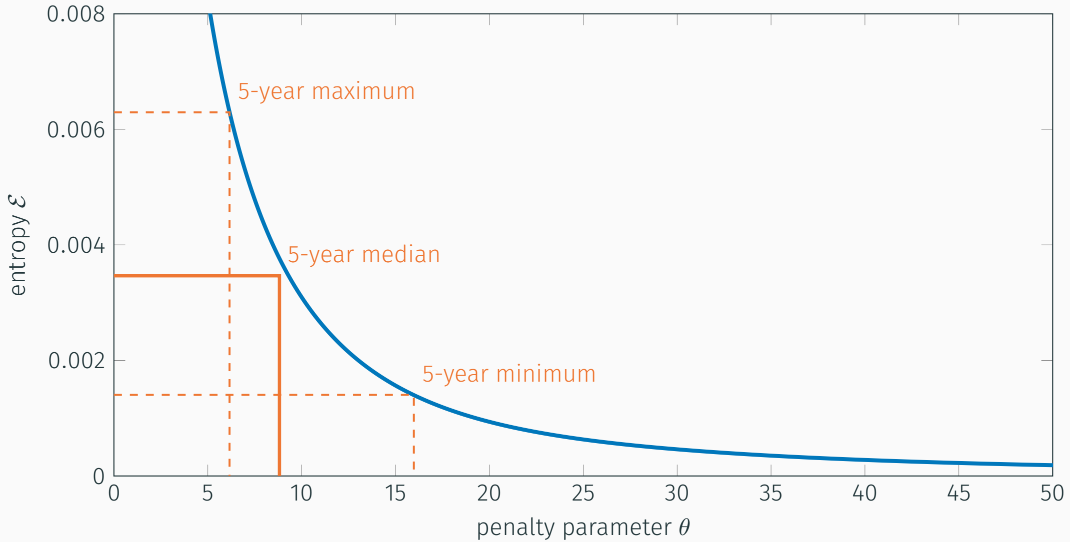

0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
σ
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4
α
1996–2000
1996
1997
1998
1999
2000
2015
2014
2013
2012
2011
2010
2009
2008
2005
2004
2002
1995
1992
1991
1987
1988
1986
1985
1984
1983
1982
1980
1979
1972
1971
1970
1966
1968
1967
1969


0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
σ
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4
α
2001–2005
2005
2004
2001
2002
2015
2014
2013
2012
2011
2009
1996
1995
1992
1991
1987
1997
1998
1988
1986
1985
1984
1983
1982
1980
1972
1971
1970
1966
1968
1967
1969


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


0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
σ
1.4
1.5
1.6
1.7
1.8
1.9
2
2.1
2.2
2.3
2.4
α
2015
2011–2015
2012
2011
2009
2008
2005
2004
2001
2002
1995
1991
1996
1992
1997
1998
1987
1988
1986
1985
1984
1983
1982
1980
1979
1972
1971
1970
1974
1966
1968
1967
1969

Mapping from entropy to θ



Alternative approach: Use survey data and detection error probabilities

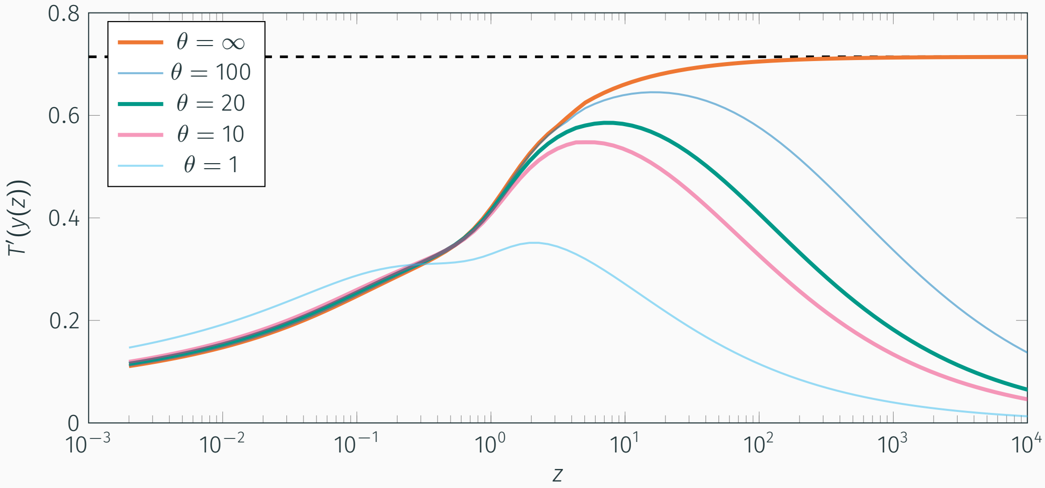
We utilize the concept of detection error probability from Anderson et al. (2003).

- measures how distinguishable are $\widetilde{f}(z)$ and $f(z)$ using a given finite sample of data of size I
- probability that a sample drawn from $f(z)$ is favored by the likelihood of $\widetilde{f}(z)$, and vice versa

Calibrate the detection error $\bar{d}$ to sample sizes in available surveys.

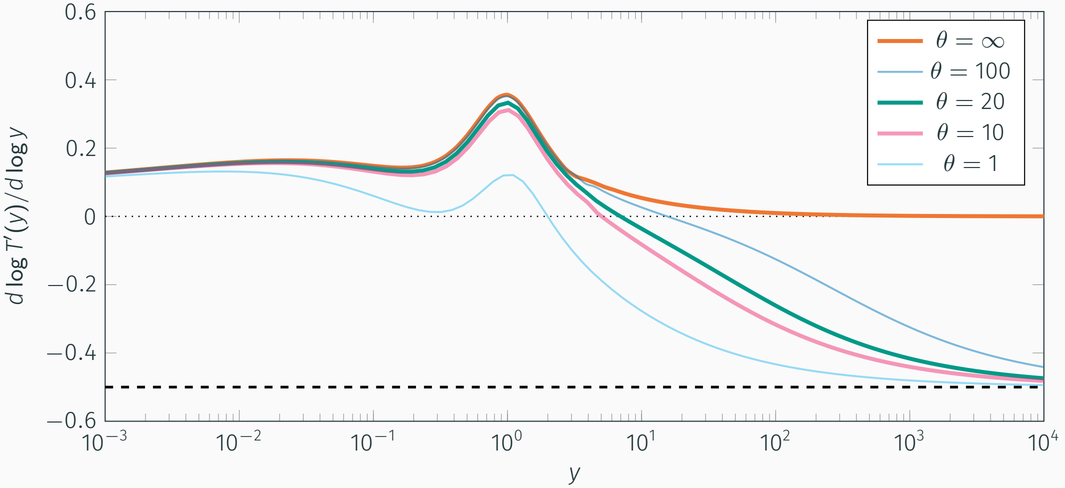
- Survey of Consumer Finances $I \approx 3,500$.
- pick θ to achieve detection error probability $\bar{d} = 10\%$

Details on DEPs

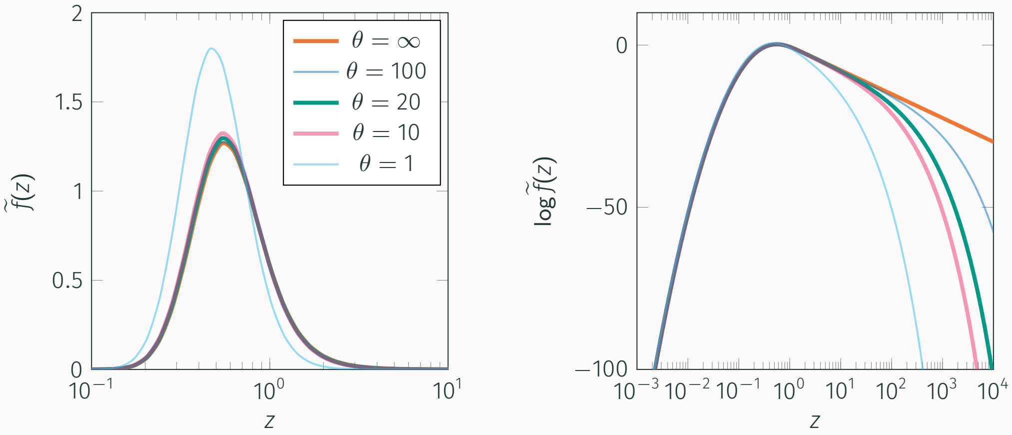


- without misspecification concerns, marginal tax rate converges to 71.4% (dashed line)

Elasticity of marginal tax rate



- theoretical limit under misspecification concerns is $-\frac{1}{2}$ (dashed line)



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

The worst-case density is characterized by the distortion

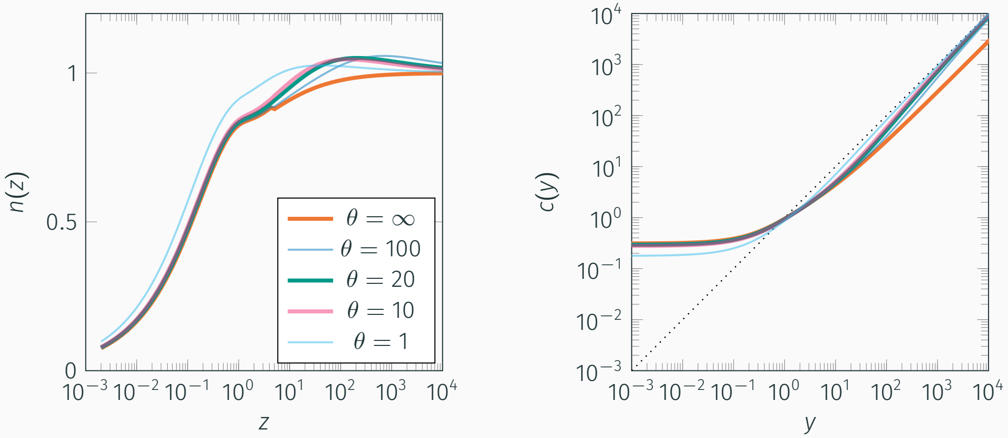
$$m(z) = \bar{m}\exp\left(-\frac{1}{\theta}\left[\mathcal{U}(z) + \mu T(y(z))\right]\right)$$

Right tail of the type distribution

- dominated by budgetary concerns since $\rho > \frac{1}{2}$
- since $\lim_{z\to\infty} T(y(z)) = \infty$, we also have $\lim_{z\to\infty} m(z) = 0$

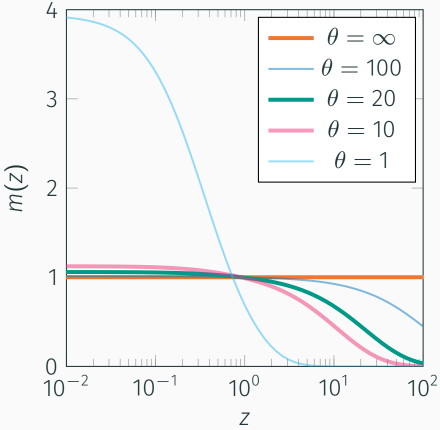
Left tail of the type distribution

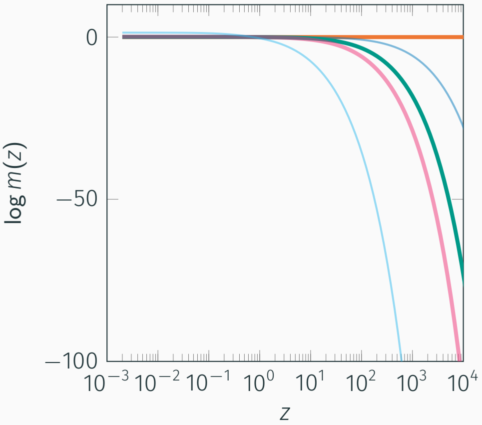
- without redistribution, we would have $\lim_{z\to 0}\mathcal{U}(z) = -\infty$, and $\lim_{z\to 0} m(z) = \infty$
- but redistributive transfers bound $\mathcal{U}(z)$ from below, and so $m(z)$ is bounded above



$n(z)$
1
0.5
0
10^{-3} 10^{-2} 10^{-1} 10^{0} 10^{1} 10^{2} 10^{3} 10^{4}
z
$\theta = \infty$
$\theta = 100$
$\theta = 20$
$\theta = 10$
$\theta = 1$
$c(y)$
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

Worst-case distortion





Moments under benchmark and worst-case distributions

moments \ θ	∞	100	20	10	1
$E[z]$	1.000	1.000	1.000	1.000	1.000
$\widetilde{E}[z]$	1.000	0.986	0.951	0.918	0.657
$E[y]$	0.824	0.828	0.838	0.848	0.918
$\widetilde{E}[y]$	0.824	0.815	0.792	0.770	0.579
μ	1.220	1.229	1.261	1.297	1.726
T_0	−0.311	−0.307	−0.292	−0.278	−0.178
$\max_y T'(y)$ (%)	71.4	64.5	58.6	55.0	35.2
$\arg\max_y T'(y)$	∞	15.541	6.955	4.609	1.992
$E[T]$	0.000	0.008	0.023	0.037	0.110
$\widetilde{E}[T]$	0.000	0.000	0.000	0.000	0.000

Multidimensional type distribution

Uncertain productivity and labor supply elasticity distribution

We now incorporate uncertainty about joint type distribution $s = (z, \gamma)$.

- productivity z, labor supply elasticity γ^{-1}

We also consider different sources of information on the type distribution.

1. Data on the joint distribution of $s = (z, \gamma)$.
   - for example, SCF with personal characteristics
2. Data on the marginal distribution of income $\mathcal{Y}\left(s; T^{US}\right)$
   - for example, income data from a tax authority

The joint type distributions $f(z,\gamma)$ and $\widetilde{f}(z,\gamma)$ have likelihood ratio

$$m(z,\gamma) = \frac{\widetilde{f}(z,\gamma)}{f(z,\gamma)}.$$

Given tax function $T(y)$, type distributions imply income distributions

$$f^y(y;T) = \int f(z,\gamma)\frac{\partial}{\partial y}\mathcal{Z}(y,\gamma;T)\,d\gamma \qquad \widetilde{f}^y(y;T) = \int \widetilde{f}(z,\gamma)\frac{\partial}{\partial y}\mathcal{Z}(y,\gamma;T)\,d\gamma$$

- $\mathcal{Z}(y,\gamma;T)$ is the inverse to the income function $\mathcal{Y}(z,\gamma;T)$.

Income distributions $f^y(y;T)$ and $\widetilde{f}^y(y;T)$ have likelihood ratio

$$m^y(y;T) = \frac{\widetilde{f}^y(y;T)}{f^y(y;T)}.$$

Entropy penalty for deviations from the joint distribution

$$\theta \int m(z,\gamma) \log m(z,\gamma)\, dF(z,\gamma)\,.$$

Entropy penalty for deviations from the income distribution

$$\theta^y \int m^y\left(y;T^{US}\right) \log m^y\left(y;T^{US}\right) dF\left(y;T^{US}\right)$$

- plausible distributions $\widetilde{f}(z,\gamma)$ should generate plausible distributions $\widetilde{f}^y\left(y;T^{US}\right)$
  … under the existing tax code $T^{US}(y)$

Penalty parameters θ and θ^y calibrated to survey and administrative data.

- perfect knowledge of $f^y\left(y;T^{US}\right)$ still preserves uncertainty about $f(z,\gamma)$

Robust government now solves

$$\max_{T} \min_{\substack{m>0 \\ E[m]=1}} \int \psi(s)\mathcal{U}(s;T) m(s)\, dF(s) + \theta \int m(s) \log m(s)\, dF(s) + \theta^{y} \int m^{y}\left(y;T^{US}\right) \log m^{y}\left(y;T^{US}\right) dF\left(y;T^{US}\right)$$

subject to

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

We restrict attention to parameterized tax functions

$$T(y) = -\tau_0 + y - \tau_1 y^{1-\tau_2}$$

- Heathcote et al. (2017), Heathcote and Tsujiyama (2021)

Preferences

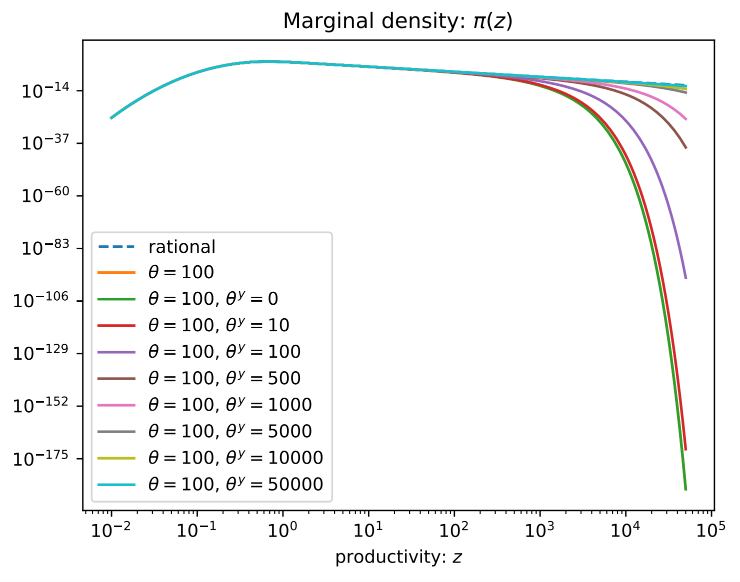
- $\rho = 1$, $\chi = 1$, $\gamma = 2$

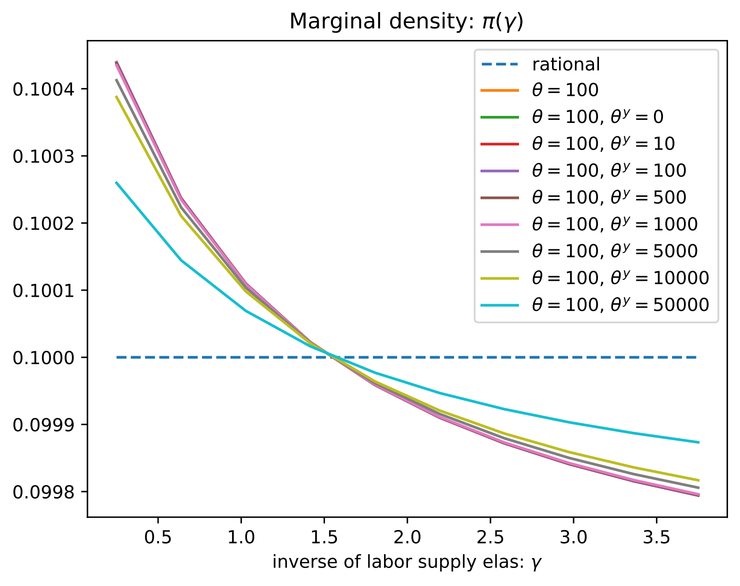
Type distribution

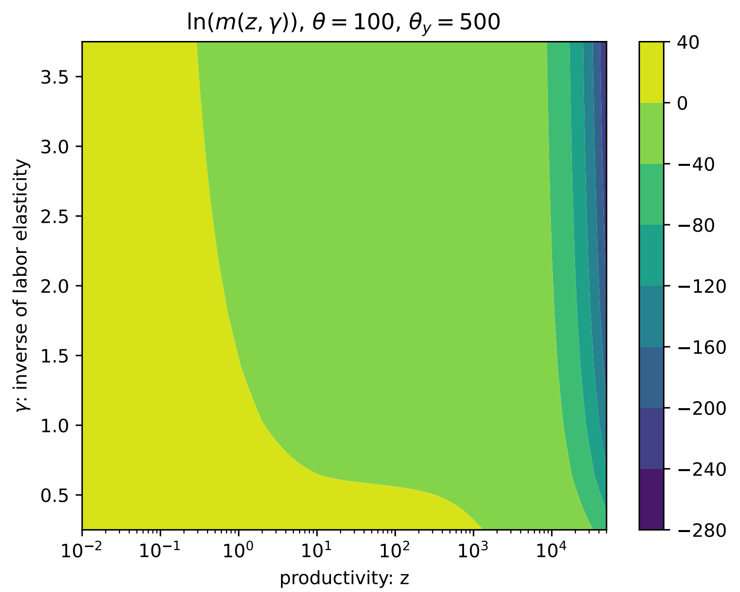
- z and γ independent under the benchmark distribution
- $\log z$ has exponentially modified Gaussian (EGM) distribution
- $\gamma \sim U\,[0.25, 3.75]$, implying $E\,[\gamma] = 2$

Other parameters

- $B = 0$


Marginal density: π(z)
10^−14
10^−37
10^−60
10^−83
10^−106
10^−129
10^−152
10^−175
rational
θ = 100
θ = 100, θ^y = 0
θ = 100, θ^y = 10
θ = 100, θ^y = 100
θ = 100, θ^y = 500
θ = 100, θ^y = 1000
θ = 100, θ^y = 5000
θ = 100, θ^y = 10000
θ = 100, θ^y = 50000
10^−2
10^−1
10^0
10^1
10^2
10^3
10^4
10^5
productivity: z


Marginal density: π(γ)
0.1004
0.1003
0.1002
0.1001
0.1000
0.0999
0.0998
0.5
1.0
1.5
2.0
2.5
3.0
3.5
inverse of labor supply elas: γ
rational
θ = 100
θ = 100, θ^y = 0
θ = 100, θ^y = 10
θ = 100, θ^y = 100
θ = 100, θ^y = 500
θ = 100, θ^y = 1000
θ = 100, θ^y = 5000
θ = 100, θ^y = 10000
θ = 100, θ^y = 50000


ln(m(z, γ)), θ = 100, θ_y = 500
γ: inverse of labor elasticity
3.5
3.0
2.5
2.0
1.5
1.0
0.5
10^-2
10^-1
10^0
10^1
10^2
10^3
10^4
productivity: z
40
0
−40
−80
−120
−160
−200
−240
−280

Reducing model misspecification concerns (higher θ and θ^y) increases optimal tax progressivity τ_2.

$\theta \setminus \theta^y$	0	10	100	500	1000	5000	10000	50000
1	0.200		0.342	0.353	0.358	0.365	0.367	0.368
10	0.350		0.393	0.398	0.398	0.400	0.400	0.401
100	0.399	0.400	0.403	0.406	0.407	0.407	0.408	0.409
1000	0.408	0.408	0.409	0.409	0.409	0.409	0.409	0.409

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
- survey evidence: detection error probabilities measure distinguishablity in finite samples

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

Multidimensional private type space.

- heterogeneity in productivity and labor supply elasticity increases misspecification concerns

Other applications with substantial uncertainty about type distribution.

- wealth taxation

Bibliography

Literature I

Anderson, Evan W., Lars Peter Hansen, and Thomas J. Sargent (2003) "A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection," *Journal of the European Economic Association*, 1 (1), 68–123.

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.

——— (2010) "Fragile Beliefs and the Price of Uncertainty," *Quantitative Economics*, 1 (1), 129–162.

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.

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.

Additional slides

First-best allocation

If government observed types z and could choose $T(z)$, it would achieve first-best.

- households' first-order conditions are not distorted by taxes

$$U_n(c,n) = -zU_c(c,n)$$

- taxes and transfers $T(z)$ equalize marginal utility to marginal social value of public funds μ

$$\frac{d}{dT(z)}\mathcal{U}(z;T(z)) = \mu$$

  where μ is the Lagrange multiplier on government budget constraint

- budgetary and utilitarian implications of misspecification concerns increase μ

If the tax function can only be conditioned on y, taxation becomes distortionary (Ramsey (1927)).

Calibration of misspecification concerns

Degree of misspecification concerns is controlled by penalty parameter θ.

- a specific value of θ does not affect the limiting tax at $y \to \infty$ but has quantitative implications across the type distribution

We utilize the concept of detection error probability from Anderson et al. (2003).

- measures how distinguishable are $\widetilde{f}(z)$ and $f(z)$ using a given finite sample of data of size I

Detection error probability

Consider a sample $\{z_i^B\}_{i=1}^I$ drawn from the benchmark distribution $f(z)$.

- evaluate probability that $\{z_i^B\}_{i=1}^I$ is assigned a higher likelihood under the worst-case $\tilde{f}(z;\theta)$

$$P\left(\sum_{i=1}^I \log \tilde{f}\left(z_i^B;\theta\right) > \sum_{i=1}^I \log f\left(z_i^B\right)\right) = P\left(\sum_{i=1}^I \log m\left(z_i^B;\theta\right) > 0\right)$$

Repeat reciprocally with a sample $\{z_i^A\}_{i=1}^I$ drawn from the alternative worst-case distribution $\tilde{f}(z;\theta)$.

The detection error probability is defined as

$$d(\theta, I) = \frac{1}{2}\left(P\left(\sum_{i=1}^I \log m\left(z_i^B;\theta\right) > 0\right) + P\left(\sum_{i=1}^I \log m\left(z_i^A;\theta\right) < 0\right)\right).$$

- chance that the likelihood ratio leads to the erroneous conclusion about which of the two distributions generated the random sample
- construction implies $0 \leq d(\theta, I) \leq \frac{1}{2}$.

Detection error probability

Imagine we have an available draw of size I.

- Survey of Consumer Finances $I \approx 3,500$.

Agree on a plausible level of the detection error probability $\bar{d}$.

- Hansen and Sargent (2010) argue that $\bar{d} = 0.2$ is conservative, we target $\bar{d} = 0.1$

Infer $\theta\left(\bar{d}, I\right)$ from the implicit equation

$$d\left(\theta\left(\bar{d}, I\right), I\right) = \bar{d}.$$

- a more informative survey with a larger I leads to a higher value $\theta\left(\bar{d}, I\right)$ for a given fixed $\bar{d}$
- larger sample sizes achieve the same detection error probabilities for distributions that are statistically more alike

Detection error probabilities

I	100	500	1000	3500	10000
$\theta = 100$	0.462	0.452	0.437	0.377	0.304
$\theta = 20$	0.432	0.298	0.247	0.090	0.010
$\theta = 10$	0.348	0.172	0.092	0.007	0.000
$\theta = 1$	0.001	0.000	0.000	0.000	0.000

Detection error probabilities for models with alternative choices of misspecification concerns parameterized by θ, and for alternative sizes of the random samples I.

Back to DEPs