Statistical Methods for Distributional Analysis

Frank Cowell[1] Emmanuel Flachaire[2]

[1]STICERD
London School of Economics

[2]GREQAM, Marseille.

Winter School, Canazei. January 2014

Outline

Introduction

Introduction

- Prepared for the *Handbook on Income Distribution*
  - volume 2
  - edited by Atkinson and Bourguignon

Introduction

- Prepared for the *Handbook on Income Distribution*
  - volume 2
  - edited by Atkinson and Bourguignon
- Not really new work
  - but perhaps some fresh insights
  - survey of theory, methods underlying good practice
  - guide to the tools available to the practitioner in this field

Introduction

- Prepared for the *Handbook on Income Distribution*
  - volume 2
  - edited by Atkinson and Bourguignon
- Not really new work
  - but perhaps some fresh insights
  - survey of theory, methods underlying good practice
  - guide to the tools available to the practitioner in this field
- Not just standard inference
  - how to model distributions
  - how to handle data problems

Notation

Notation

- *Income y*. (earnings, wealth, consumption...) Belongs to a set $\mathbb{Y} = [\underline{y}, \bar{y}) \subseteq \mathbb{R}$.

Notation

- *Income y*. (earnings, wealth, consumption...) Belongs to a set $\mathbb{Y}=[\underline{y},\bar{y})\subseteq\mathbb{R}$.
- *Population proportion* $q\in\mathbb{Q}:=[0,1]$.

Notation

- *Income y*. (earnings, wealth, consumption...) Belongs to a set $\mathbb{Y} = [\underline{y}, \bar{y}) \subseteq \mathbb{R}$.
- *Population proportion* $q \in \mathbb{Q} := [0,1]$.
- *Distribution F*. Set of all distribution functions will be denoted $\mathbb{F}$.

Notation

- *Income* y. (earnings, wealth, consumption...) Belongs to a set $\mathbb{Y} = [\underline{y}, \bar{y}) \subseteq \mathbb{R}$.
- *Population proportion* $q \in \mathbb{Q} := [0,1]$.
- *Distribution* F. Set of all distribution functions will be denoted $\mathbb{F}$.
- *Indicator function* $\iota(\cdot)$. For logical condition D:

$$\iota(D) = \begin{cases} 1 & \text{if } D \text{ is true} \\ 0 & \text{if } D \text{ is not true} \end{cases}$$

Data types

Data types

- Complete enumeration

Data types

- Complete enumeration
- Sample: Administrative data
  - summaries of income distributions in grouped form

Data types

- Complete enumeration
- Sample: Administrative data
  - summaries of income distributions in grouped form
  - official micro-data similar size to sample surveys

Data types

- Complete enumeration
- Sample: Administrative data
  - summaries of income distributions in grouped form
  - official micro-data similar size to sample surveys
  - very large data sets: complete collections of admin data

Data types

- Complete enumeration
- Sample: Administrative data
  - summaries of income distributions in grouped form
  - official micro-data similar size to sample surveys
  - very large data sets: complete collections of admin data
    - only what is legally permissible and convenient

Data types

- Complete enumeration
- Sample: Administrative data
  - summaries of income distributions in grouped form
  - official micro-data similar size to sample surveys
  - very large data sets: complete collections of admin data
    - only what is legally permissible and convenient
    - design may not be ideal for economist

Data types

- Complete enumeration
- Sample: Administrative data
  - summaries of income distributions in grouped form
  - official micro-data similar size to sample surveys
  - very large data sets: complete collections of admin data
    - only what is legally permissible and convenient
    - design may not be ideal for economist
- Survey data
  - usually purpose-built

Data types

- Complete enumeration
- Sample: Administrative data
    - summaries of income distributions in grouped form
    - official micro-data similar size to sample surveys
    - very large data sets: complete collections of admin data
        - only what is legally permissible and convenient
        - design may not be ideal for economist
- Survey data
    - usually purpose-built
    - smaller size and worse response rate than administrative-data

Data types

- Complete enumeration
- Sample: Administrative data
  - summaries of income distributions in grouped form
  - official micro-data similar size to sample surveys
  - very large data sets: complete collections of admin data
    - only what is legally permissible and convenient
    - design may not be ideal for economist
- Survey data
  - usually purpose-built
  - smaller size and worse response rate than administrative-data
  - may exclude some sections of the population.

Data design

Data design

- Simple design

Data design

- Simple design
  - sample designed so each member of the population has equal probability of being included in sample

Data design

- Simple design
  - sample designed so each member of the population has equal probability of being included in sample
  - ideal case that enables one to focus on the central issues of statistical inference

Data design

- Simple design
  - sample designed so each member of the population has equal probability of being included in sample
  - ideal case that enables one to focus on the central issues of statistical inference
  - but sampling frame could be out of date or exclude part of the population

Data design

- Simple design
  - sample designed so each member of the population has equal probability of being included in sample
  - ideal case that enables one to focus on the central issues of statistical inference
  - but sampling frame could be out of date or exclude part of the population
- Complex design

Data design

- Simple design
  - sample designed so each member of the population has equal probability of being included in sample
  - ideal case that enables one to focus on the central issues of statistical inference
  - but sampling frame could be out of date or exclude part of the population
- Complex design
  - *Clustering* observations by geographical location may reduce the costs of running the survey

Data design

- Simple design
  - sample designed so each member of the population has equal probability of being included in sample
  - ideal case that enables one to focus on the central issues of statistical inference
  - but sampling frame could be out of date or exclude part of the population
- Complex design
  - *Clustering* observations by geographical location may reduce the costs of running the survey
  - *Stratification*: oversampling certain categories to ensure that adequate representation of certain types

Data Problems

Data Problems

- *Measurement error*
  - similar to measurement error in other contexts

Data Problems

- *Measurement error*
  - similar to measurement error in other contexts
  - observed income = true income adjusted by error term

Data Problems

- *Measurement error*
  - similar to measurement error in other contexts
  - observed income = true income adjusted by error term
  - model resembles problem of factor-source decomposition

Data Problems

- *Measurement error*
  - similar to measurement error in other contexts
  - observed income = true income adjusted by error term
  - model resembles problem of factor-source decomposition
- *Data contamination*

Data Problems

- *Measurement error*
  - similar to measurement error in other contexts
  - observed income = true income adjusted by error term
  - model resembles problem of factor-source decomposition
- *Data contamination*
  - mixture of true distribution and contamination distribution

Data Problems

- *Measurement error*
  - similar to measurement error in other contexts
  - observed income = true income adjusted by error term
  - model resembles problem of factor-source decomposition
- *Data contamination*
  - mixture of true distribution and contamination distribution
  - model resembles problem of subgroup decomposition

Data Problems

- *Measurement error*
  - similar to measurement error in other contexts
  - observed income = true income adjusted by error term
  - model resembles problem of factor-source decomposition
- *Data contamination*
  - mixture of true distribution and contamination distribution
  - model resembles problem of subgroup decomposition
- *Incomplete data*

Data Problems

- *Measurement error*
  - similar to measurement error in other contexts
  - observed income = true income adjusted by error term
  - model resembles problem of factor-source decomposition
- *Data contamination*
  - mixture of true distribution and contamination distribution
  - model resembles problem of subgroup decomposition
- *Incomplete data*
  1. Subset of $\mathbb{Y}$ is specified: income-boundaries $(\underline{z}, \bar{z})$

Data Problems

- *Measurement error*
  - similar to measurement error in other contexts
  - observed income = true income adjusted by error term
  - model resembles problem of factor-source decomposition
- *Data contamination*
  - mixture of true distribution and contamination distribution
  - model resembles problem of subgroup decomposition
- *Incomplete data*
  1. Subset of $\mathbb{Y}$ is specified: income-boundaries $(\underline{z}, \bar{z})$
  2. Subset of $\mathbb{Q}$ is specified: proportions $\left(\underline{\beta}, \bar{\beta}\right)$ in tails

Data Problems

- *Measurement error*
  - similar to measurement error in other contexts
  - observed income = true income adjusted by error term
  - model resembles problem of factor-source decomposition
- *Data contamination*
  - mixture of true distribution and contamination distribution
  - model resembles problem of subgroup decomposition
- *Incomplete data*
  1. Subset of $\mathbb{Y}$ is specified: income-boundaries $(\underline{z}, \bar{z})$
  2. Subset of $\mathbb{Q}$ is specified: proportions $\left(\underline{\beta}, \bar{\beta}\right)$ in tails

Incomplete Data types

Incomplete Data types

	Information re Excluded Sample		
	None	*Sample proportion*	*Multiple statistics*
limits $(\underline{z}, \bar{z})$ *fixed;* $\left(\underline{\beta}, \bar{\beta}\right)$ *unknown*	**A**	**B**	**C**
proportions $\left(\underline{\beta}, \bar{\beta}\right)$ *fixed;* $(\underline{z}, \bar{z})$ *unknown*	**D**	**(E)**	**(F)**

Incomplete Data types

	Information re Excluded Sample		
	None	*Sample proportion*	*Multiple statistics*
limits $(\underline{z}, \bar{z})$ *fixed;* $(\underline{\beta}, \bar{\beta})$ *unknown*	**A**	**B**	**C**
proportions $(\underline{\beta}, \bar{\beta})$ *fixed;* $(\underline{z}, \bar{z})$ *unknown*	**D**	**(E)**	**(F)**

- A: standard form of truncation

Incomplete Data types

	Information re Excluded Sample		
	None	*Sample proportion*	*Multiple statistics*
limits $(\underline{z}, \bar{z})$ *fixed;* $\left(\underline{\beta}, \bar{\beta}\right)$ *unknown*	**A**	**B**	**C**
proportions $\left(\underline{\beta}, \bar{\beta}\right)$ *fixed;* $(\underline{z}, \bar{z})$ *unknown*	**D**	**(E)**	**(F)**

- A: standard form of truncation
- B: "censoring". Point masses at $(\underline{z}, \bar{z})$ estimate the population-share of the excluded part.

Incomplete Data types

	Information re Excluded Sample		
	None	*Sample proportion*	*Multiple statistics*
limits $(\underline{z}, \bar{z})$ *fixed;* $\left(\underline{\beta}, \bar{\beta}\right)$ *unknown*	**A**	**B**	**C**
proportions $\left(\underline{\beta}, \bar{\beta}\right)$ *fixed;* $(\underline{z}, \bar{z})$ *unknown*	**D**	**(E)**	**(F)**

- A: standard form of truncation
- B: "censoring". Point masses at $(\underline{z}, \bar{z})$ estimate the population-share of the excluded part.
- C: Extension of estimation problem with grouped data

Incomplete Data types

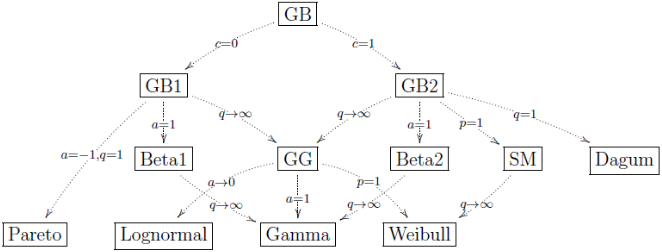
	Information re Excluded Sample		
	None	*Sample proportion*	*Multiple statistics*
limits $(\underline{z}, \bar{z})$ *fixed;* $\left(\underline{\beta}, \bar{\beta}\right)$ *unknown*	**A**	**B**	**C**
proportions $\left(\underline{\beta}, \bar{\beta}\right)$ *fixed;* $(\underline{z}, \bar{z})$ *unknown*	**D**	**(E)**	**(F)**

- A: standard form of truncation
- B: "censoring". Point masses at $(\underline{z}, \bar{z})$ estimate the population-share of the excluded part.
- C: Extension of estimation problem with grouped data
- D: Trimming

Outline

Parametric density estimation

Most of the standard parametric income distributions are special cases of the Generalized Beta distribution:



Goodness of Fit

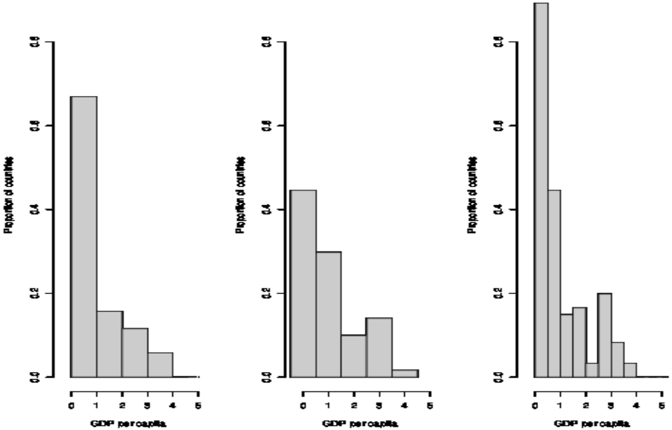
- Density based statistic: Pearson chi-squared (χ^2)
- EDF based statistic: $\begin{cases} \text{Kolmogorov-Smirnov} \\ \text{Anderson-Darling} \\ \text{Cramér-von-Mises} \end{cases}$
- Cowell et al. (2011) developed a GoF test based on axiomatic discussion, with social-welfare foundations:

$$G_\xi = \frac{1}{\xi^2 - \xi} \sum_{i=1}^{n} \left[\left[\frac{u_i}{\mu_u} \right]^{\xi} \left[\frac{2i}{n+1} \right]^{1-\xi} - 1 \right],$$

where $u_i = F(y_{(i)}; \hat{\theta})$ and $y_{(i)}$ is the ith smallest observat.
- The pearson χ^2 statistic has poor finite sample properties

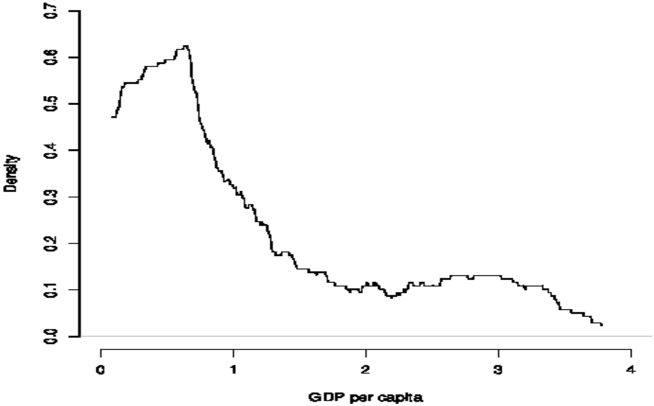
Outline

Histograms



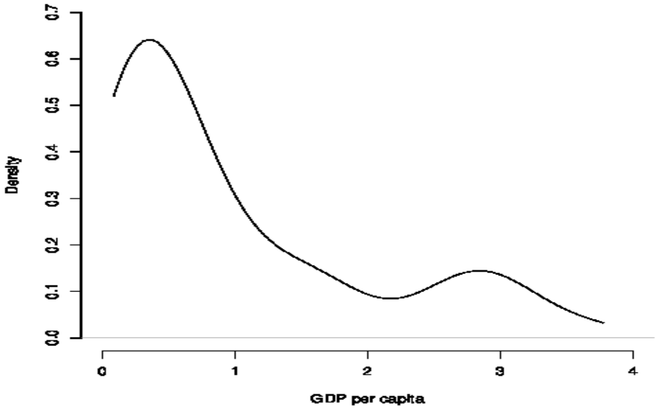
Histogram's sensitivity to the position and the number of bins

Naive estimator



Naive estimator of GDP per capita

Kernel estimator



Kernel estimator of GDP per capita

Bandwidth selection

The kernel density estimator

$$\hat{f}(y) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{y - y_i}{h}\right)$$

is not really affected by the choice of the kernel $K()$, but it is sensitive to the choice of the bandwidth h

Bandwidth selection:

- Silverman's rule-of-thumb, $\hat{h}_{opt} = 0.9 \min\left(\hat{\sigma}\,; \frac{\hat{q}_3 - \hat{q}_1}{1.349}\right) n^{-\frac{1}{5}}$.
- Plug-in method
- Cross-validation

Adaptive kernel

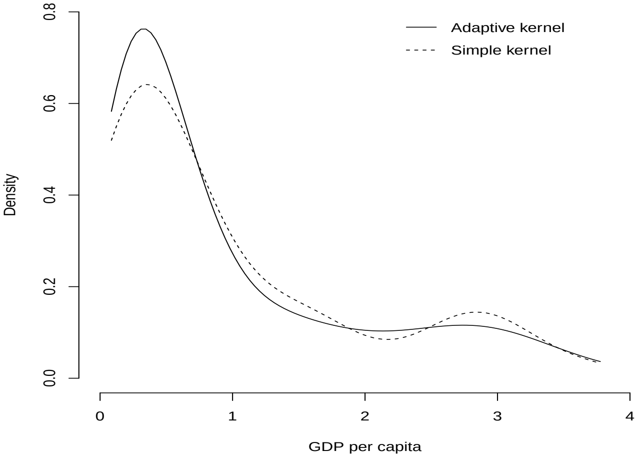
- When the concentration of the data is markedly heterogeneous, a fixed bandwidth may be quite restrictive.

- The adaptive kernel estimator is defined as follows:

$$\hat{f}(y) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h\lambda_i}K\left(\frac{y-y_i}{h\lambda_i}\right),$$

where λ_i is a parameter that varies with the local concentration of the data, $\lambda_i = [g/\tilde{f}(y_i)]^{\alpha}$.

Adaptive kernel



Outline

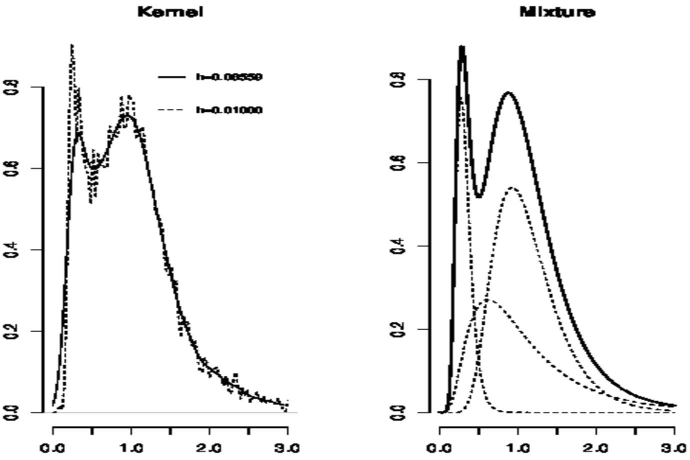
Finite-mixture models

- Under regularity conditions, any distribution can be consistently estimated by a mixture of Normal distributions
- Estimate any income distrib. with a mixture of lognormals :

$$f(\log y; \Theta) = \sum_{k=1}^{K} \pi_k \Phi(y_k; \mu_k, \sigma_k)$$

- Interpretation: a (heterogeneous) population can be decomposed into several distinct (homogeneous) groups
- Brings out the link between parametric and nonparametric estimator ($K = 1$ and $K = n$, $\pi_k = 1/n$)

Finite-mixture models



Income distribution in the United Kingdom in 1973

Finite-mixture models with covariates

- Covariates can be introduced in probabilities to characterized group profiles

$$f(\log y|z;\Theta) = \sum_{k=1}^{K} \pi_k(z_k;\alpha_k)\Phi(y_k;\mu_k,\sigma_k)$$

- Covariates can be introduced into the modeling of the densities in each of the groups, leading us to consider mixture of regression models

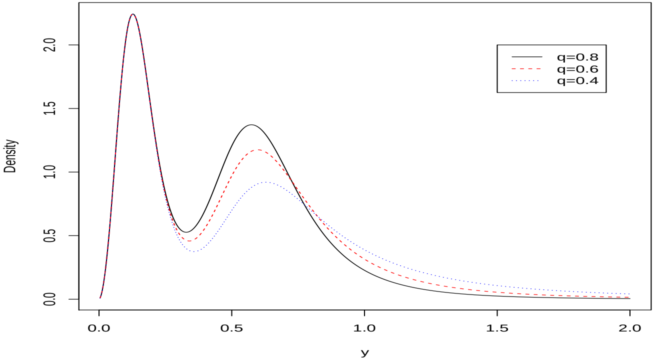
$$f(\log y|x;\Theta) = \sum_{k=1}^{K} \pi_k\Phi(y_k|x_k;\mu_k,\sigma_k)$$

- Covariates can be introduced in both components

Outline

Finite sample properties

- Quality of the fit: MIAE $= E\left(\int_0^\infty \left|\hat{f}(y) - f(y)\right| \mathrm{d}y\right)$.
- Data are drawn from lognormals, Singh-Maddala and mixtures of two SM distributions.



Finite sample properties

	Standard kernel			Adaptive kernel			Mixture
	Silv.	CV	Plug-in	Silv.	CV	Plug-in	lognormal
Lognormal							
$\sigma = 0.5$	0.1044	0.1094	0.1033	0.0982	0.1098	0.1028	0.0407
$\sigma = 0.75$	0.1326	0.1326	0.1252	0.1098	0.1283	0.1179	0.0407
$\sigma = 1$	0.1643	0.1716	0.1522	0.1262	0.1609	0.1362	0.0407
Singh-Maddala							
$q = 1.7$	0.0942	0.1009	0.0951	0.0915	0.0994	0.0934	0.0840
$q = 1.2$	0.1039	0.1100	0.1048	0.0947	0.1050	0.0994	0.0920
$q = 0.7$	0.1346	0.1482	0.1326	0.1049	0.1349	0.1175	0.0873
Mixture of two Singh-Maddala							
$q = 0.8$	0.2080	0.1390	0.1328	0.1577	0.1356	0.1224	0.1367
$q = 0.6$	0.2458	0.1528	0.1463	0.1896	0.1457	0.1293	0.1464
$q = 0.4$	0.2885	0.1953	0.1733	0.2234	0.1812	0.1450	0.1366

Two key functionals

Two key functionals

- Used repeatedly in distributional analysis

Two key functionals

- Used repeatedly in distributional analysis
- Quantile functional
  - $Q : \mathbb{F} \times \mathbb{Q} \to \mathbb{Y}$ given by $Q(F;q) := \inf\{y|F(y) \geq q\}$

Two key functionals

- Used repeatedly in distributional analysis
- Quantile functional
  - $Q : \mathbb{F} \times \mathbb{Q} \to \mathbb{Y}$ given by $Q(F;q) := \inf\{y | F(y) \geq q\}$
  - $y_q := Q(F;q)$

Two key functionals

- Used repeatedly in distributional analysis
- Quantile functional
  - $Q : \mathbb{F} \times \mathbb{Q} \to \mathbb{Y}$ given by $Q(F;q) := \inf\{y | F(y) \geq q\}$
  - $y_q := Q(F;q)$
  - Examples:
    - $q = 0.5$ gives $Q(F;0.5)$, median of distribution F
    - bottom decile: $Q(F;0.1)$
    - upper quartile: b $Q(F;0.75)$,

Two key functionals

- Used repeatedly in distributional analysis
- Quantile functional
  - $Q : \mathbb{F} \times \mathbb{Q} \to \mathbb{Y}$ given by $Q(F;q) := \inf\{y | F(y) \geq q\}$
  - $y_q := Q(F;q)$
  - Examples:
    - $q = 0.5$ gives $Q(F;0.5)$, median of distribution F
    - bottom decile: $Q(F;0.1)$
    - upper quartile: b $Q(F;0.75)$,
- Cumulation functional
  - $C : \mathbb{F} \times \mathbb{Q} \to \mathbb{Y}$ given by $C(F;q) := \int_{\underline{y}}^{y_q} y \, \mathrm{d}F(y)$

Two key functionals

- Used repeatedly in distributional analysis
- Quantile functional
  - $Q : \mathbb{F} \times \mathbb{Q} \to \mathbb{Y}$ given by $Q(F;q) := \inf\{y | F(y) \geq q\}$
  - $y_q := Q(F;q)$
  - Examples:
    - $q = 0.5$ gives $Q(F;0.5)$, median of distribution F
    - bottom decile: $Q(F;0.1)$
    - upper quartile: b $Q(F;0.75)$,
- Cumulation functional
  - $C : \mathbb{F} \times \mathbb{Q} \to \mathbb{Y}$ given by $C(F;q) := \int_{\underline{y}}^{y_q} y \, \mathrm{d}F(y)$
  - $c_q := C(F;q)$

Two key functionals

- Used repeatedly in distributional analysis
- Quantile functional
  - $Q : \mathbb{F} \times \mathbb{Q} \to \mathbb{Y}$ given by $Q(F;q) := \inf\{y | F(y) \geq q\}$
  - $y_q := Q(F;q)$
  - Examples:
    - $q = 0.5$ gives $Q(F;0.5)$, median of distribution F
    - bottom decile: $Q(F;0.1)$
    - upper quartile: b $Q(F;0.75)$,
- Cumulation functional
  - $C : \mathbb{F} \times \mathbb{Q} \to \mathbb{Y}$ given by $C(F;q) := \int_{\underline{y}}^{y_q} y \mathrm{d}F(y)$
  - $c_q := C(F;q)$
  - Examples:
    - $c_1 = C(F;1) = \mu(F)$, mean of distribution F
    - $\frac{c_q}{c_1} = \frac{C(F;q)}{C(F;1)}$, income share of bottom $100q$ percent

Welfare functionals

Welfare functionals

- Fundamental assessment tools of distributional analysis
  - Inequality and poverty indices as special cases

Welfare functionals

- Fundamental assessment tools of distributional analysis
  - Inequality and poverty indices as special cases
- Simplest class: $W_{\mathrm{AD}}(F) := \int \phi(y)\, \mathrm{d}F(y)$ (up to transformation involving μ)

Welfare functionals

- Fundamental assessment tools of distributional analysis
  - Inequality and poverty indices as special cases
- Simplest class: $W_{\mathrm{AD}}(F) := \int \phi(y)\, \mathrm{d}F(y)$ (up to transformation involving μ)
  - for grouped data $\sum_{i=1}^{m} f_i \phi(y_i)$

Welfare functionals

- Fundamental assessment tools of distributional analysis
  - Inequality and poverty indices as special cases
- Simplest class: $W_{\text{AD}}(F) := \int \phi(y)\, \mathrm{d}F(y)$ (up to transformation involving μ)
  - for grouped data $\sum_{i=1}^{m} f_i \phi(y_i)$
  - GE measures, Atkinson

Welfare functionals

- Fundamental assessment tools of distributional analysis
  - Inequality and poverty indices as special cases
- Simplest class:$W_{\mathrm{AD}}(F) := \int \phi(y)\, \mathrm{d}F(y)$ (up to transformation involving μ)
  - for grouped data $\sum_{i=1}^{m} f_i \phi(y_i)$
  - GE measures, Atkinson
- Slightly broader class:$W_{\mathrm{QAD}}(F) := \int \varphi(y, \mu(F))\, \mathrm{d}F(y)$
  - includes $W_{\mathrm{AD}}(F)$

Welfare functionals

- Fundamental assessment tools of distributional analysis
  - Inequality and poverty indices as special cases
- Simplest class:$W_{\text{AD}}(F) := \int \phi(y)\,\mathrm{d}F(y)$ (up to transformation involving μ)
  - for grouped data $\sum_{i=1}^{m} f_i \phi(y_i)$
  - GE measures, Atkinson
- Slightly broader class:$W_{\text{QAD}}(F) := \int \varphi(y, \mu(F))\,\mathrm{d}F(y)$
  - includes $W_{\text{AD}}(F)$
  - RMD, Pietra ratio

Welfare functionals

- Fundamental assessment tools of distributional analysis
  - Inequality and poverty indices as special cases
- Simplest class:$W_{\mathrm{AD}}(F) := \int \phi(y)\,\mathrm{d}F(y)$ (up to transformation involving μ)
  - for grouped data $\sum_{i=1}^{m} f_i \phi(y_i)$
  - GE measures, Atkinson
- Slightly broader class:$W_{\mathrm{QAD}}(F) := \int \varphi(y, \mu(F))\,\mathrm{d}F(y)$
  - includes $W_{\mathrm{AD}}(F)$
  - RMD, Pietra ratio
- Rank-dependence: $W_{\mathrm{RD}}(F) := \int \psi(y, \mu(F), F(y))\,\mathrm{d}F(y)$,

Welfare functionals

- Fundamental assessment tools of distributional analysis
    - Inequality and poverty indices as special cases
- Simplest class: $W_{\mathrm{AD}}(F) := \int \phi(y)\, \mathrm{d}F(y)$ (up to transformation involving μ)
    - for grouped data $\sum_{i=1}^{m} f_i \phi(y_i)$
    - GE measures, Atkinson
- Slightly broader class: $W_{\mathrm{QAD}}(F) := \int \varphi(y, \mu(F))\, \mathrm{d}F(y)$
    - includes $W_{\mathrm{AD}}(F)$
    - RMD, Pietra ratio
- Rank-dependence: $W_{\mathrm{RD}}(F) := \int \psi(y, \mu(F), F(y))\, \mathrm{d}F(y)$,
    - Gini
    - generalised Gini

Welfare functionals

- Fundamental assessment tools of distributional analysis
  - Inequality and poverty indices as special cases
- Simplest class: $W_{\mathrm{AD}}(F) := \int \phi(y)\, \mathrm{d}F(y)$ (up to transformation involving μ)
  - for grouped data $\sum_{i=1}^{m} f_i \phi(y_i)$
  - GE measures, Atkinson
- Slightly broader class: $W_{\mathrm{QAD}}(F) := \int \varphi(y, \mu(F))\, \mathrm{d}F(y)$
  - includes $W_{\mathrm{AD}}(F)$
  - RMD, Pietra ratio
- Rank-dependence: $W_{\mathrm{RD}}(F) := \int \psi(y, \mu(F), F(y))\, \mathrm{d}F(y)$,
  - Gini
  - generalised Gini

Outline

The Influence Function

- Introducing a fundamental tool

The Influence Function

- Introducing a fundamental tool
- Mixture distributions

The Influence Function

- Introducing a fundamental tool
- Mixture distributions
  - point mass at z: $H^{(z)}(y) = \iota(y \geq z)$

The Influence Function

- Introducing a fundamental tool
- Mixture distributions
  - point mass at z: $H^{(z)}(y) = \iota(y \geq z)$
  - the mixture: $G = [1-\delta]F + \delta H^{(z)}$

The Influence Function

- Introducing a fundamental tool
- Mixture distributions
  - point mass at z: $H^{(z)}(y) = \iota(y \geq z)$
  - the mixture: $G = [1-\delta]F + \delta H^{(z)}$
  - δ : importance of point mass in mixture

The Influence Function

- Introducing a fundamental tool
- Mixture distributions
  - point mass at z: $H^{(z)}(y) = \iota(y \geq z)$
  - the mixture: $G = [1 - \delta]F + \delta H^{(z)}$
  - δ : importance of point mass in mixture
- How "important" is point mass at z?

The Influence Function

- Introducing a fundamental tool
- Mixture distributions
  - point mass at z: $H^{(z)}(y) = \iota(y \geq z)$
  - the mixture: $G = [1-\delta]F + \delta H^{(z)}$
  - δ : importance of point mass in mixture
- How "important" is point mass at z?
  - define this with reference to a given statistic T

The Influence Function

- Introducing a fundamental tool
- Mixture distributions
  - point mass at z: $H^{(z)}(y) = \iota(y \geq z)$
  - the mixture: $G = [1 - \delta]F + \delta H^{(z)}$
  - δ : importance of point mass in mixture
- How "important" is point mass at z?
  - define this with reference to a given statistic T
  - Influence Function: $IF(z; T, F) := \lim_{\delta \downarrow 0} \left[\frac{T(G) - T(F)}{\delta} \right]$

The Influence Function

- Introducing a fundamental tool
- Mixture distributions
  - point mass at z: $H^{(z)}(y) = \iota(y \geq z)$
  - the mixture: $G = [1-\delta]F + \delta H^{(z)}$
  - δ : importance of point mass in mixture
- How "important" is point mass at z?
  - define this with reference to a given statistic T
  - Influence Function: $IF(z;T,F) := \lim_{\delta\downarrow 0}\left[\frac{T(G)-T(F)}{\delta}\right]$
  - if differentiable: $IF(z;T,F) := \left.\frac{\partial}{\partial\delta}T(G)\right|_{\delta\to 0}$

IF and AV

- A simple decomposition:
  - $T(G) = T(F) + \int IF(y; T, F)\, \mathrm{d}(G - F)(y) + \text{remainder}$

IF and AV

- A simple decomposition:
  - $T(G) = T(F) + \int IF(y;T,F)\,\mathrm{d}(G-F)(y) + \text{remainder}$
  - $T(F^{(n)}) \approx T(F) + \frac{1}{n}\sum_{i=1}^{n} IF(y_i;T,F) + \text{remainder}$
- **Lemma**
  - $\sqrt{n}\left(T(F^{(n)}) - T(F)\right)$ is asymptotically normal
  - asymptotic covariance matrix $\int IF(y;T,F)IF^{\top}(y;T,F)\,\mathrm{d}F(y)$

IF and AV

- A simple decomposition:
  - $T(G) = T(F) + \int IF(y;T,F)\,\mathrm{d}(G-F)(y) + \text{remainder}$
  - $T(F^{(n)}) \approx T(F) + \frac{1}{n}\sum_{i=1}^{n} IF(y_i;T,F) + \text{remainder}$
- **Lemma**
  - $\sqrt{n}\left(T(F^{(n)}) - T(F)\right)$ is asymptotically normal
  - asymptotic covariance matrix $\int IF(y;T,F)IF^{\top}(y;T,F)\,\mathrm{d}F(y)$
- A short-cut to AV formula:

IF and AV

- A simple decomposition:
  - $T(G) = T(F) + \int IF(y;T,F)\,\mathrm{d}(G-F)(y) + \text{remainder}$
  - $T(F^{(n)}) \approx T(F) + \frac{1}{n}\sum_{i=1}^{n} IF(y_i;T,F) + \text{remainder}$
- **Lemma**
  - $\sqrt{n}\left(T(F^{(n)}) - T(F)\right)$ is asymptotically normal
  - asymptotic covariance matrix $\int IF(y;T,F)IF^{\top}(y;T,F)\,\mathrm{d}F(y)$
- A short-cut to AV formula:
  - can usually find Z such that: $IF(y,T,F) = Z - E(Z)$

IF and AV

- A simple decomposition:
  - $T(G) = T(F) + \int IF(y;T,F)\,\mathrm{d}(G-F)(y) + \text{remainder}$
  - $T(F^{(n)}) \approx T(F) + \frac{1}{n}\sum_{i=1}^{n} IF(y_i;T,F) + \text{remainder}$
- **Lemma**
  - $\sqrt{n}\left(T(F^{(n)}) - T(F)\right)$ is asymptotically normal
  - asymptotic covariance matrix $\int IF(y;T,F) IF^{\top}(y;T,F)\,\mathrm{d}F(y)$
- A short-cut to AV formula:
  - can usually find Z such that: $IF(y,T,F) = Z - E(Z)$
  - therefore: $\int IF(y,T,F)^2 dF(y) = \int (Z - E(Z))^2\,\mathrm{d}F(Z)$

IF and AV

- A simple decomposition:
  - $T(G) = T(F) + \int IF(y;T,F)\,\mathrm{d}(G-F)(y) + \text{remainder}$
  - $T(F^{(n)}) \approx T(F) + \frac{1}{n}\sum_{i=1}^{n} IF(y_i;T,F) + \text{remainder}$
- **Lemma**
  - $\sqrt{n}\left(T(F^{(n)}) - T(F)\right)$ is asymptotically normal
  - asymptotic covariance matrix $\int IF(y;T,F)IF^{\top}(y;T,F)\,\mathrm{d}F(y)$
- A short-cut to AV formula:
  - can usually find Z such that: $IF(y,T,F) = Z - E(Z)$
  - therefore: $\int IF(y,T,F)^2 dF(y) = \int (Z-E(Z))^2\,\mathrm{d}F(Z)$
  - $\widehat{\text{var}}\left(T(F^{(n)})\right) = \frac{1}{n}\widehat{\text{var}}(Z) = \frac{1}{n^2}\sum_{i=1}^{n}(Z_i - \bar{Z})^2$

IF: Application to basics 1

- To apply the IF method:

IF: Application to basics 1

- To apply the IF method:
  - evaluate Q or C for the mixture distribution

IF: Application to basics 1

- To apply the IF method:
  - evaluate Q or C for the mixture distribution
  - differentiate wrt δ

IF: Application to basics 1

- To apply the IF method:
    - evaluate Q or C for the mixture distribution
    - differentiate wrt δ
    - let δ go to zero

IF: Application to basics 1

- To apply the IF method:
  - evaluate Q or C for the mixture distribution
  - differentiate wrt δ
  - let δ go to zero
- Quantile (mixture): $Q(G,q) = Q\left(F, \frac{q-\iota(y_q \geq z)\delta}{1-\delta}\right)$

IF: Application to basics 1

- To apply the IF method:
  - evaluate Q or C for the mixture distribution
  - differentiate wrt δ
  - let δ go to zero
- Quantile (mixture): $Q(G,q) = Q\left(F, \frac{q - \iota(y_q \geq z)\delta}{1-\delta}\right)$
  - $y_q = Q(F,q)$ is qth quantile for the (unmixed) distribution

IF: Application to basics 1

- To apply the IF method:
  - evaluate Q or C for the mixture distribution
  - differentiate wrt δ
  - let δ go to zero
- Quantile (mixture): $Q(G,q) = Q\left(F, \frac{q-\iota(y_q \geq z)\delta}{1-\delta}\right)$
  - $y_q = Q(F,q)$ is qth quantile for the (unmixed) distribution
  - $IF(z;Q(\cdot,q),F) = \frac{q-\iota(Q(F;q)\geq z)}{f(Q(F;q))} = \frac{q-\iota(y_q\geq z)}{f(y_q)}$

IF: Application to basics 1

- To apply the IF method:
  - evaluate Q or C for the mixture distribution
  - differentiate wrt δ
  - let δ go to zero
- Quantile (mixture): $Q(G,q) = Q\left(F, \frac{q-\iota(y_q \geq z)\delta}{1-\delta}\right)$
  - $y_q = Q(F,q)$ is qth quantile for the (unmixed) distribution
  - $IF(z;Q(\cdot,q),F) = \frac{q-\iota(Q(F;q)\geq z)}{f(Q(F;q))} = \frac{q-\iota(y_q \geq z)}{f(y_q)}$
- Cumulation (mixture): $C(G;q) = [1-\delta]\int_{\underline{y}}^{Q(G,q)} y\,\mathrm{d}F(y) + \delta z$

IF: Application to basics 1

- To apply the IF method:
  - evaluate Q or C for the mixture distribution
  - differentiate wrt δ
  - let δ go to zero
- Quantile (mixture): $Q(G,q) = Q\left(F, \frac{q-\iota(y_q \geq z)\delta}{1-\delta}\right)$
  - $y_q = Q(F,q)$ is qth quantile for the (unmixed) distribution
  - $IF(z;Q(\cdot,q),F) = \frac{q-\iota(Q(F;q)\geq z)}{f(Q(F;q))} = \frac{q-\iota(y_q \geq z)}{f(y_q)}$
- Cumulation (mixture): $C(G;q) = [1-\delta]\int_{\underline{y}}^{Q(G,q)} y\,\mathrm{d}F(y) + \delta z$
  - differentiating wrt δ and setting $\delta = 0$ we get

IF: Application to basics 1

- To apply the IF method:
  - evaluate Q or C for the mixture distribution
  - differentiate wrt δ
  - let δ go to zero
- Quantile (mixture): $Q(G,q) = Q\left(F, \frac{q-\iota(y_q \geq z)\delta}{1-\delta}\right)$
  - $y_q = Q(F,q)$ is qth quantile for the (unmixed) distribution
  - $IF(z;Q(\cdot,q),F) = \frac{q-\iota(Q(F;q)\geq z)}{f(Q(F;q))} = \frac{q-\iota(y_q \geq z)}{f(y_q)}$
- Cumulation (mixture): $C(G;q) = [1-\delta]\int_{\underline{y}}^{Q(G,q)} y\,\mathrm{d}F(y) + \delta z$
  - differentiating wrt δ and setting $\delta = 0$ we get
  - $qQ(F,q) - C(F,q) + \iota(q \geq F(z))[z - Q(F,q)]$

IF: Application to basics 1

- To apply the IF method:
  - evaluate Q or C for the mixture distribution
  - differentiate wrt δ
  - let δ go to zero
- Quantile (mixture): $Q(G,q) = Q\left(F, \frac{q-\iota(y_q \geq z)\delta}{1-\delta}\right)$
  - $y_q = Q(F,q)$ is qth quantile for the (unmixed) distribution
  - $IF(z;Q(\cdot,q),F) = \frac{q-\iota(Q(F;q)\geq z)}{f(Q(F;q))} = \frac{q-\iota(y_q \geq z)}{f(y_q)}$
- Cumulation (mixture): $C(G;q) = [1-\delta]\int_{\underline{y}}^{Q(G,q)} y\,\mathrm{d}F(y) + \delta z$
  - differentiating wrt δ and setting $\delta = 0$ we get
  - $qQ(F,q) - C(F,q) + \iota(q \geq F(z))[z - Q(F,q)]$
  - $IF(z;C(\cdot,q),F) = qy_q - c_q + \iota(y_q \geq z)[z - y_q]$

IF: Application to basics 1

- To apply the IF method:
  - evaluate Q or C for the mixture distribution
  - differentiate wrt δ
  - let δ go to zero
- Quantile (mixture): $Q(G,q) = Q\left(F, \frac{q-\iota(y_q \geq z)\delta}{1-\delta}\right)$
  - $y_q = Q(F,q)$ is qth quantile for the (unmixed) distribution
  - $IF(z;Q(\cdot,q),F) = \frac{q-\iota(Q(F;q)\geq z)}{f(Q(F;q))} = \frac{q-\iota(y_q \geq z)}{f(y_q)}$
- Cumulation (mixture): $C(G;q) = [1-\delta]\int_{\underline{y}}^{Q(G,q)} y\,\mathrm{d}F(y) + \delta z$
  - differentiating wrt δ and setting $\delta = 0$ we get
  - $qQ(F,q) - C(F,q) + \iota(q \geq F(z))[z - Q(F,q)]$
  - $IF(z;C(\cdot,q),F) = qy_q - c_q + \iota(y_q \geq z)[z - y_q]$

IF: Application to basics 2

IF: Application to basics 2

- $\widehat{W}_{\text{QAD}} := W_{\text{QAD}}(F^{(n)}) = \frac{1}{n}\sum_{i=1}^{n} \varphi(y_i, \hat{\mu})$

IF: Application to basics 2

- $\widehat{W}_{\text{QAD}} := W_{\text{QAD}}(F^{(n)}) = \frac{1}{n}\sum_{i=1}^{n} \varphi(y_i, \hat{\mu})$
  - sample mean: $\hat{\mu} := \mu\left(F^{(n)}\right) = \frac{1}{n}\sum_{i=1}^{n} y_i$.

IF: Application to basics 2

- $\widehat{W}_{\text{QAD}} := W_{\text{QAD}}(F^{(n)}) = \frac{1}{n}\sum_{i=1}^{n} \varphi(y_i, \hat{\mu})$
  - sample mean: $\hat{\mu} := \mu\left(F^{(n)}\right) = \frac{1}{n}\sum_{i=1}^{n} y_i$.
- Same procedure as before

IF: Application to basics 2

- $\widehat{W}_{\text{QAD}} := W_{\text{QAD}}(F^{(n)}) = \frac{1}{n}\sum_{i=1}^{n} \varphi(y_i, \hat{\mu})$
  - sample mean: $\hat{\mu} := \mu\left(F^{(n)}\right) = \frac{1}{n}\sum_{i=1}^{n} y_i$.
- Same procedure as before
  - evaluate W_{QAD} for the mixture distribution
  - differentiate wrt δ
  - let δ go to zero
- $$IF(z; W_{\text{QAD}}, F) = \varphi(z, \mu(F)) - W_{\text{QAD}}(F) + [z - \mu(F)] \int \varphi_\mu (z, \mu($$
  - where φ_μ denotes the partial derivative

IF: Application to basics 2

- $\widehat{W}_{\text{QAD}} := W_{\text{QAD}}(F^{(n)}) = \frac{1}{n}\sum_{i=1}^{n} \varphi(y_i, \hat{\mu})$
  - sample mean: $\hat{\mu} := \mu\left(F^{(n)}\right) = \frac{1}{n}\sum_{i=1}^{n} y_i$.
- Same procedure as before
  - evaluate W_{QAD} for the mixture distribution
  - differentiate wrt δ
  - let δ go to zero
- $$IF(z; W_{\text{QAD}}, F) = \varphi(z, \mu(F)) - W_{\text{QAD}}(F) + [z - \mu(F)] \int \varphi_\mu (z, \mu($$
  - where φ_μ denotes the partial derivative
  - $IF(y, W_{\text{QAD}}, F) = Z - E(Z)$
  - where $Z = \varphi(y, \mu(F)) + y \int \varphi_\mu (y, \mu(F))\, \mathrm{d}F(y)$

IF: Application to basics 2

- $\widehat{W}_{\text{QAD}} := W_{\text{QAD}}(F^{(n)}) = \frac{1}{n}\sum_{i=1}^{n} \varphi\left(y_i, \hat{\mu}\right)$
  - sample mean: $\hat{\mu} := \mu\left(F^{(n)}\right) = \frac{1}{n}\sum_{i=1}^{n} y_i$.
- Same procedure as before
  - evaluate W_{QAD} for the mixture distribution
  - differentiate wrt δ
  - let δ go to zero
- $$IF(z; W_{\text{QAD}}, F) = \varphi\left(z, \mu(F)\right) - W_{\text{QAD}}(F) + [z - \mu(F)] \int \varphi_\mu\left(z, \mu\right.$$
  - where φ_μ denotes the partial derivative
  - $IF(y, W_{\text{QAD}}, F) = Z - E(Z)$
  - where $Z = \varphi\left(y, \mu(F)\right) + y \int \varphi_\mu\left(y, \mu(F)\right) \mathrm{d}F(y)$
- AV of $\sqrt{n}(\widehat{W}_{\text{QAD}} - W_{\text{QAD}})$ is the variance of Z.

IF: Application to basics 2

- $\widehat{W}_{\text{QAD}} := W_{\text{QAD}}(F^{(n)}) = \frac{1}{n}\sum_{i=1}^{n} \varphi\left(y_i, \hat{\mu}\right)$
  - sample mean: $\hat{\mu} := \mu\left(F^{(n)}\right) = \frac{1}{n}\sum_{i=1}^{n} y_i$.
- Same procedure as before
  - evaluate W_{QAD} for the mixture distribution
  - differentiate wrt δ
  - let δ go to zero
- $$IF(z; W_{\text{QAD}}, F) = \varphi\left(z, \mu(F)\right) - W_{\text{QAD}}(F) + [z - \mu(F)] \int \varphi_{\mu}\left(z, \mu\right.$$
  - where φ_{μ} denotes the partial derivative
  - $IF(y, W_{\text{QAD}}, F) = Z - E(Z)$
  - where $Z = \varphi\left(y, \mu(F)\right) + y \int \varphi_{\mu}\left(y, \mu(F)\right) \mathrm{d}F(y)$
- AV of $\sqrt{n}(\widehat{W}_{\text{QAD}} - W_{\text{QAD}})$ is the variance of Z.
- Provides key to large class of indices used in economics

Outline

Application: GE class

Application: GE class

- $I_{\mathrm{GE}}^{\xi}(F) = \frac{1}{\xi^2 - \xi}\left[\int_{\underline{y}}^{\overline{y}} \left[\frac{y}{\mu(F)}\right]^{\xi} \mathrm{d}F(y) - 1\right]$

Application: GE class

- $I_{\mathrm{GE}}^{\xi}(F) = \frac{1}{\xi^2 - \xi}\left[\int_{\underline{y}}^{\overline{y}} \left[\frac{y}{\mu(F)}\right]^{\xi} \mathrm{d}F(y) - 1\right]$
  - $I_{\mathrm{GE}}^{0}(F) = -\int_{\underline{y}}^{\overline{y}} \log\left(\frac{y}{\mu(F)}\right) \mathrm{d}F(y)$

Application: GE class

- $I_{\mathrm{GE}}^{\xi}(F) = \frac{1}{\xi^2-\xi}\left[\int_{\underline{y}}^{\overline{y}}\left[\frac{y}{\mu(F)}\right]^{\xi}\mathrm{d}F(y) - 1\right]$
  - $I_{\mathrm{GE}}^{0}(F) = -\int_{\underline{y}}^{\overline{y}}\log\left(\frac{y}{\mu(F)}\right)\mathrm{d}F(y)$
  - $I_{\mathrm{GE}}^{1}(F) = \int_{\underline{y}}^{\overline{y}}\frac{y}{\mu(F)}\log\left(\frac{y}{\mu(F)}\right)\mathrm{d}F(y)$

Application: GE class

- $I_{\text{GE}}^{\xi}(F) = \frac{1}{\xi^2-\xi}\left[\int_{\underline{y}}^{\overline{y}}\left[\frac{y}{\mu(F)}\right]^{\xi} \mathrm{d}F(y) - 1\right]$
  - $I_{\text{GE}}^{0}(F) = -\int_{\underline{y}}^{\overline{y}} \log\left(\frac{y}{\mu(F)}\right) \mathrm{d}F(y)$
  - $I_{\text{GE}}^{1}(F) = \int_{\underline{y}}^{\overline{y}} \frac{y}{\mu(F)} \log\left(\frac{y}{\mu(F)}\right) \mathrm{d}F(y)$
- We have $\varphi\left(y, \mu(F)\right) = \frac{1}{\xi^2-\xi}\left[\left[\frac{y}{\mu(F)}\right]^{\xi} - 1\right]$

Application: GE class

- $I_{\text{GE}}^{\xi}(F) = \frac{1}{\xi^2-\xi}\left[\int_{\underline{y}}^{\overline{y}}\left[\frac{y}{\mu(F)}\right]^{\xi}\mathrm{d}F(y) - 1\right]$
  - $I_{\text{GE}}^{0}(F) = -\int_{\underline{y}}^{\overline{y}}\log\left(\frac{y}{\mu(F)}\right)\mathrm{d}F(y)$
  - $I_{\text{GE}}^{1}(F) = \int_{\underline{y}}^{\overline{y}}\frac{y}{\mu(F)}\log\left(\frac{y}{\mu(F)}\right)\mathrm{d}F(y)$
- We have $\varphi\left(y,\mu(F)\right) = \frac{1}{\xi^2-\xi}\left[\left[\frac{y}{\mu(F)}\right]^{\xi} - 1\right]$
  - $\varphi_{\mu}\left(y,\mu(F)\right) = \frac{-\xi}{\xi^2-\xi}\left[\frac{y^{\xi}}{\mu(F)^{\xi+1}}\right] = -\frac{\xi}{\mu}\left(\varphi\left(y,\mu(F)\right) + \frac{1}{\xi^2-\xi}\right)$
- $\widehat{\text{var}}(\hat{I}_{\text{GE}}^{\xi}) = \frac{1}{n^2}\sum_{i=1}^{n}(Z_i-\bar{Z})^2$ where Z_i is

Application: GE class

- $I_{\mathrm{GE}}^{\xi}(F) = \frac{1}{\xi^2-\xi}\left[\int_{\underline{y}}^{\overline{y}}\left[\frac{y}{\mu(F)}\right]^{\xi}\mathrm{d}F(y) - 1\right]$
  - $I_{\mathrm{GE}}^{0}(F) = -\int_{\underline{y}}^{\overline{y}}\log\left(\frac{y}{\mu(F)}\right)\mathrm{d}F(y)$
  - $I_{\mathrm{GE}}^{1}(F) = \int_{\underline{y}}^{\overline{y}}\frac{y}{\mu(F)}\log\left(\frac{y}{\mu(F)}\right)\mathrm{d}F(y)$
- We have $\varphi\left(y,\mu(F)\right) = \frac{1}{\xi^2-\xi}\left[\left[\frac{y}{\mu(F)}\right]^{\xi} - 1\right]$
  - $\varphi_{\mu}\left(y,\mu(F)\right) = \frac{-\xi}{\xi^2-\xi}\left[\frac{y^{\xi}}{\mu(F)^{\xi+1}}\right] = -\frac{\xi}{\mu}\left(\varphi\left(y,\mu(F)\right) + \frac{1}{\xi^2-\xi}\right)$
- $\widehat{\mathrm{var}}(\hat{I}_{\mathrm{GE}}^{\xi}) = \frac{1}{n^2}\sum_{i=1}^{n}(Z_i-\bar{Z})^2$ where Z_i is
  - $(\xi^2-\xi)^{-1}(y_i/\hat{\mu})^{\xi} - \xi(y_i/\hat{\mu})\left[\hat{I}_{\mathrm{GE}}^{\xi} + (\xi^2-\xi)^{-1}\right]$ for $\xi \neq 0, 1$

Application: GE class

- $I_{\text{GE}}^{\xi}(F) = \frac{1}{\xi^2-\xi}\left[\int_{\underline{y}}^{\overline{y}}\left[\frac{y}{\mu(F)}\right]^{\xi} \mathrm{d}F(y) - 1\right]$
  - $I_{\text{GE}}^{0}(F) = -\int_{\underline{y}}^{\overline{y}} \log\left(\frac{y}{\mu(F)}\right) \mathrm{d}F(y)$
  - $I_{\text{GE}}^{1}(F) = \int_{\underline{y}}^{\overline{y}} \frac{y}{\mu(F)} \log\left(\frac{y}{\mu(F)}\right) \mathrm{d}F(y)$
- We have $\varphi\left(y, \mu(F)\right) = \frac{1}{\xi^2-\xi}\left[\left[\frac{y}{\mu(F)}\right]^{\xi} - 1\right]$
  - $\varphi_{\mu}\left(y, \mu(F)\right) = \frac{-\xi}{\xi^2-\xi}\left[\frac{y^{\xi}}{\mu(F)^{\xi+1}}\right] = -\frac{\xi}{\mu}\left(\varphi\left(y, \mu(F)\right) + \frac{1}{\xi^2-\xi}\right)$
- $\widehat{\text{var}}(\hat{I}_{\text{GE}}^{\xi}) = \frac{1}{n^2}\sum_{i=1}^{n}(Z_i - \bar{Z})^2$ where Z_i is
  - $(\xi^2-\xi)^{-1}(y_i/\hat{\mu})^{\xi} - \xi(y_i/\hat{\mu})\left[\hat{I}_{\text{GE}}^{\xi} + (\xi^2-\xi)^{-1}\right]$ for $\xi \neq 0, 1$
  - $(y_i/\hat{\mu}) - \log y_i$

Application: GE class

- $I_{\mathrm{GE}}^{\xi}(F) = \frac{1}{\xi^2-\xi}\left[\int_{\underline{y}}^{\overline{y}}\left[\frac{y}{\mu(F)}\right]^{\xi}\mathrm{d}F(y) - 1\right]$
  - $I_{\mathrm{GE}}^{0}(F) = -\int_{\underline{y}}^{\overline{y}}\log\left(\frac{y}{\mu(F)}\right)\mathrm{d}F(y)$
  - $I_{\mathrm{GE}}^{1}(F) = \int_{\underline{y}}^{\overline{y}}\frac{y}{\mu(F)}\log\left(\frac{y}{\mu(F)}\right)\mathrm{d}F(y)$
- We have $\varphi\left(y,\mu(F)\right) = \frac{1}{\xi^2-\xi}\left[\left[\frac{y}{\mu(F)}\right]^{\xi} - 1\right]$
  - $\varphi_{\mu}\left(y,\mu(F)\right) = \frac{-\xi}{\xi^2-\xi}\left[\frac{y^{\xi}}{\mu(F)^{\xi+1}}\right] = -\frac{\xi}{\mu}\left(\varphi\left(y,\mu(F)\right) + \frac{1}{\xi^2-\xi}\right)$
- $\widehat{\mathrm{var}}(\hat{I}_{\mathrm{GE}}^{\xi}) = \frac{1}{n^2}\sum_{i=1}^{n}(Z_i - \bar{Z})^2$ where Z_i is
  - $(\xi^2-\xi)^{-1}(y_i/\hat{\mu})^{\xi} - \xi(y_i/\hat{\mu})\left[\hat{I}_{\mathrm{GE}}^{\xi} + (\xi^2-\xi)^{-1}\right]$ for $\xi \neq 0, 1$
  - $(y_i/\hat{\mu}) - \log y_i$
  - $(y_i/\hat{\mu})\left[\log(y_i/\hat{\mu}) - \hat{I}_{\mathrm{GE}}^{1} - 1\right]$

Application: GE class

- $I_{\mathrm{GE}}^{\xi}(F) = \frac{1}{\xi^2-\xi}\left[\int_{\underline{y}}^{\overline{y}}\left[\frac{y}{\mu(F)}\right]^{\xi}\mathrm{d}F(y) - 1\right]$
  - $I_{\mathrm{GE}}^{0}(F) = -\int_{\underline{y}}^{\overline{y}}\log\left(\frac{y}{\mu(F)}\right)\mathrm{d}F(y)$
  - $I_{\mathrm{GE}}^{1}(F) = \int_{\underline{y}}^{\overline{y}}\frac{y}{\mu(F)}\log\left(\frac{y}{\mu(F)}\right)\mathrm{d}F(y)$
- We have $\varphi(y,\mu(F)) = \frac{1}{\xi^2-\xi}\left[\left[\frac{y}{\mu(F)}\right]^{\xi} - 1\right]$
  - $\varphi_{\mu}(y,\mu(F)) = \frac{-\xi}{\xi^2-\xi}\left[\frac{y^{\xi}}{\mu(F)^{\xi+1}}\right] = -\frac{\xi}{\mu}\left(\varphi(y,\mu(F)) + \frac{1}{\xi^2-\xi}\right)$
- $\widehat{\mathrm{var}}(\hat{I}_{\mathrm{GE}}^{\xi}) = \frac{1}{n^2}\sum_{i=1}^{n}(Z_i-\bar{Z})^2$ where Z_i is
  - $(\xi^2-\xi)^{-1}(y_i/\hat{\mu})^{\xi} - \xi(y_i/\hat{\mu})\left[\hat{I}_{\mathrm{GE}}^{\xi} + (\xi^2-\xi)^{-1}\right]$ for $\xi \neq 0,1$
  - $(y_i/\hat{\mu}) - \log y_i$
  - $(y_i/\hat{\mu})\left[\log(y_i/\hat{\mu}) - \hat{I}_{\mathrm{GE}}^{1} - 1\right]$

Application: Gini

Application: Gini

- Gini has multiple equivalent forms

Application: Gini

- Gini has multiple equivalent forms
  - From the Lorenz curve
  - $I_{\mathrm{Gini}}(F) = 1 - 2\int_0^1 L(F;q)\,\mathrm{d}q$

Application: Gini

- Gini has multiple equivalent forms
  - From the Lorenz curve
  - $I_{\text{Gini}}(F) = 1 - 2\int_0^1 L(F;q)\,\mathrm{d}q$
  - We can use results on $C(\cdot;q)$

Application: Gini

- Gini has multiple equivalent forms
  - From the Lorenz curve
  - $I_{\text{Gini}}(F) = 1 - 2\int_0^1 L(F;q)\,\mathrm{d}q$
  - We can use results on $C(\cdot;q)$
- Standard form for $IF(z;I_{\text{Gini}},F)$:

Application: Gini

- Gini has multiple equivalent forms
  - From the Lorenz curve
  - $I_{\text{Gini}}(F) = 1 - 2\int_0^1 L(F;q)\,\mathrm{d}q$
  - We can use results on $C(\cdot;q)$
- Standard form for $IF(z; I_{\text{Gini}}, F)$:
  - $1 - I_{\text{Gini}}(F) - \frac{2C(F;F(z))}{\mu(F)} + z\frac{1-I_{\text{Gini}}(F)-2[1-F(z)]}{\mu(F)}$

Application: Gini

- Gini has multiple equivalent forms
  - From the Lorenz curve
  - $I_{\text{Gini}}(F) = 1 - 2\int_0^1 L(F;q)\,\mathrm{d}q$
  - We can use results on $C(\cdot;q)$
- Standard form for $IF(z;I_{\text{Gini}},F)$:
  - $1 - I_{\text{Gini}}(F) - \frac{2C(F;F(z))}{\mu(F)} + z\frac{1-I_{\text{Gini}}(F)-2[1-F(z)]}{\mu(F)}$
- Alternative form

Application: Gini

- Gini has multiple equivalent forms
  - From the Lorenz curve
  - $I_{\text{Gini}}(F) = 1 - 2\int_0^1 L(F;q)\,\mathrm{d}q$
  - We can use results on $C(\cdot;q)$
- Standard form for $IF(z;I_{\text{Gini}},F)$:
  - $1 - I_{\text{Gini}}(F) - \frac{2C(F;F(z))}{\mu(F)} + z\frac{1-I_{\text{Gini}}(F)-2[1-F(z)]}{\mu(F)}$
- Alternative form
  - note $E[C(F;F(z))] = E[z[1-F(z)]] = \frac{1-I_{\text{Gini}}(F)}{2}\mu(F)$

Application: Gini

- Gini has multiple equivalent forms
  - From the Lorenz curve
  - $I_{\text{Gini}}(F) = 1 - 2\int_0^1 L(F;q)\,\mathrm{d}q$
  - We can use results on $C(\cdot;q)$
- Standard form for $IF(z;I_{\text{Gini}},F)$:
  - $1 - I_{\text{Gini}}(F) - \frac{2C(F;F(z))}{\mu(F)} + z\frac{1-I_{\text{Gini}}(F)-2[1-F(z)]}{\mu(F)}$
- Alternative form
  - note $E[C(F;F(z))] = E[z[1-F(z)]] = \frac{1-I_{\text{Gini}}(F)}{2}\mu(F)$
  - let $Z = [1-I_{\text{Gini}}(F)]z - 2[C(F;F(z)) + z(1-F(z))]$

Application: Gini

- Gini has multiple equivalent forms
  - From the Lorenz curve
  - $I_{\text{Gini}}(F) = 1 - 2\int_0^1 L(F;q)\,\mathrm{d}q$
  - We can use results on $C(\cdot;q)$
- Standard form for $IF(z;I_{\text{Gini}},F)$:
  - $1 - I_{\text{Gini}}(F) - \frac{2C(F;F(z))}{\mu(F)} + z\frac{1-I_{\text{Gini}}(F)-2[1-F(z)]}{\mu(F)}$
- Alternative form
  - note $E[C(F;F(z))] = E[z[1-F(z)]] = \frac{1-I_{\text{Gini}}(F)}{2}\mu(F)$
  - let $Z = [1-I_{\text{Gini}}(F)]z - 2[C(F;F(z)) + z(1-F(z))]$
  - then $IF(z;I_{\text{Gini}},F) = (Z - E(Z))/\mu(F)$

Application: Gini

- Gini has multiple equivalent forms
  - From the Lorenz curve
  - $I_{\text{Gini}}(F) = 1 - 2\int_0^1 L(F;q)\,\mathrm{d}q$
  - We can use results on $C(\cdot;q)$
- Standard form for $IF(z;I_{\text{Gini}},F)$:
  - $1 - I_{\text{Gini}}(F) - \frac{2C(F;F(z))}{\mu(F)} + z\frac{1-I_{\text{Gini}}(F)-2[1-F(z)]}{\mu(F)}$
- Alternative form
  - note $E[C(F;F(z))] = E[z[1-F(z)]] = \frac{1-I_{\text{Gini}}(F)}{2}\mu(F)$
  - let $Z = [1-I_{\text{Gini}}(F)]z - 2[C(F;F(z)) + z(1-F(z))]$
  - then $IF(z;I_{\text{Gini}},F) = (Z - E(Z))/\mu(F)$
- $\text{var}\left(\sqrt{n}(I_{\text{Gini}}(F^{(n)}) - I_{\text{Gini}}(F))\right) = \text{var}(Z)/\mu(F)^2$

Outline

Stat methods

FAC-EF

Application: ASP class

Application: ASP class

- $P(F) := \int p(y, \zeta(F))\, \mathrm{d}F(y)$
  - p is non-increasing in y; is zero for $y \geq \zeta(F)$

Application: ASP class

- $P(F) := \int p(y, \zeta(F))\, \mathrm{d}F(y)$
  - p is non-increasing in y; is zero for $y \geq \zeta(F)$
  - large class of poverty measures

Application: ASP class

- $P(F) := \int p(y, \zeta(F))\, \mathrm{d}F(y)$
  - p is non-increasing in y; is zero for $y \geq \zeta(F)$
  - large class of poverty measures
  - example: $P^{\xi}_{\text{FGT}}(F) = \int_0^{\zeta_0} \left(\frac{\zeta_0 - y}{\zeta_0}\right)^{\xi} \mathrm{d}F(y) \qquad \xi \geq 0$
- $IF(z; P, F) = p(z, \zeta(F)) - P(F) + \int p_{\zeta}(y, \zeta)\, \mathrm{d}F(y) IF(z; \zeta, F)$

Application: ASP class

- $P(F) := \int p\left(y, \zeta\left(F\right)\right) \mathrm{d}F(y)$
  - p is non-increasing in y; is zero for $y \geq \zeta(F)$
  - large class of poverty measures
  - example: $P^{\xi}_{\text{FGT}}(F) = \int_0^{\zeta_0} \left(\frac{\zeta_0 - y}{\zeta_0}\right)^{\xi} \mathrm{d}F(y) \qquad \xi \geq 0$
- $IF(z; P, F) = p\left(z, \zeta(F)\right) - P(F) + \int p_{\zeta}\left(y, \zeta\right) \mathrm{d}F(y) IF(z; \zeta, F)$
- Case 1: $\zeta(F) = \zeta_0$:

Application: ASP class

- $P(F) := \int p(y, \zeta(F))\, \mathrm{d}F(y)$
  - p is non-increasing in y; is zero for $y \geq \zeta(F)$
  - large class of poverty measures
  - example: $P_{\mathrm{FGT}}^{\xi}(F) = \int_0^{\zeta_0} \left(\frac{\zeta_0 - y}{\zeta_0}\right)^{\xi} \mathrm{d}F(y) \qquad \xi \geq 0$
- $IF(z; P, F) = p(z, \zeta(F)) - P(F) + \int p_{\zeta}(y, \zeta)\, \mathrm{d}F(y) IF(z; \zeta, F)$
- Case 1: $\zeta(F) = \zeta_0$:
  - $IF(y; P, F) = Z - E(Z)$
  - $Z = p(y, \zeta_0)$

Application: ASP class

- $P(F) := \int p(y, \zeta(F))\,\mathrm{d}F(y)$
  - p is non-increasing in y; is zero for $y \geq \zeta(F)$
  - large class of poverty measures
  - example: $P^{\xi}_{\mathrm{FGT}}(F) = \int_0^{\zeta_0} \left(\frac{\zeta_0 - y}{\zeta_0}\right)^{\xi} \mathrm{d}F(y) \qquad \xi \geq 0$
- $IF(z; P, F) = p(z, \zeta(F)) - P(F) + \int p_{\zeta}(y, \zeta)\,\mathrm{d}F(y) IF(z; \zeta, F)$
- Case 1: $\zeta(F) = \zeta_0$:
  - $IF(y; P, F) = Z - E(Z)$
  - $Z = p(y, \zeta_0)$
  - AV is $\int p(z, \zeta_0)^2\,\mathrm{d}F(z) - P(F)^2$

Application: ASP class

- $P(F) := \int p\left(y, \zeta\left(F\right)\right) \mathrm{d}F(y)$
  - p is non-increasing in y; is zero for $y \geq \zeta(F)$
  - large class of poverty measures
  - example: $P_{\mathrm{FGT}}^{\xi}(F) = \int_0^{\zeta_0} \left(\frac{\zeta_0 - y}{\zeta_0}\right)^{\xi} \mathrm{d}F(y) \qquad \xi \geq 0$
- $IF(z;P,F) = p\left(z, \zeta\left(F\right)\right) - P(F) + \int p_{\zeta}\left(y, \zeta\right) \mathrm{d}F(y) IF(z;\zeta,F)$
- Case 1: $\zeta(F) = \zeta_0$:
  - $IF(y;P,F) = Z - E(Z)$
  - $Z = p(y, \zeta_0)$
  - AV is $\int p\left(z, \zeta_0\right)^2 \mathrm{d}F(z) - P(F)^2$
- Case 2: $\zeta(F) = \zeta_0 + \gamma y_q$

Application: ASP class

- $P(F) := \int p(y, \zeta(F))\, \mathrm{d}F(y)$
  - p is non-increasing in y; is zero for $y \geq \zeta(F)$
  - large class of poverty measures
  - example: $P_{\mathrm{FGT}}^{\xi}(F) = \int_0^{\zeta_0} \left(\frac{\zeta_0 - y}{\zeta_0}\right)^{\xi} \mathrm{d}F(y) \qquad \xi \geq 0$
- $IF(z; P, F) = p(z, \zeta(F)) - P(F) + \int p_{\zeta}(y, \zeta)\, \mathrm{d}F(y) IF(z; \zeta, F)$
- Case 1: $\zeta(F) = \zeta_0$:
  - $IF(y; P, F) = Z - E(Z)$
  - $Z = p(y, \zeta_0)$
  - AV is $\int p(z, \zeta_0)^2\, \mathrm{d}F(z) - P(F)^2$
- Case 2: $\zeta(F) = \zeta_0 + \gamma y_q$
  - $IF(z; \zeta, F) = \gamma \frac{q - \iota(y_q \geq z)}{f(y_q)}$

Application: ASP class

- $P(F) := \int p(y, \zeta(F))\, \mathrm{d}F(y)$
  - p is non-increasing in y; is zero for $y \geq \zeta(F)$
  - large class of poverty measures
  - example: $P_{\mathrm{FGT}}^{\xi}(F) = \int_0^{\zeta_0} \left(\frac{\zeta_0 - y}{\zeta_0}\right)^{\xi} \mathrm{d}F(y) \qquad \xi \geq 0$
- $IF(z; P, F) = p(z, \zeta(F)) - P(F) + \int p_{\zeta}(y, \zeta)\, \mathrm{d}F(y) IF(z; \zeta, F)$
- Case 1: $\zeta(F) = \zeta_0$:
  - $IF(y; P, F) = Z - E(Z)$
  - $Z = p(y, \zeta_0)$
  - AV is $\int p(z, \zeta_0)^2\, \mathrm{d}F(z) - P(F)^2$
- Case 2: $\zeta(F) = \zeta_0 + \gamma y_q$
  - $IF(z; \zeta, F) = \gamma \frac{q - \iota(y_q \geq z)}{f(y_q)}$
- Case 3: $\zeta(F) = \zeta_0 + \gamma \mu(F)$

Application: ASP class

- $P(F) := \int p\left(y, \zeta\left(F\right)\right) \mathrm{d}F(y)$
  - p is non-increasing in y; is zero for $y \geq \zeta(F)$
  - large class of poverty measures
  - example: $P_{\mathrm{FGT}}^{\xi}(F) = \int_0^{\zeta_0} \left(\frac{\zeta_0 - y}{\zeta_0}\right)^{\xi} \mathrm{d}F(y) \qquad \xi \geq 0$
- $IF(z;P,F) = p\left(z, \zeta(F)\right) - P(F) + \int p_{\zeta}\left(y, \zeta\right) \mathrm{d}F(y) IF(z;\zeta,F)$
- Case 1: $\zeta(F) = \zeta_0$:
  - $IF(y;P,F) = Z - E(Z)$
  - $Z = p(y, \zeta_0)$
  - AV is $\int p\left(z, \zeta_0\right)^2 \mathrm{d}F(z) - P(F)^2$
- Case 2: $\zeta(F) = \zeta_0 + \gamma y_q$
  - $IF(z;\zeta,F) = \gamma \frac{q - \iota(y_q \geq z)}{f(y_q)}$
- Case 3: $\zeta(F) = \zeta_0 + \gamma \mu(F)$
  - $IF(z;\zeta,F) = \gamma \frac{q - \iota(y_q \geq z)}{f(y_q)}$

Application: Sen poverty index

Application: Sen poverty index

- $P_{\text{Sen}}(F) = P^0_{\text{FGT}} I^p_{\text{Gini}} + P^1_{F\text{GT}}(1 - I^p_{\text{Gini}})$

Application: Sen poverty index

- $P_{\text{Sen}}(F) = P^0_{\text{FGT}} I^p_{\text{Gini}} + P^1_{F\text{GT}}(1 - I^p_{\text{Gini}})$
  - $P_{\text{Sen}}(F) = \frac{2}{\zeta_0 F(\zeta_0)} \int_0^{\zeta_0} (\zeta_0 - y)(F(\zeta_0) - F(y))\,\mathrm{d}F(y)$

Application: Sen poverty index

- $P_{\text{Sen}}(F) = P^0_{\text{FGT}} I^p_{\text{Gini}} + P^1_{F\text{GT}}(1 - I^p_{\text{Gini}})$
  - $P_{\text{Sen}}(F) = \frac{2}{\zeta_0 F(\zeta_0)} \int_0^{\zeta_0} (\zeta_0 - y)(F(\zeta_0) - F(y))\, \mathrm{d}F(y)$
- Consistent estimate:
  - $\hat{P}_{\text{Sen}} := P_{\text{Sen}}\left(F^{(n)}\right) = \frac{2}{n n_p \zeta_0} \sum_{i=1}^{n_p} (\zeta_0 - y_{(i)}) \left(n_p - i + \frac{1}{2}\right)$
  - $F(y_{(i)})$ estimated by $F^{(n)}(y_{(i)}) = \frac{2i-1}{2n}$

Application: Sen poverty index

- $P_{\text{Sen}}(F) = P^0_{\text{FGT}} I^p_{\text{Gini}} + P^1_{F\text{GT}}(1 - I^p_{\text{Gini}})$
  - $P_{\text{Sen}}(F) = \frac{2}{\zeta_0 F(\zeta_0)} \int_0^{\zeta_0} (\zeta_0 - y)(F(\zeta_0) - F(y))\, \mathrm{d}F(y)$
- Consistent estimate:
  - $\hat{P}_{\text{Sen}} := P_{\text{Sen}}\left(F^{(n)}\right) = \frac{2}{nn_p\zeta_0} \sum_{i=1}^{n_p} (\zeta_0 - y_{(i)})\left(n_p - i + \frac{1}{2}\right)$
  - $F(y_{(i)})$ estimated by $F^{(n)}(y_{(i)}) = \frac{2i-1}{2n}$
- $IF(z, P_{\text{Sen}}, F) = \frac{2}{\zeta_0 F(\zeta_0)}(Z - E(Z))$

Application: Sen poverty index

- $P_{\text{Sen}}(F) = P^0_{\text{FGT}} I^p_{\text{Gini}} + P^1_{FGT}(1 - I^p_{\text{Gini}})$
  - $P_{\text{Sen}}(F) = \frac{2}{\zeta_0 F(\zeta_0)} \int_0^{\zeta_0} (\zeta_0 - y)(F(\zeta_0) - F(y))\, \mathrm{d}F(y)$
- Consistent estimate:
  - $\hat{P}_{\text{Sen}} := P_{\text{Sen}}\left(F^{(n)}\right) = \frac{2}{n n_p \zeta_0} \sum_{i=1}^{n_p} (\zeta_0 - y_{(i)}) \left(n_p - i + \frac{1}{2}\right)$
  - $F(y_{(i)})$ estimated by $F^{(n)}(y_{(i)}) = \frac{2i-1}{2n}$
- $IF(z, P_{\text{Sen}}, F) = \frac{2}{\zeta_0 F(\zeta_0)}(Z - E(Z))$
  - $Z = \left[\zeta_0 F(\zeta_0) - \frac{\zeta_0 P_S}{2} - zF(\zeta_0) + zF(z) - C(F; F(z))\right] \iota(z \leq \zeta_0)$

Application: Sen poverty index

- $P_{\text{Sen}}(F) = P^0_{\text{FGT}} I^p_{\text{Gini}} + P^1_{FGT}(1 - I^p_{\text{Gini}})$
  - $P_{\text{Sen}}(F) = \frac{2}{\zeta_0 F(\zeta_0)} \int_0^{\zeta_0} (\zeta_0 - y)(F(\zeta_0) - F(y))\,\mathrm{d}F(y)$
- Consistent estimate:
  - $\hat{P}_{\text{Sen}} := P_{\text{Sen}}\left(F^{(n)}\right) = \frac{2}{n n_p \zeta_0} \sum_{i=1}^{n_p} (\zeta_0 - y_{(i)})\left(n_p - i + \frac{1}{2}\right)$
  - $F(y_{(i)})$ estimated by $F^{(n)}(y_{(i)}) = \frac{2i-1}{2n}$
- $IF(z, P_{\text{Sen}}, F) = \frac{2}{\zeta_0 F(\zeta_0)}(Z - E(Z))$
  - $Z = \left[\zeta_0 F(\zeta_0) - \frac{\zeta_0 P_S}{2} - zF(\zeta_0) + zF(z) - C(F; F(z))\right] \iota(z \leq \zeta_0)$
- $\widehat{\text{var}}\left(\hat{P}_{\text{Sen}}\right) = \frac{4}{(\zeta_0 n_p)^2} \sum_{i=1}^{n} (Z_i - \bar{Z})^2$

Application: Sen poverty index

- $P_{\text{Sen}}(F) = P^0_{\text{FGT}} I^p_{\text{Gini}} + P^1_{FGT}(1 - I^p_{\text{Gini}})$
  - $P_{\text{Sen}}(F) = \frac{2}{\zeta_0 F(\zeta_0)} \int_0^{\zeta_0} (\zeta_0 - y)(F(\zeta_0) - F(y))\, \mathrm{d}F(y)$
- Consistent estimate:
  - $\hat{P}_{\text{Sen}} := P_{\text{Sen}}\left(F^{(n)}\right) = \frac{2}{nn_p\zeta_0} \sum_{i=1}^{n_p} (\zeta_0 - y_{(i)}) \left(n_p - i + \frac{1}{2}\right)$
  - $F(y_{(i)})$ estimated by $F^{(n)}(y_{(i)}) = \frac{2i-1}{2n}$
- $IF(z, P_{\text{Sen}}, F) = \frac{2}{\zeta_0 F(\zeta_0)}(Z - E(Z))$
  - $Z = \left[\zeta_0 F(\zeta_0) - \frac{\zeta_0 P_S}{2} - zF(\zeta_0) + zF(z) - C(F; F(z))\right] \iota(z \leq \zeta_0)$
- $\widehat{\text{var}}\left(\hat{P}_{\text{Sen}}\right) = \frac{4}{(\zeta_0 n_p)^2} \sum_{i=1}^{n} (Z_i - \bar{Z})^2$
  - $Z_i = \frac{\zeta_0}{2}\left(\frac{2n_p}{n} - \hat{P}_{\text{Sen}}\right) - \frac{2n_p - 2i + 1}{2n} y_{(i)} - \frac{1}{n}\sum_{j=1}^{i} y_{(j)}$ for $i \leq n_p$
  - $\bar{Z} = n^{-1} \sum_{i=1}^{n} Z_i$

Outline

Finite sample properties

	Theil		Gini		SST	
	asym	boot	asym	boot	asym	boot
Lognormal						
$\sigma = 0.5$	0.927	0.936	0.942	0.943	0.926	0.952
$\sigma = 1.0$	0.871	0.913	0.922	0.936	0.945	0.940
$\sigma = 1.5$	0.746	0.854	0.876	0.920	0.964	0.937
Singh-Maddala						
$q = 1.7$	0.915	0.931	0.945	0.944	0.945	0.950
$q = 1.2$	0.856	0.905	0.925	0.934	0.945	0.951
$q = 0.7$	0.647	0.802	0.847	0.906	0.939	0.946

Table: Coverage of asymptotic and bootstrap confidence intervals at the 95% level for the Theil, Gini and SST indices, $n = 500$.

Inequality indices: unreliable CI with heavy-tailed distributions!

Inference with heavy-tailed distributions

	asym	boot	varstab	semip	mixture
Lognormal					
$\sigma = 0.5$	0.927	0.936	0.939	0.937	0.942
$\sigma = 1.0$	0.871	0.913	0.907	0.921	0.946
$\sigma = 1.5$	0.746	0.854	0.850	0.915	0.944
Singh-Maddala					
$q = 1.7$	0.915	0.931	0.933	0.926	0.928
$q = 1.2$	0.856	0.905	0.899	0.905	0.912
$q = 0.7$	0.647	0.802	0.796	0.871	0.789

Table: Coverage of asymptotic and bootstrap confidence intervals at the 95% level for the Theil index, for several bootstrap approaches, $n = 500$.

Testing equality of inequality measures

- Null hypothesis:

$$H_0 : W_x = W_y$$

- Independent samples: $X = \{x_1, \ldots, x_n\}, Y = \{y_1, \ldots, y_m\}$
- Test statistic:

$$\tau = (\hat{W}_x - \hat{W}_y)/[\widehat{\text{var}}(\hat{W}_x) + \widehat{\text{var}}(\hat{W}_y)]^{1/2}$$

- Monte Carlo permutation tests:
  - $F_x = F_y$: exact inference!![1]
  - $F_x \neq F_y$: not valid
- Dufour et al. (2013) propose a new bootstrap method:
  - with exact inference when $F_x = F_y$
  - valid when $F_x \neq F_y$

[1] even with very heavy-tailed distr. and very small samples

Standard bootstrap

- With independent samples, we test $H_0 : W_x = W_y$ with

$$\tau = (\hat{W}_x - \hat{W}_y)/[\widehat{\text{var}}(\hat{W}_x) + \widehat{\text{var}}(\hat{W}_y)]^{1/2}$$

- Bootstrap samples:
  - $X^\star$: resample with replacement n observations from X.
  - $Y^\star$: resample with replacement m observations from Y.
- Bootstrap test:

$$\tau_b^\star = [\hat{W}_{x_b^\star} - \hat{W}_{y_b^\star} - (\hat{W}_x - \hat{W}_y)]/[\widehat{\text{var}}(\hat{W}_{x_b^\star}) + \widehat{\text{var}}(\hat{W}_{y_b^\star})]^{1/2}$$

- Bootstrap distribution:

  EDF of the B bootstrap statistics, $\tau_b^\star$ for $b = 1, \ldots, B$

New bootstrap method

- Dufour et al (2013) propose generating bootstrap samples
  - $X^{\star\star}$: resample with replacement n observations from Z.
  - $Y^{\star\star}$: resample with replacement m observations from Z.

$$Z = \left\{ \frac{x_1}{\bar{x}}, \ldots, \frac{x_n}{\bar{x}}, \frac{y_1}{\bar{y}}, \ldots, \frac{y_m}{\bar{y}} \right\}$$

where $\bar{x}$ and $\bar{y}$ are sample means. The bootstrap test is

$$\tau_b^{\star\star} = [\hat{W}_{x_b^{\star\star}} - \hat{W}_{y_b^{\star\star}}]/[\widehat{\text{var}}(\hat{W}_{x_b^{\star\star}}) + \widehat{\text{var}}(\hat{W}_{y_b^{\star\star}})]^{1/2}$$

- This bootstrap procedure is
  - closely related to permutation test when $F_x = F_y$
  - still valid when $F_x \neq F_y$
  - respects the null hypothesis (Golden Rule)

Finite sample properties

$\alpha_x - \alpha_y$	asym	boot	bootH_0
0.00	0.0770	0.0614	0.0500
0.58	0.0796	0.0612	0.0496
0.96	0.0825	0.0639	0.0510
1.23	0.0865	0.0668	0.0539
1.57	0.0956	0.0719	0.0586
1.97	0.1138	0.0824	0.0705
2.57	0.1289	0.0911	0.0805

Table: Rejection frequencies for the Gini index, $H_0 : I_{\text{Gini}}(F_x) = I_{\text{Gini}}(F_y)$, as F_x moves away from F_y (as $\alpha_x - \alpha_y$ increases), at nominal level 0.05, $n = 50$.

Outline

First-order Dominance

First-order Dominance

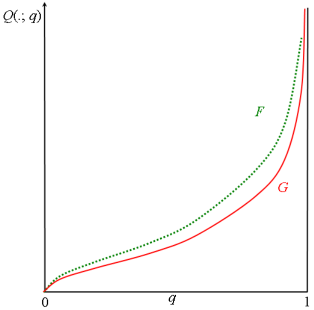
- "F first-order dominates G"

First-order Dominance

- "F first-order dominates G"
  - $\forall q \in \mathbb{Q} : Q(F,q) \geq Q(G,q)$
  - $\exists q \in \mathbb{Q} : Q(F,q) > Q(G,q)$

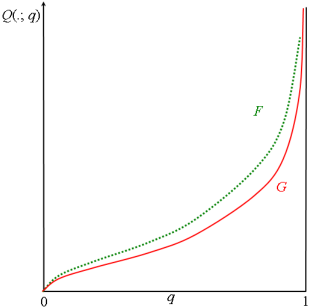
First-order Dominance

- "F first-order dominates G"
  - $\forall q \in \mathbb{Q} : Q(F,q) \geq Q(G,q)$
  - $\exists q \in \mathbb{Q} : Q(F,q) > Q(G,q)$
- Pen's Parade



-

First-order Dominance

- "F first-order dominates G"
  - $\forall q \in \mathbb{Q} : Q(F,q) \geq Q(G,q)$
  - $\exists q \in \mathbb{Q} : Q(F,q) > Q(G,q)$
- Pen's Parade
  - 

- "$W(F) \geq W(G)$, for any $W \in \mathbb{W}_1$"
  - $\mathbb{W}_1 := \{W | W(F) = \int \phi(y) \, \mathrm{d}F(y), \phi'(y) > 0\}$

Second-order Dominance

Second-order Dominance

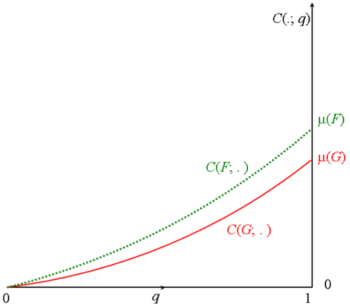
- "F second-order dominates G"

Second-order Dominance

- "F second-order dominates G"
  - $\forall q \in \mathbb{Q}: C(F,q) \geq C(G,q)$
  - $\exists q \in \mathbb{Q}: C(F,q) > C(G,q)$

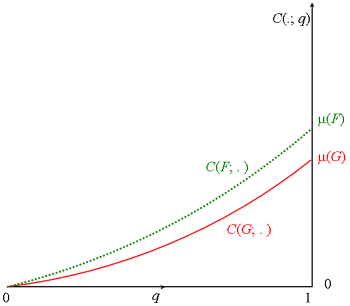
Second-order Dominance

- "*F* second-order dominates *G*"
  - $\forall q \in \mathbb{Q} : C(F,q) \geq C(G,q)$
  - $\exists q \in \mathbb{Q} : C(F,q) > C(G,q)$
- Generalised Lorenz Curve



-

Second-order Dominance

- "F second-order dominates G"
  - $\forall q \in \mathbb{Q} : C(F,q) \geq C(G,q)$
  - $\exists q \in \mathbb{Q} : C(F,q) > C(G,q)$
- Generalised Lorenz Curve
  - 

- "$W(F) \geq W(G)$, for any $W \in \mathbb{W}_2$"
  - $\mathbb{W}_2 := \{W | W(F) = \int \phi(y)\, dF(y), \phi'(y) > 0, \phi''(y) \leq 0\}$

Second-order: extensions

Second-order: extensions

- Second-order comparisons *scale independent*?

Second-order: extensions

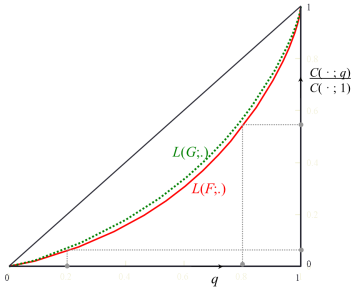
- Second-order comparisons *scale independent*?
  - for any $\lambda > 0$ distribution of y and of y/λ are equivalent

Second-order: extensions

- Second-order comparisons *scale independent*?
  - for any $\lambda > 0$ distribution of y and of y/λ are equivalent
  - Relative LC: $L(F;q) := \frac{C(F;q)}{\mu(F)} = \frac{C(F;q)}{C(F;1)}$

Second-order: extensions

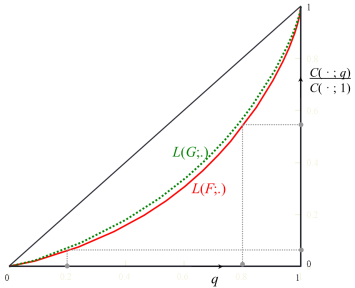
- Second-order comparisons *scale independent*?
  - for any $\lambda > 0$ distribution of y and of y/λ are equivalent
  - Relative LC: $L(F;q) := \frac{C(F;q)}{\mu(F)} = \frac{C(F;q)}{C(F;1)}$



-

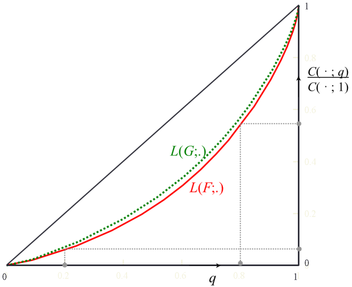
Second-order: extensions

- Second-order comparisons *scale independent*?
  - for any $\lambda > 0$ distribution of y and of y/λ are equivalent
  - Relative LC: $L(F;q) := \frac{C(F;q)}{\mu(F)} = \frac{C(F;q)}{C(F;1)}$



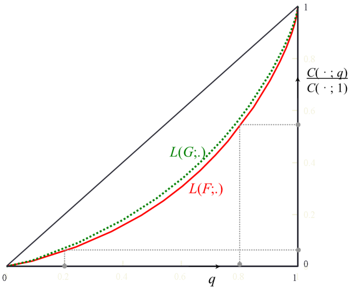
- Second-order comparisons *translation independent*?

Second-order: extensions

- Second-order comparisons *scale independent*?
  - for any $\lambda > 0$ distribution of y and of y/λ are equivalent
  - Relative LC: $L(F;q) := \frac{C(F;q)}{\mu(F)} = \frac{C(F;q)}{C(F;1)}$
  - 

- Second-order comparisons *translation independent*?
  - for any $\delta \in \mathbb{R}$ distribution of y and of $y + \delta$ are equivalent

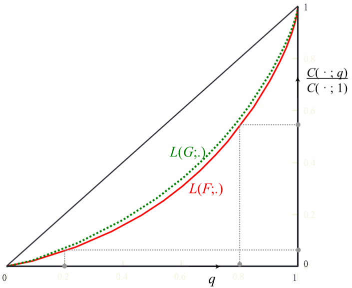
Second-order: extensions

- Second-order comparisons *scale independent*?
  - for any $\lambda > 0$ distribution of y and of y/λ are equivalent
  - Relative LC: $L(F;q) := \frac{C(F;q)}{\mu(F)} = \frac{C(F;q)}{C(F;1)}$
  - 

- Second-order comparisons *translation independent*?
  - for any $\delta \in \mathbb{R}$ distribution of y and of $y + \delta$ are equivalent
  - Absolute LC: $A(F;q) := C(F;q) - q\mu(F)$.

Second-order: extensions

- Second-order comparisons *scale independent*?
  - for any $\lambda > 0$ distribution of y and of y/λ are equivalent
  - Relative LC: $L(F;q) := \frac{C(F;q)}{\mu(F)} = \frac{C(F;q)}{C(F;1)}$
-



- Second-order comparisons *translation independent*?
  - for any $\delta \in \mathbb{R}$ distribution of y and of $y + \delta$ are equivalent
  - Absolute LC: $A(F;q) := C(F;q) - q\mu(F)$.

Dominance (general)

Dominance (general)

- Define $D_F^s(y) := \frac{1}{(s-1)!}\int_0^y (y-t)^{s-1}\,\mathrm{d}F(t)$
- General s-order dominance:

Dominance (general)

- Define $D_F^s(y) := \frac{1}{(s-1)!}\int_0^y (y-t)^{s-1}\,\mathrm{d}F(t)$
- General s-order dominance:
  - $\forall y \in \mathbb{R} : D_F^s(y) \leq D_G^s(y)$
  - $\exists y \in \mathbb{R} : D_F^s(y) < D_G^s(y)$

Dominance (general)

- Define $D_F^s(y) := \frac{1}{(s-1)!}\int_0^y (y-t)^{s-1}\,\mathrm{d}F(t)$
- General s-order dominance:
  - $\forall y \in \mathbb{R} : D_F^s(y) \leq D_G^s(y)$
  - $\exists y \in \mathbb{R} : D_F^s(y) < D_G^s(y)$
- Contains earlier dominance concepts
  - $s = 1$: first-order dominance
  - $s = 2$: second-order dominance

Dominance (general)

- Define $D_F^s(y) := \frac{1}{(s-1)!} \int_0^y (y-t)^{s-1}\, \mathrm{d}F(t)$
- General s-order dominance:
  - $\forall y \in \mathbb{R} : D_F^s(y) \leq D_G^s(y)$
  - $\exists y \in \mathbb{R} : D_F^s(y) < D_G^s(y)$
- Contains earlier dominance concepts
  - $s = 1$: first-order dominance
  - $s = 2$: second-order dominance
- $D_F^s(\zeta_0)$ is equal to the FGT poverty index, up to a scale factor

Dominance (general)

- Define $D_F^s(y) := \frac{1}{(s-1)!}\int_0^y (y-t)^{s-1}\,\mathrm{d}F(t)$
- General s-order dominance:
  - $\forall y \in \mathbb{R} : D_F^s(y) \leq D_G^s(y)$
  - $\exists y \in \mathbb{R} : D_F^s(y) < D_G^s(y)$
- Contains earlier dominance concepts
  - $s = 1$: first-order dominance
  - $s = 2$: second-order dominance
- $D_F^s(\zeta_0)$ is equal to the FGT poverty index, up to a scale factor
  - If, for all $[\zeta_0^-; \zeta_0^+]$, $D_F^s(\zeta_0) < D_G^s(\zeta_0)$:
  - then poverty lower in F than in G.

Dominance (general)

- Define $D_F^s(y) := \frac{1}{(s-1)!} \int_0^y (y-t)^{s-1}\, \mathrm{d}F(t)$
- General s-order dominance:
  - $\forall y \in \mathbb{R} : D_F^s(y) \leq D_G^s(y)$
  - $\exists y \in \mathbb{R} : D_F^s(y) < D_G^s(y)$
- Contains earlier dominance concepts
  - $s = 1$: first-order dominance
  - $s = 2$: second-order dominance
- $D_F^s(\zeta_0)$ is equal to the FGT poverty index, up to a scale factor
  - If, for all $[\zeta_0^-; \zeta_0^+]$, $D_F^s(\zeta_0) < D_G^s(\zeta_0)$:
  - then poverty lower in F than in G.

Outline

Implementation: basics

Implementation: basics

1. Choose a finite collection of population proportions $\Theta \subset \mathbb{Q}$

Implementation: basics

1. Choose a finite collection of population proportions $\Theta \subset \mathbb{Q}$
2. For each $q \in \Theta$ compute sample quantiles, cumulations:

Implementation: basics

1. Choose a finite collection of population proportions $\Theta \subset \mathbb{Q}$
2. For each $q \in \Theta$ compute sample quantiles, cumulations:
   - let $\kappa(n, q)$ be largest integer $\leq nq - q + 1$

Implementation: basics

1. Choose a finite collection of population proportions $\Theta \subset \mathbb{Q}$
2. For each $q \in \Theta$ compute sample quantiles, cumulations:
   - let $\kappa(n,q)$ be largest integer $\leq nq - q + 1$
   - quantiles: $\hat{y}_q := Q\left(F^{(n)}; q\right) = y_{(\kappa(n,q))}$

Implementation: basics

1. Choose a finite collection of population proportions $\Theta \subset \mathbb{Q}$
2. For each $q \in \Theta$ compute sample quantiles, cumulations:
   - let $\kappa(n,q)$ be largest integer $\leq nq - q + 1$
   - quantiles: $\hat{y}_q := Q\left(F^{(n)}; q\right) = y_{(\kappa(n,q))}$
   - cumulations: $\hat{c}_q := C\left(F^{(n)}; q\right) = \frac{1}{n}\sum_{i=1}^{\kappa(n,q)} y_{(i)}$

Implementation: basics

1. Choose a finite collection of population proportions $\Theta \subset \mathbb{Q}$
2. For each $q \in \Theta$ compute sample quantiles, cumulations:
   - let $\kappa(n,q)$ be largest integer $\leq nq - q + 1$
   - quantiles: $\hat{y}_q := Q\left(F^{(n)}; q\right) = y_{(\kappa(n,q))}$
   - cumulations: $\hat{c}_q := C\left(F^{(n)}; q\right) = \frac{1}{n}\sum_{i=1}^{\kappa(n,q)} y_{(i)}$
3. Compute the variances and covariances of
   - sample quantiles (first-order)
   - income cumulations (second order)

Implementation: basics

1. Choose a finite collection of population proportions $\Theta \subset \mathbb{Q}$
2. For each $q \in \Theta$ compute sample quantiles, cumulations:
   - let $\kappa(n,q)$ be largest integer $\leq nq - q + 1$
   - quantiles: $\hat{y}_q := Q\left(F^{(n)}; q\right) = y_{(\kappa(n,q))}$
   - cumulations: $\hat{c}_q := C\left(F^{(n)}; q\right) = \frac{1}{n}\sum_{i=1}^{\kappa(n,q)} y_{(i)}$
3. Compute the variances and covariances of
   - sample quantiles (first-order)
   - income cumulations (second order)
4. Specify carefully the ranking hypothesis to be tested

Covariances, Parade and GLC

Covariances, Parade and GLC

- For any $q, q' \in \mathbb{Q}$, compute covariances of ordinates

Covariances, Parade and GLC

- For any $q, q' \in \mathbb{Q}$, compute covariances of ordinates
- $\sqrt{n}\hat{y}_q$, $\sqrt{n}\hat{y}_{q'}$ asymp normally distributed, cov is $\frac{q[1-q']}{f(y_q)f(y_{q'})}$

Covariances, Parade and GLC

- For any $q, q' \in \mathbb{Q}$, compute covariances of ordinates
- $\sqrt{n}\hat{y}_q$, $\sqrt{n}\hat{y}_{q'}$ asymp normally distributed, cov is $\frac{q[1-q']}{f(y_q)f(y_{q'})}$
- $\sqrt{n}\hat{c}_q$, $\sqrt{n}\hat{c}_{q'}$ asymp normally distributed; cov is:
    - $\omega_{qq'} := s_q + [qy_q - c_q]\left[y_{q'} - q'y_{q'} + c_{q'}\right] - y_q c_q,\ q \le q'$
    - $S(F;q) := \int_{\underline{y}}^{y_q} y^2 \,\mathrm{d}F(y) =: s_q$

Covariances, Parade and GLC

- For any $q, q' \in \mathbb{Q}$, compute covariances of ordinates
- $\sqrt{n}\hat{y}_q$, $\sqrt{n}\hat{y}_{q'}$ asymp normally distributed, cov is $\frac{q[1-q']}{f(y_q)f(y_{q'})}$
- $\sqrt{n}\hat{c}_q$, $\sqrt{n}\hat{c}_{q'}$ asymp normally distributed; cov is:
  - $\omega_{qq'} := s_q + [qy_q - c_q]\left[y_{q'} - q'y_{q'} + c_{q'}\right] - y_q c_q,\ q \leq q'$
  - $S(F;q) := \int_{\underline{y}}^{y_q} y^2 \, \mathrm{d}F(y) =: s_q$
  - $\widehat{s}_q := S\left(F^{(n)};q\right) = \frac{1}{n}\sum_{i=1}^{\kappa(n,q)} y_{(i)}^2$

Covariances, Parade and GLC

- For any $q, q' \in \mathbb{Q}$, compute covariances of ordinates
- $\sqrt{n}\hat{y}_q$, $\sqrt{n}\hat{y}_{q'}$ asymp normally distributed, cov is $\frac{q[1-q']}{f(y_q)f(y_{q'})}$
- $\sqrt{n}\hat{c}_q$, $\sqrt{n}\hat{c}_{q'}$ asymp normally distributed; cov is:
  - $\omega_{qq'} := s_q + [qy_q - c_q]\left[y_{q'} - q'y_{q'} + c_{q'}\right] - y_q c_q,\, q \leq q'$
  - $S(F;q) := \int_{\underline{y}}^{y_q} y^2 \, \mathrm{d}F(y) =: s_q$
  - $\widehat{s}_q := S\left(F^{(n)};q\right) = \frac{1}{n}\sum_{i=1}^{\kappa(n,q)} y_{(i)}^2$
- Derivation:
  - $\omega_{qq'} = \int IF(z;C(F;q),F)IF(z;C(F;q'),F)\,\mathrm{d}F(z))$

Covariances, Parade and GLC

- For any $q, q' \in \mathbb{Q}$, compute covariances of ordinates
- $\sqrt{n}\hat{y}_q$, $\sqrt{n}\hat{y}_{q'}$ asymp normally distributed, cov is $\frac{q[1-q']}{f(y_q)f(y_{q'})}$
- $\sqrt{n}\hat{c}_q$, $\sqrt{n}\hat{c}_{q'}$ asymp normally distributed; cov is:
  - $\omega_{qq'} := s_q + [qy_q - c_q]\left[y_{q'} - q'y_{q'} + c_{q'}\right] - y_q c_q,\ q \le q'$
  - $S(F;q) := \int_{\underline{y}}^{y_q} y^2 \,\mathrm{d}F(y) =: s_q$
  - $\widehat{s}_q := S\left(F^{(n)};q\right) = \frac{1}{n}\sum_{i=1}^{\kappa(n,q)} y_{(i)}^2$
- Derivation:
  - $\omega_{qq'} = \int IF(z;C(F;q),F)IF(z;C(F;q'),F)\,\mathrm{d}F(z))$
  - but $IF(z;C(F;q),F) = qy_q - c_q + \iota(y_q \ge z)[z - y_q]$

Covariances, Parade and GLC

- For any $q, q' \in \mathbb{Q}$, compute covariances of ordinates
- $\sqrt{n}\hat{y}_q$, $\sqrt{n}\hat{y}_{q'}$ asymp normally distributed, cov is $\frac{q[1-q']}{f(y_q)f(y_{q'})}$
- $\sqrt{n}\hat{c}_q$, $\sqrt{n}\hat{c}_{q'}$ asymp normally distributed; cov is:
  - $\omega_{qq'} := s_q + [qy_q - c_q]\left[y_{q'} - q'y_{q'} + c_{q'}\right] - y_q c_q,\ q \le q'$
  - $S(F;q) := \int_{\underline{y}}^{y_q} y^2 \,\mathrm{d}F(y) =: s_q$
  - $\widehat{s}_q := S\left(F^{(n)};q\right) = \frac{1}{n}\sum_{i=1}^{\kappa(n,q)} y_{(i)}^2$
- Derivation:
  - $\omega_{qq'} = \int IF(z;C(F;q),F)IF(z;C(F;q'),F)\,\mathrm{d}F(z))$
  - but $IF(z;C(F;q),F) = qy_q - c_q + \iota(y_q \ge z)[z - y_q]$
  - So, given that $\iota(x_{q'} \ge z) = 1$ whenever $\iota(x_q \ge z) = 1$:
  - $\omega_{qq'} = [qy_q - c_q]\left[q'y_{q'} - c_{q'}\right] + \int_{\underline{y}}^{y_{q'}} [qy_q - c_q][z - y_{q'}]\,\mathrm{d}F(z) + \int_{\underline{y}}^{y_q}\left[q'y_{q'} - c_{q'} + z - y_{q'}\right][z - y_q]\,\mathrm{d}F(z)$

Covariances, GLC (2)

- We can also use the "short form" of the IF method:

Covariances, GLC (2)

- We can also use the “short form” of the IF method:
  - $IF(z; C(F,q), F) = Z_q - E(Z_q)$
  - $Z_q = [z - y_q]\iota(z \leq y_q)$.

Covariances, GLC (2)

- We can also use the "short form" of the IF method:
  - $IF(z; C(F,q), F) = Z_q - E(Z_q)$
  - $Z_q = [z - y_q]\iota(z \leq y_q)$.
- Asymptotic covariance of $\sqrt{n}\hat{c}_q$ and $\sqrt{n}\hat{c}_{q'}$:

Covariances, GLC (2)

- We can also use the “short form” of the IF method:
  - $IF(z;C(F,q),F) = Z_q - E(Z_q)$
  - $Z_q = [z-y_q]\iota(z \leq y_q)$.
- Asymptotic covariance of $\sqrt{n}\hat{c}_q$ and $\sqrt{n}\hat{c}_{q'}$:
  - $\omega_{qq'} = \text{cov}(Z_q, Z_{q'})$

Covariances, GLC (2)

- We can also use the "short form" of the IF method:
  - $IF(z;C(F,q),F) = Z_q - E(Z_q)$
  - $Z_q = [z - y_q]\iota(z \le y_q)$.
- Asymptotic covariance of $\sqrt{n}\hat{c}_q$ and $\sqrt{n}\hat{c}_{q'}$:
  - $\omega_{qq'} = \mathrm{cov}(Z_q, Z_{q'})$
  - $\widehat{\mathrm{cov}}(\hat{c}_q, \hat{c}_{q'}) = \frac{1}{n}\widehat{\omega}_{qq'} = \frac{1}{n^2}\sum_{i=1}^{n}(Z_{iq} - \bar{Z}_q)(Z_{iq'} - \bar{Z}_{q'})$

Covariances, GLC (2)

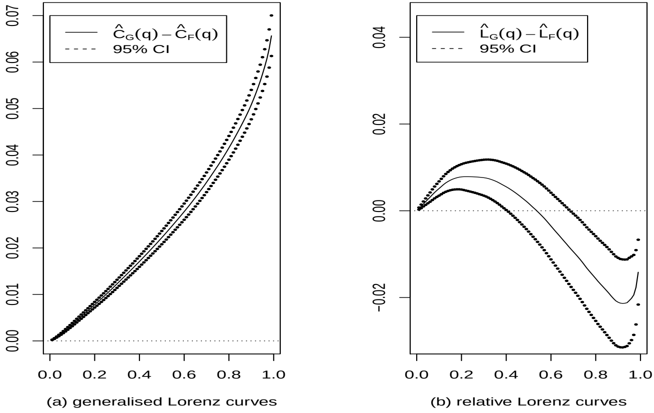
- We can also use the “short form” of the IF method:
  - $IF(z; C(F,q), F) = Z_q - E(Z_q)$
  - $Z_q = [z - y_q]\iota(z \leq y_q)$.
- Asymptotic covariance of $\sqrt{n}\hat{c}_q$ and $\sqrt{n}\hat{c}_{q'}$:
  - $\omega_{qq'} = \text{cov}(Z_q, Z_{q'})$
  - $\widehat{\text{cov}}(\hat{c}_q, \hat{c}_{q'}) = \frac{1}{n}\widehat{\omega}_{qq'} = \frac{1}{n^2}\sum_{i=1}^{n}(Z_{iq} - \bar{Z}_q)(Z_{iq'} - \bar{Z}_{q'})$
  - $Z_{iq} = [y_i - \hat{y}_q]\,\iota(y_i \leq \hat{y}_q)$

Covariances, GLC (2)

- We can also use the "short form" of the IF method:
  - $IF(z;C(F,q),F) = Z_q - E(Z_q)$
  - $Z_q = [z - y_q]\iota(z \le y_q)$.
- Asymptotic covariance of $\sqrt{n}\hat{c}_q$ and $\sqrt{n}\hat{c}_{q'}$:
  - $\omega_{qq'} = \text{cov}(Z_q, Z_{q'})$
  - $\widehat{\text{cov}}(\hat{c}_q, \hat{c}_{q'}) = \frac{1}{n}\widehat{\omega}_{qq'} = \frac{1}{n^2}\sum_{i=1}^{n}(Z_{iq} - \bar{Z}_q)(Z_{iq'} - \bar{Z}_{q'})$
  - $Z_{iq} = [y_i - \hat{y}_q]\,\iota(y_i \le \hat{y}_q)$
- Consistent estimate:
  $\widehat{\omega}_{qq'} := \widehat{s}_q + [q\widehat{y}_q - \widehat{c}_q]\left[\widehat{y}_{q'} - q'\widehat{y}_{q'} + \widehat{c}_{q'}\right] - \hat{y}_q\widehat{c}_q$

Outline

Dominance: an intuitive application



Difference between two empirical Lorenz curves, $n = 5\,000$

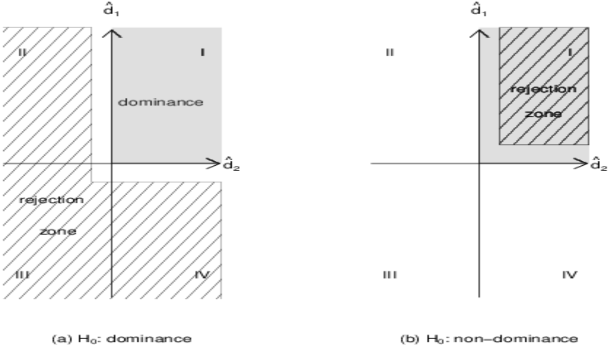
Dominance: an intuitive application

	Distribution F		Distribution G	
	Index	$CI_{95\%}$	Index	$CI_{95\%}$
Poverty measures*				
P^0_{FGT}	0.1134	*[0.1046;0.1222]*	0.0260	*[0.0216;0.0304]*
P^1_{FGT}	0.0299	*[0.0270;0.0329]*	0.0053	*[0.0042;0.0065]*
P_{Sen}	0.0426	*[0.0385;0.0466]*	0.0077	*[0.0061;0.0093]*
P_{SST}	0.0579	*[0.0523;0.0635]*	0.0106	*[0.0083;0.0129]*
Generalised Entropy measures				
I^{-1}_{GE}	0.1803	*[0.1694;0.1913]*	0.1568	*[0.1468;0.1667]*
I^{0}_{GE}	0.1416	*[0.1351;0.1481]*	0.1420	*[0.1324;0.1516]*
I^{1}_{GE}	0.1360	*[0.1289;0.1430]*	0.1570	*[0.1411;0.1729]*
I^{2}_{GE}	0.1548	*[0.1431;0.1665]*	0.2266	*[0.1798;0.2734]*
I_{Gini}	0.2849	*[0.2785;0.2913]*	0.2909	*[0.2816;0.3001]*

* The poverty line is half the median of the sample drawn from distribution F: $\zeta_0 = 0.07565776$.

Outline

The null hypothesis: dominance or non-dominance



Tests of dominance and non-dominance. The first quadrant, I, corresponds to dominance of G by F in the sample (grey area). The quadrants II, III and IV correspond to non-dominance.

Outline

Hypothesis testing

- Under the null of dominance (F dominates G), we have

$$\begin{aligned} H_0: &\quad D_F^s(y) \leq D_G^s(y), \quad \text{for all } y \in \mathbb{Y}, \\ H_1: &\quad D_F^s(y) > D_G^s(y), \quad \text{for some } y \in \mathbb{Y}. \end{aligned}$$

  Test based on the supremum of individual differences:

$$\tau = \sup_{y \in \mathbb{Y}} \left(\hat{D}_F^s(y) - \hat{D}_G^s(y) \right).$$

- Under the null of non-dominance (F does not dominate G):

$$\begin{aligned} H_0: &\quad D_F^s(y) \geq D_G^s(y), \quad \text{for some } y \in \mathbb{Y}, \\ H_1: &\quad D_F^s(y) < D_G^s(y), \quad \text{for all } y \in \mathbb{Y}. \end{aligned}$$

  Test based on the infinum of individual differences:

$$\tau' = \inf_{y \in \mathbb{Y}^r} \left(\hat{D}_G^s(y) - \hat{D}_F^s(y) \right).$$

Overview

Overview

- Focus on data issues that go beyond sampling

Overview

- Focus on data issues that go beyond sampling
  - Robustness
  - Incomplete data
  - Semi-parametric modelling

Overview

- Focus on data issues that go beyond sampling
  - Robustness
  - Incomplete data
  - Semi-parametric modelling
- Reuse tools from earlier material

Overview

- Focus on data issues that go beyond sampling
  - Robustness
  - Incomplete data
  - Semi-parametric modelling
- Reuse tools from earlier material
- Apply similar techniques

Outline

Robustness (1)

- Suppose true distribution is mixed with contamination

Robustness (1)

- Suppose true distribution is mixed with contamination
  - point mass at z: $H^{(z)}(y) = \iota(y \geq z)$
  - the mixture: $G = [1-\delta]F + \delta H^{(z)}$
  - δ : "size" of contamination

Robustness (1)

- Suppose true distribution is mixed with contamination
  - point mass at z: $H^{(z)}(y) = \iota(y \geq z)$
  - the mixture: $G = [1-\delta]F + \delta H^{(z)}$
  - δ : "size" of contamination
- Use IF to see effect of infinitesimal contamination at z

Robustness (1)

- Suppose true distribution is mixed with contamination
  - point mass at z: $H^{(z)}(y) = \iota(y \geq z)$
  - the mixture: $G = [1-\delta]F + \delta H^{(z)}$
  - δ : "size" of contamination
- Use IF to see effect of infinitesimal contamination at z

1. Example: the mean

Robustness (1)

- Suppose true distribution is mixed with contamination
  - point mass at z: $H^{(z)}(y) = \iota(y \geq z)$
  - the mixture: $G = [1-\delta]F + \delta H^{(z)}$
  - δ : "size" of contamination
- Use IF to see effect of infinitesimal contamination at z

1. Example: the mean
   - $\mu(G) = \mu\left([1-\delta]F + \delta H^{(z)}\right) = [1-\delta]\mu(F) + \delta\mu\left(H^{(z)}\right)$

Robustness (1)

- Suppose true distribution is mixed with contamination
  - point mass at z: $H^{(z)}(y) = \iota(y \geq z)$
  - the mixture: $G = [1-\delta]F + \delta H^{(z)}$
  - δ : "size" of contamination
- Use IF to see effect of infinitesimal contamination at z

1. Example: the mean
   - $\mu(G) = \mu\left([1-\delta]F + \delta H^{(z)}\right) = [1-\delta]\mu(F) + \delta\mu\left(H^{(z)}\right)$
   - $\mu(G) = [1-\delta]\mu(F) + \delta z$

Robustness (1)

- Suppose true distribution is mixed with contamination
  - point mass at z: $H^{(z)}(y) = \iota(y \geq z)$
  - the mixture: $G = [1-\delta]F + \delta H^{(z)}$
  - δ : "size" of contamination
- Use IF to see effect of infinitesimal contamination at z

1. Example: the mean
   - $\mu(G) = \mu\left([1-\delta]F + \delta H^{(z)}\right) = [1-\delta]\mu(F) + \delta\mu\left(H^{(z)}\right)$
   - $\mu(G) = [1-\delta]\mu(F) + \delta z$
   - $IF(z; \mu, F) = z - \mu(F)$

Robustness (1)

- Suppose true distribution is mixed with contamination
  - point mass at z: $H^{(z)}(y) = \iota(y \geq z)$
  - the mixture: $G = [1-\delta]F + \delta H^{(z)}$
  - δ : "size" of contamination

- Use IF to see effect of infinitesimal contamination at z

1. Example: the mean
   - $\mu(G) = \mu\left([1-\delta]F + \delta H^{(z)}\right) = [1-\delta]\mu(F) + \delta\mu\left(H^{(z)}\right)$
   - $\mu(G) = [1-\delta]\mu(F) + \delta z$
   - $IF(z;\mu,F) = z - \mu(F)$
2. Example: the median

Robustness (1)

- Suppose true distribution is mixed with contamination
  - point mass at z: $H^{(z)}(y) = \iota(y \geq z)$
  - the mixture: $G = [1-\delta]F + \delta H^{(z)}$
  - δ : "size" of contamination
- Use IF to see effect of infinitesimal contamination at z

1. Example: the mean
   - $\mu(G) = \mu\left([1-\delta]F + \delta H^{(z)}\right) = [1-\delta]\mu(F) + \delta\mu\left(H^{(z)}\right)$
   - $\mu(G) = [1-\delta]\mu(F) + \delta z$
   - $IF(z;\mu,F) = z - \mu(F)$
2. Example: the median
   - $IF(z;Q(\cdot,0.5),F) = \frac{q-\iota(q\geq F(z))}{f(Q(F,0.5))} = \frac{q-\iota(y_{0.5}\geq z)}{f(y_{0.5})}$

Robustness (2)

- Inequality and poverty indices respond differently to contamination
- Consider two important members of W_{QAD} class

Robustness (2)

- Inequality and poverty indices respond differently to contamination
- Consider two important members of W_{QAD} class
  - IF in general case:
  - $\varphi(z,\mu(F)) - W_{\mathrm{QAD}}(F) + [z-\mu(F)]\int \varphi_{\mu}(z,\mu(F))\,\mathrm{d}F(z)$
- Inequality

Robustness (2)

- Inequality and poverty indices respond differently to contamination
- Consider two important members of W_{QAD} class
  - IF in general case:
  - $\varphi\left(z, \mu(F)\right) - W_{\mathrm{QAD}}(F) + \left[z - \mu(F)\right] \int \varphi_{\mu}\left(z, \mu(F)\right) \mathrm{d}F(z)$
- Inequality
  - Compute IF for GE:

Robustness (2)

- Inequality and poverty indices respond differently to contamination
- Consider two important members of W_{QAD} class
  - IF in general case:
  - $\varphi(z, \mu(F)) - W_{\text{QAD}}(F) + [z - \mu(F)] \int \varphi_\mu(z, \mu(F)) \, \mathrm{d}F(z)$
- Inequality
  - Compute IF for GE:
  - $\varphi(z, \mu(F)) = \frac{[z/\mu(F)]^\xi - 1}{\xi^2 - \xi}$
  - unbounded for all values of ξ

Robustness (2)

- Inequality and poverty indices respond differently to contamination
- Consider two important members of W_{QAD} class
  - IF in general case:
  - $\varphi(z, \mu(F)) - W_{\mathrm{QAD}}(F) + [z - \mu(F)] \int \varphi_\mu (z, \mu(F))\, \mathrm{d}F(z)$
- Inequality
  - Compute IF for GE:
  - $\varphi(z, \mu(F)) = \frac{[z/\mu(F)]^\xi - 1}{\xi^2 - \xi}$
  - unbounded for all values of ξ
  - also unbounded effect on mean
- Poverty

Robustness (2)

- Inequality and poverty indices respond differently to contamination
- Consider two important members of W_{QAD} class
  - IF in general case:
  - $\varphi(z,\mu(F)) - W_{\text{QAD}}(F) + [z-\mu(F)] \int \varphi_\mu(z,\mu(F))\,\mathrm{d}F(z)$
- Inequality
  - Compute IF for GE:
  - $\varphi(z,\mu(F)) = \frac{[z/\mu(F)]^{\xi}-1}{\xi^2-\xi}$
  - unbounded for all values of ξ
  - also unbounded effect on mean
- Poverty
  - take ASP case (fixed poverty line ζ_0):
  - $IF(z;P,F) = p(z,\zeta_0) - P(F)$

Robustness (2)

- Inequality and poverty indices respond differently to contamination
- Consider two important members of W_{QAD} class
  - IF in general case:
  - $\varphi(z, \mu(F)) - W_{\text{QAD}}(F) + [z - \mu(F)] \int \varphi_\mu(z, \mu(F))\, \mathrm{d}F(z)$
- Inequality
  - Compute IF for GE:
  - $\varphi(z, \mu(F)) = \frac{[z/\mu(F)]^\xi - 1}{\xi^2 - \xi}$
  - unbounded for all values of ξ
  - also unbounded effect on mean
- Poverty
  - take ASP case (fixed poverty line ζ_0):
  - $IF(z; P, F) = p(z, \zeta_0) - P(F)$
  - example (FGT): $p(z, \zeta_0) = [\max(1 - z/\zeta_0, 0)]^\xi$

Robustness (3)

- Use parametric $f(y;\theta)$ for part of the income distribution?

Robustness (3)

- Use parametric $f(y;\theta)$ for part of the income distribution?
- MLE are efficient but usually non-robust

Robustness (3)

- Use parametric $f(y;\theta)$ for part of the income distribution?
- MLE are efficient but usually non-robust
- *M-estimators* characterised by $\sum_{i=1}^{n} \psi(y_i;\theta) = 0, \psi : \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}^p$

Robustness (3)

- Use parametric $f(y;\theta)$for part of the income distribution?
- MLE are efficient but usually non-robust
- *M-estimators* characterised by $\sum_{i=1}^{n}\psi(y_i;\theta)=0, \psi:\mathbb{R}\times\mathbb{R}^p\rightarrow\mathbb{R}^p$
- OBRE defined as solution in θ of
  - $\sum_{i=1}^{n}\psi(x_i;\theta)=\sum_{i=1}^{n}[s(x_i;\theta)-a(\theta)]\cdot W_c(x_i;\theta)=0$

Robustness (3)

- Use parametric $f(y;\theta)$ for part of the income distribution?
- MLE are efficient but usually non-robust
- *M-estimators* characterised by $\sum_{i=1}^{n} \psi(y_i;\theta) = 0, \psi : \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}^p$
- OBRE defined as solution in θ of
  - $\sum_{i=1}^{n} \psi(x_i;\theta) = \sum_{i=1}^{n} [s(x_i;\theta) - a(\theta)] \cdot W_c(x_i;\theta) = 0$
  - $c \geq \sqrt{p}$, fixed a bound on the *IF*
  - weights: $W_c(x;\theta) = \min\left\{ 1 \; ; \frac{c}{\|A(\theta)[s(x;\theta)-a(\theta)]\|} \right\}$
  - scores function, $s(x;\theta) = \partial/\partial\theta \log f(x;\theta)$

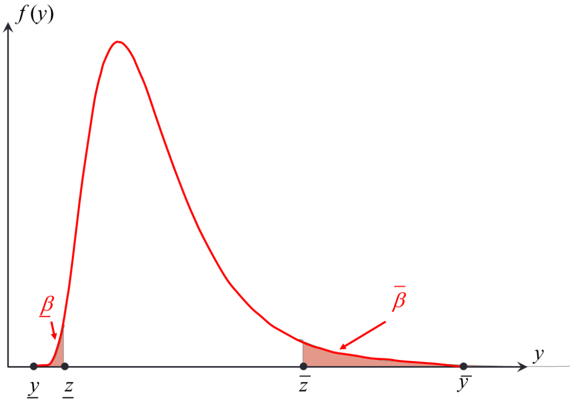
Robustness (3)

- Use parametric $f(y;\theta)$ for part of the income distribution?
- MLE are efficient but usually non-robust
- *M-estimators* characterised by $\sum_{i=1}^{n} \psi(y_i;\theta) = 0, \psi : \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}^p$
- OBRE defined as solution in θ of
  - $\sum_{i=1}^{n} \psi(x_i;\theta) = \sum_{i=1}^{n} [s(x_i;\theta) - a(\theta)] \cdot W_c(x_i;\theta) = 0$
  - $c \geq \sqrt{p}$, fixed a bound on the *IF*
  - weights: $W_c(x;\theta) = \min\left\{ 1 \, ; \frac{c}{\|A(\theta)[s(x;\theta)-a(\theta)]\|} \right\}$
  - scores function, $s(x;\theta) = \partial/\partial\theta \log f(x;\theta)$
- $p \times p$ matrix $A(\theta)$ and $a(\theta) \in \mathbb{R}^p$:
  - $E\left[\psi(x;\theta)\psi(x;\theta)^T\right] = \left[A(\theta)^T A(\theta)\right]^{-1}; E[\psi(x;\theta)] = 0$
  - c: regulator between efficiency (high) and robustness (low)

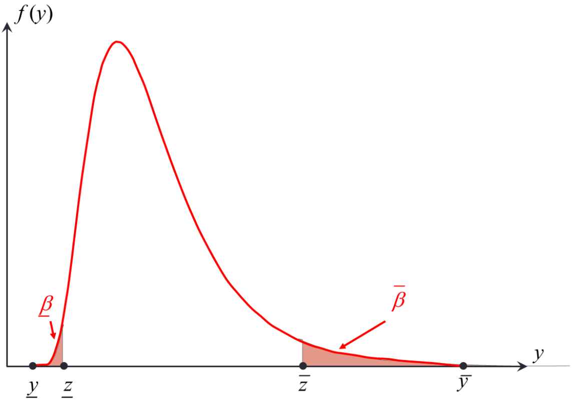
Outline

Censoring and Truncation

Censoring and Truncation

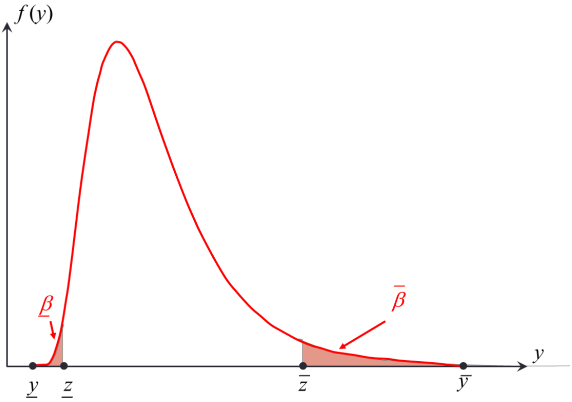


Censoring and Truncation



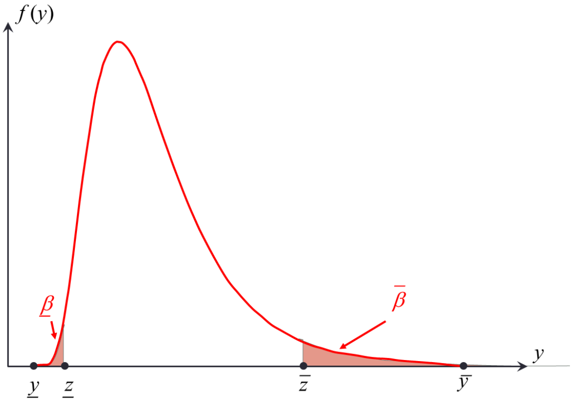
- Empirical distribution is random

Censoring and Truncation



- Empirical distribution is random
- Fixed boundaries $(\underline{z}, \overline{z})$ on excluded portion

Censoring and Truncation



- Empirical distribution is random
- Fixed boundaries $(\underline{z}, \bar{z})$ on excluded portion
- Therefore size of excluded portions $(\underline{\beta}, 1 - \bar{\beta})$ is random

Censoring and Truncation

Censoring and Truncation

- A: replace support $(\underline{y}, \overline{y})$ by narrower truncation limits $(\underline{z}, \overline{z})$

Censoring and Truncation

- A: replace support $(\underline{y}, \bar{y})$ by narrower truncation limits $(\underline{z}, \bar{z})$
  - then as full info
- B: Censoring with minimal information

Censoring and Truncation

- A: replace support $(\underline{y}, \overline{y})$ by narrower truncation limits $(\underline{z}, \overline{z})$
  - then as full info
- B: Censoring with minimal information
  - if we do not use the observed point masses at $\underline{z}$ and $\overline{z}$, this could be just treated as case A

Censoring and Truncation

- A: replace support $(\underline{y}, \overline{y})$ by narrower truncation limits $(\underline{z}, \overline{z})$
  - then as full info
- B: Censoring with minimal information
  - if we do not use the observed point masses at $\underline{z}$ and $\overline{z}$, this could be just treated as case A
  - need: n (the full sample size), $\underline{n}$ (#observations equal to $\underline{z}$) and $\overline{n}$ (#observations equal to $\overline{z}$)

Censoring and Truncation

- A: replace support $(\underline{y}, \overline{y})$ by narrower truncation limits $(\underline{z}, \overline{z})$
  - then as full info
- B: Censoring with minimal information
  - if we do not use the observed point masses at $\underline{z}$ and $\overline{z}$, this could be just treated as case A
  - need: n (the full sample size), $\underline{n}$ (#observations equal to $\underline{z}$) and $\overline{n}$ (#observations equal to $\overline{z}$)
- C: Censoring with rich information

Censoring and Truncation

- A: replace support $(\underline{y}, \overline{y})$ by narrower truncation limits $(\underline{z}, \overline{z})$
  - then as full info
- B: Censoring with minimal information
  - if we do not use the observed point masses at $\underline{z}$ and $\overline{z}$, this could be just treated as case A
  - need: n (the full sample size), $\underline{n}$ (#observations equal to $\underline{z}$) and $\overline{n}$ (#observations equal to $\overline{z}$)
- C: Censoring with rich information
  - carry out inference on Lorenz-curve ordinates and some welfare indices

Censoring, Case C

- Need to modify statistics to take account of missing portions

Censoring, Case C

- Need to modify statistics to take account of missing portions
- At the bottom of the distribution:
  - $\hat{c}_{\text{low}} := \frac{1}{n}\sum_{i=1}^{n} y_{(i)}$
  - $\hat{s}_{\text{low}} := \frac{1}{n}\sum_{i=1}^{n} y_{(i)}^2$

Censoring, Case C

- Need to modify statistics to take account of missing portions
- At the bottom of the distribution:
  - $\hat{c}_{\text{low}} := \frac{1}{n}\sum_{i=1}^{n} y_{(i)}$
  - $\hat{s}_{\text{low}} := \frac{1}{n}\sum_{i=1}^{n} y_{(i)}^2$
- At the top of the distribution:
  - $\hat{c}_{\text{high}} := \frac{1}{n}\sum_{n-\overline{n}+1}^{n} y_{(i)}$
  - $\hat{s}_{\text{high}} := \frac{1}{n}\sum_{n-\overline{n}+1}^{n} y_{(i)}^2$

Censoring, Case C

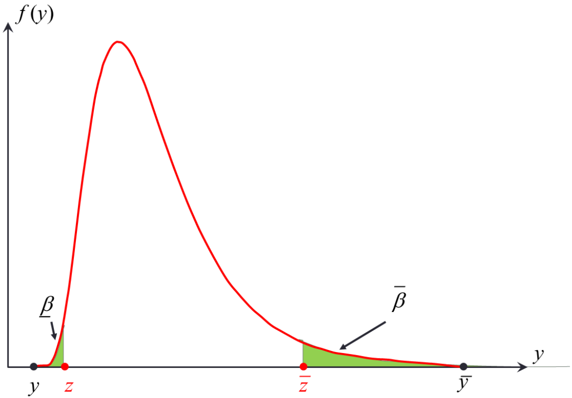
- Need to modify statistics to take account of missing portions
- At the bottom of the distribution:
  - $\hat{c}_{\text{low}} := \frac{1}{n}\sum_{i=1}^{n} y_{(i)}$
  - $\hat{s}_{\text{low}} := \frac{1}{n}\sum_{i=1}^{n} y_{(i)}^2$
- At the top of the distribution:
  - $\hat{c}_{\text{high}} := \frac{1}{n}\sum_{n-\overline{n}+1}^{n} y_{(i)}$
  - $\hat{s}_{\text{high}} := \frac{1}{n}\sum_{n-\overline{n}+1}^{n} y_{(i)}^2$
- Asymptotic covariance:

Censoring, Case C

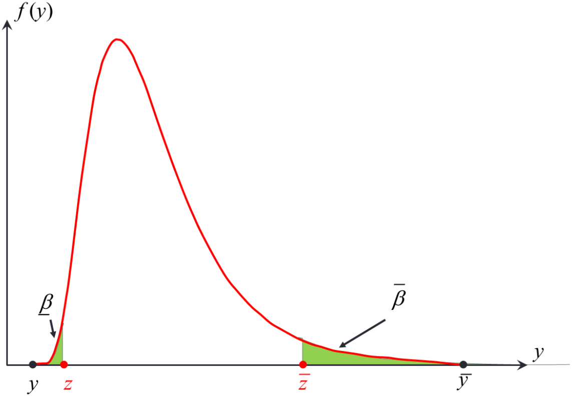
- Need to modify statistics to take account of missing portions
- At the bottom of the distribution:
  - $\hat{c}_{\text{low}} := \frac{1}{n}\sum_{i=1}^{\underline{n}} y_{(i)}$
  - $\hat{s}_{\text{low}} := \frac{1}{n}\sum_{i=1}^{\underline{n}} y_{(i)}^2$
- At the top of the distribution:
  - $\hat{c}_{\text{high}} := \frac{1}{n}\sum_{n-\overline{n}+1}^{n} y_{(i)}$
  - $\hat{s}_{\text{high}} := \frac{1}{n}\sum_{n-\overline{n}+1}^{n} y_{(i)}^2$
- Asymptotic covariance:
  - $\widehat{\omega}_{qq'} := \widehat{s}_q + \left[q\widehat{y}_q - \widehat{c}_q\right]\left[\widehat{y}_{q'} - q'\widehat{y}_{q'} + \widehat{c}_{q'}\right] - \hat{y}_q\widehat{c}_q$
  - $\widehat{c}_q := \widehat{c}_{\text{low}} + \frac{1}{n}\sum_{i=\kappa(n,\underline{\beta})+1}^{\kappa(n,q)} y_{(i)}$
  - $\widehat{s}_q := \hat{s}_{\text{low}} + \frac{1}{n}\sum_{i=\kappa(n,\underline{\beta})+1}^{\kappa(n,q)} y_{(i)}^2$

Trimming

Trimming

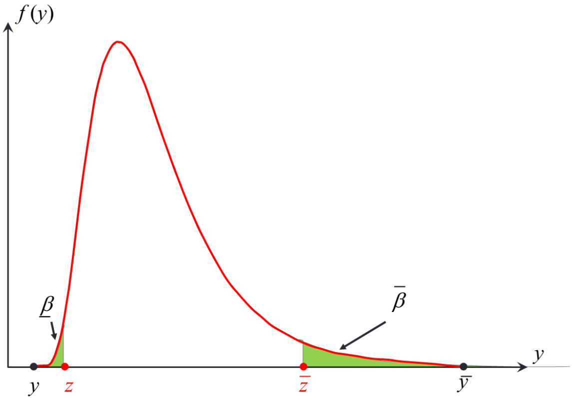


Trimming



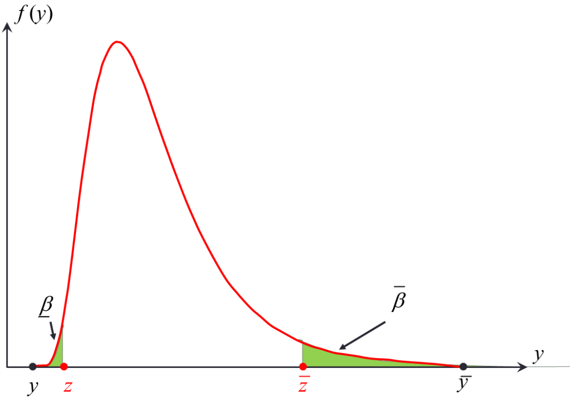
- Fixed proportion of the sample discarded

Trimming



- Fixed proportion of the sample discarded
  - remove outliers for robustness reasons?

Trimming



- Fixed proportion of the sample discarded
  - remove outliers for robustness reasons?
  - proportions $(\underline{\beta}, 1 - \bar{\beta})$ removed from the (bottom, top)
  - $y_{\underline{\beta}}$ and $y_{\bar{\beta}}$ are *random*

Trimming: inference

Trimming: inference

- Inference on *full distribution*, known proportions trimmed

Trimming: inference

- Inference on *full distribution*, known proportions trimmed
  - The *trimmed distribution*

$$\tilde{F}_{\beta}(y) := \begin{cases} 0 & \text{if } y < Q(F,\underline{\beta}) \\ b\left[F(y) - \underline{\beta}\right] & \text{if } Q(F,\underline{\beta}) \le y < Q(F,\overline{\beta}) \\ 1 & \text{if } y \ge Q(F,\overline{\beta}) \end{cases} .$$

  - $b := 1/\left[\bar{\beta} - \underline{\beta}\right]$

Trimming: inference

- Inference on *full distribution*, known proportions trimmed
  - The *trimmed distribution*

$$\tilde{F}_\beta(y) := \begin{cases} 0 & \text{if } y < Q(F,\underline{\beta}) \\ b\left[F(y) - \underline{\beta}\right] & \text{if } Q(F,\underline{\beta}) \le y < Q(F,\overline{\beta}) \\ 1 & \text{if } y \ge Q(F,\overline{\beta}) \end{cases} .$$

  - $b := 1/\left[\bar{\beta} - \underline{\beta}\right]$
- Key statistics:
  - income cumulations $c_{\beta,q} := C(\tilde{F}_\beta; q) = b\int_{y_{\underline{\beta}}}^{y_q} y\,\mathrm{d}F(y)$
  - mean $\mu_\beta := \mu(\tilde{F}_\beta)$
  - $s_{\beta,q} := S(\tilde{F}_\beta; q) := b\int_{y_{\underline{\beta}}}^{y_q} y^2\,\mathrm{d}F(y)$

Trimming: GLC (1)

Trimming: GLC (1)

- Drawing GLC is easy because
  - $C(\tilde{F}_{\beta};q) = b\left[C(F;q) - C(F;\underline{\beta})\right]$

Trimming: GLC (1)

- Drawing GLC is easy because
  - $C(\tilde{F}_\beta; q) = b\left[C(F;q) - C(F;\underline{\beta})\right]$
- For inference on GLC or RLC again use the IF method
  - Need to evaluate $\int IF(z; C(\cdot;q), \tilde{F}_\beta) IF(z; C(\cdot;q'), \tilde{F}_\beta)\, \mathrm{d}F(z)$

Trimming: GLC (1)

- Drawing GLC is easy because
  - $C(\tilde{F}_\beta; q) = b\left[C(F;q) - C(F;\underline{\beta})\right]$
- For inference on GLC or RLC again use the IF method
  - Need to evaluate $\int IF(z; C(\cdot;q), \tilde{F}_\beta) IF(z; C(\cdot;q'), \tilde{F}_\beta)\, \mathrm{d}F(z)$
- $IF(z; C(\cdot;q), \tilde{F}_\beta) =$
  $-c_{\beta,q} + b\left[q y_q - \underline{\beta} y_{\underline{\beta}} + \iota(y_q \geq z)[z - y_q] - \iota(y_{\underline{\beta}} \geq z)[z - y_{\underline{\beta}}]\right]$
  - $= -c_{\beta,q} + b\left[q y_q - \underline{\beta} y_{\underline{\beta}} - \iota(y_q \geq z) y_q + \iota(y_{\underline{\beta}} \geq z) y_{\underline{\beta}}\right] +$
    $b\left[\iota(y_q \geq z) - \iota(y_{\underline{\beta}} \geq z)\right]$

Trimming: GLC (1)

- Drawing GLC is easy because
  - $C(\tilde{F}_\beta; q) = b\left[C(F;q) - C(F;\underline{\beta})\right]$
- For inference on GLC or RLC again use the IF method
  - Need to evaluate $\int IF(z; C(\cdot;q), \tilde{F}_\beta) IF(z; C(\cdot;q'), \tilde{F}_\beta)\, \mathrm{d}F(z)$
- $IF(z; C(\cdot;q), \tilde{F}_\beta) =$
  $-c_{\beta,q} + b\left[qy_q - \underline{\beta}y_{\underline{\beta}} + \iota(y_q \geq z)[z - y_q] - \iota(y_{\underline{\beta}} \geq z)[z - y_{\underline{\beta}}]\right]$
  - $= -c_{\beta,q} + b\left[qy_q - \underline{\beta}y_{\underline{\beta}} - \iota(y_q \geq z)y_q + \iota(y_{\underline{\beta}} \geq z)y_{\underline{\beta}}\right] +$
    $b\left[\iota(y_q \geq z) - \iota(y_{\underline{\beta}} \geq z)\right]$
- So the asymptotic covariance of $\sqrt{n}\hat{c}_{\beta,q}$, $\sqrt{n}\hat{c}_{\beta,q'}$ ($q \leq q'$) is
  - $\varpi_{qq'} = b^2\left[\omega_{qq'} + \omega_{\underline{\beta}\underline{\beta}} - \omega_{\underline{\beta}q} - \omega_{\underline{\beta}q'}\right]$

Trimming: GLC (2)

- To implement we need the sample analogues

Trimming: GLC (2)

- To implement we need the sample analogues
- Sample estimates of cumulations
  - $\hat{\mu}_\beta := \mu(\tilde{F}_\beta^{(n)}) = \frac{b}{n}\sum_{i=1}^{n} y_{(i)}\iota\left(\kappa(n,\underline{\beta})+1 < i < \kappa(n,\overline{\beta})\right)$

Trimming: GLC (2)

- To implement we need the sample analogues
- Sample estimates of cumulations
  - $\hat{\mu}_\beta := \mu(\tilde{F}_\beta^{(n)}) = \frac{b}{n}\sum_{i=1}^n y_{(i)}\iota\left(\kappa(n,\underline{\beta})+1<i<\kappa(n,\overline{\beta})\right)$
- Covariance of $\sqrt{n}\hat{c}_{\beta,q}$, $\sqrt{n}\hat{c}_{\beta,q'}$ $(q \leq q'$) estimated by
  - $\widehat{\varpi}_{q_iq_j} = \left[q_iy_{(i)} - \underline{\beta}y_{(1)} - \sum_{k=1}^i \frac{y_{(k)}}{bn_\beta}\right] \times \left[[1-q_j]\,y_{(j)} - \left[1-\underline{\beta}\right]y_{(1)} + \sum_{k=1}^j \frac{y_{(k)}}{bn_\beta}\right] - \sum_{k=1}^i \frac{y_{(i)}y_{(k)} - y_{(k)}^2}{bn_\beta} + y_{(1)}\left[q_iy_{(i)} - \underline{\beta}y_{(i)} - \sum_{k=1}^i \frac{y_{(i)}}{bn_\beta}\right]$

Trimming: QAD Welfare

Trimming: QAD Welfare

- $W_{\text{QAD}}(\tilde{F}_\beta) = b \int \varphi\left(x, \mu(\tilde{F}_\beta)\right) \,\mathrm{d}F(x)$

Trimming: QAD Welfare

- $W_{\text{QAD}}(\tilde{F}_\beta) = b \int \varphi\left(x, \mu(\tilde{F}_\beta)\right) \mathrm{d}F(x)$
  - $\hat{w}_{\text{QAD},\beta} := W_{\text{QAD}}(\tilde{F}_\beta^{(n)}) := \frac{b}{n} \sum_{i=1}^{n} \varphi\left(y_{(i)}, \hat{\mu}_\beta\right) \iota(\kappa(n, \underline{\beta}) + 1 < i < \kappa(n, \overline{\beta}))$

Trimming: QAD Welfare

- $W_{\mathrm{QAD}}(\tilde{F}_\beta) = b \int \varphi\left(x, \mu(\tilde{F}_\beta)\right) \mathrm{d}F(x)$
  - $\hat{w}_{\mathrm{QAD},\beta} := W_{\mathrm{QAD}}(\tilde{F}_\beta^{(n)}) := \frac{b}{n} \sum_{i=1}^{n} \varphi\left(y_{(i)}, \hat{\mu}_\beta\right) \iota(\kappa(n,\underline{\beta}) + 1 < i < \kappa(n,\overline{\beta}))$
  - $IF(z; W_{\mathrm{QAD}}, \tilde{F}_\beta) = b\varphi\left(\max\left(y_{\underline{\beta}}, \min(z, y_{\overline{\beta}})\right), \mu(\tilde{F}_\beta)\right) - W_{\mathrm{QAD}}(\tilde{F}_\beta) + bIF(z, C(\cdot;\overline{\beta}), \tilde{F}_\beta) \int_{Q(F,\underline{\beta})}^{Q(F,\overline{\beta})} \varphi_\mu\left(x, \mu(\tilde{F}_\beta)\right) \mathrm{d}F(x)$

Trimming: QAD Welfare

- $W_{\text{QAD}}(\tilde{F}_\beta) = b \int \varphi\left(x, \mu(\tilde{F}_\beta)\right) \mathrm{d}F(x)$
    - $\hat{w}_{\text{QAD},\beta} := W_{\text{QAD}}(\tilde{F}_\beta^{(n)}) := \frac{b}{n} \sum_{i=1}^{n} \varphi\left(y_{(i)}, \hat{\mu}_\beta\right) \iota(\kappa(n, \underline{\beta}) + 1 < i < \kappa(n, \overline{\beta}))$
    - $IF(z; W_{\text{QAD}}, \tilde{F}_\beta) = b\varphi\left(\max\left(y_{\underline{\beta}}, \min(z, y_{\overline{\beta}})\right), \mu(\tilde{F}_\beta)\right) - W_{\text{QAD}}(\tilde{F}_\beta) + bIF(z, C(\cdot; \overline{\beta}), \tilde{F}_\beta) \int_{Q(F,\underline{\beta})}^{Q(F,\overline{\beta})} \varphi_\mu\left(x, \mu(\tilde{F}_\beta)\right) \mathrm{d}F(x)$
- Estimate of AV of $\sqrt{n} W_{\text{QAD}}(\tilde{F}_\beta^{(n)})$ found by computing the mean of squares of $IF(z; W_{\text{QAD}}, \tilde{F}_\beta)$, $z = y_i$, $i = 1, \ldots, n$.

Trimming: QAD Welfare

- $W_{\text{QAD}}(\tilde{F}_\beta) = b\int \varphi\left(x, \mu(\tilde{F}_\beta)\right) \mathrm{d}F(x)$
  - $\hat{w}_{\text{QAD},\beta} := W_{\text{QAD}}(\tilde{F}_\beta^{(n)}) := \frac{b}{n}\sum_{i=1}^n \varphi\left(y_{(i)}, \hat{\mu}_\beta\right) \iota(\kappa(n,\underline{\beta})+1 < i < \kappa(n,\overline{\beta}))$
  - $IF(z; W_{\text{QAD}}, \tilde{F}_\beta) = b\varphi\left(\max\left(y_{\underline{\beta}}, \min(z, y_{\overline{\beta}})\right), \mu(\tilde{F}_\beta)\right) - W_{\text{QAD}}(\tilde{F}_\beta) + bIF(z, C(\cdot;\overline{\beta}), \tilde{F}_\beta)\int_{Q(F,\underline{\beta})}^{Q(F,\overline{\beta})} \varphi_\mu\left(x, \mu(\tilde{F}_\beta)\right) \mathrm{d}F(x)$
- Estimate of AV of $\sqrt{n}W_{\text{QAD}}(\tilde{F}_\beta^{(n)})$ found by computing the mean of squares of $IF(z; W_{\text{QAD}}, \tilde{F}_\beta)$, $z = y_i$, $i = 1, \ldots, n$.
  - $F^*_\beta(y) = F(y), \quad Q(F,\underline{\beta}) \le y < Q(F,\overline{\beta})$

Trimming: QAD Welfare

- $W_{\text{QAD}}(\tilde{F}_\beta) = b \int \varphi\left(x, \mu(\tilde{F}_\beta)\right) \mathrm{d}F(x)$
  - $\hat{w}_{\text{QAD},\beta} := W_{\text{QAD}}(\tilde{F}_\beta^{(n)}) := \frac{b}{n} \sum_{i=1}^{n} \varphi\left(y_{(i)}, \hat{\mu}_\beta\right) \iota(\kappa(n, \underline{\beta}) + 1 < i < \kappa(n, \overline{\beta}))$
  - $IF(z; W_{\text{QAD}}, \tilde{F}_\beta) = b\varphi\left(\max\left(y_{\underline{\beta}}, \min(z, y_{\overline{\beta}})\right), \mu(\tilde{F}_\beta)\right) - W_{\text{QAD}}(\tilde{F}_\beta) + bIF(z, C(\cdot; \overline{\beta}), \tilde{F}_\beta) \int_{Q(F,\underline{\beta})}^{Q(F,\overline{\beta})} \varphi_\mu\left(x, \mu(\tilde{F}_\beta)\right) \mathrm{d}F(x)$
- Estimate of AV of $\sqrt{n} W_{\text{QAD}}(\tilde{F}_\beta^{(n)})$ found by computing the mean of squares of $IF(z; W_{\text{QAD}}, \tilde{F}_\beta)$, $z = y_i$, $i = 1, \ldots, n$.
  - $F_\beta^*(y) = F(y), \quad Q(F, \underline{\beta}) \le y < Q(F, \overline{\beta})$
  - The asymptotic variance of $\sqrt{n} W_{\text{QAD}}(\tilde{F}_\beta^{(n)})$
    $b^2 \text{var}(\varphi\left(x, \mu(\tilde{F}_\beta)\right); F_\beta^*) +$
    $2b^3 \text{cov}\left(x, \varphi\left(x, \mu(\tilde{F}_\beta)\right); F_\beta^*\right) \int_{Q(F,\underline{\beta})}^{Q(F,\overline{\beta})} \varphi_\mu\left(x, \mu(\tilde{F}_\beta)\right) \mathrm{d}F(x) +$
    $b^4 \text{var}(x; F_\beta^*) \left[\int_{Q(F,\underline{\beta})}^{Q(F,\overline{\beta})} \varphi_\mu\left(x, \mu(\tilde{F}_\beta)\right) \mathrm{d}F(x)\right]^2$

Trimming: Gini

- $I_{\text{Gini}}(\tilde{F}_\beta) = 1 - 2\int_{\underline{\beta}}^{\overline{\beta}} \frac{C(\tilde{F}_\beta, q)}{C(\tilde{F}_\beta, \overline{\beta})}\,\mathrm{d}q$

Trimming: Gini

- $I_{\text{Gini}}(\tilde{F}_\beta) = 1 - 2\int_{\underline{\beta}}^{\overline{\beta}} \frac{C(\tilde{F}_\beta, q)}{C(\tilde{F}_\beta, \overline{\beta})}\,\mathrm{d}q$
- Asymptotic variance of $\sqrt{n} I_{\text{Gini}}(\tilde{F}_\beta^{(n)})$ is
  - $4b^2 \vartheta_\beta / \mu_\beta^4$
  - $\vartheta_\beta = \mu_\beta^2 \int_{\underline{\beta}}^{\overline{\beta}} \int_{\underline{\beta}}^{q} \varpi_{q'q}\,\mathrm{d}q'\,\mathrm{d}q + \mu_\beta^2 \int_{\underline{\beta}}^{\overline{\beta}} \int_{q}^{\overline{\beta}} \varpi_{qq'}\,\mathrm{d}q\,\mathrm{d}q + \varpi_{\overline{\beta}\overline{\beta}} \left[\int_{\underline{\beta}}^{\overline{\beta}} c_{\beta,q}\,\mathrm{d}q\right]^2 - 2\mu_\beta \int_{\underline{\beta}}^{\overline{\beta}} c_{\beta,q}\,\mathrm{d}q \int_{\underline{\beta}}^{\overline{\beta}} \varpi_{q\overline{\beta}}\,\mathrm{d}q$

Outline

Semi-parametric modelling

- An approach to robustness / incomplete information

Semi-parametric modelling

- An approach to robustness / incomplete information
- Semi -parametric model:
  - apply to proportion $\beta \in \mathbb{Q}$ of upper incomes
  - use EDF for remaining $1 - \beta$ of lower incomes

Semi-parametric modelling

- An approach to robustness / incomplete information
- Semi -parametric model:
  - apply to proportion $\beta \in \mathbb{Q}$ of upper incomes
  - use EDF for remaining $1-\beta$ of lower incomes
- Main issues
  1. What parametric model should be used for the tail?
  2. How should the model be estimated?
  3. How should the proportion β be chosen?
  4. What implications for welfare indices, dominance criteria?

Semi-parametric distribution

Semi-parametric distribution

- Pareto model has two parameters:

Semi-parametric distribution

- Pareto model has two parameters:
  - y_0 determined by quantile $Q(F; 1-\beta)$

Semi-parametric distribution

- Pareto model has two parameters:
  - y_0 determined by quantile $Q(F; 1-\beta)$
  - dispersion parameter α estimated from the data

Semi-parametric distribution

- Pareto model has two parameters:
  - y_0 determined by quantile $Q(F; 1-\beta)$
  - dispersion parameter α estimated from the data
- The *semi-parametric distribution* :

$$\widetilde{F}(y) = \begin{cases} F(y) & y \leq Q(F; 1-\beta) \\ 1-\beta\left(\frac{y}{Q(F;1-\beta)}\right)^{-\alpha} & y > Q(F; 1-\beta) \end{cases}$$

Semi-parametric distribution

- Pareto model has two parameters:
  - y_0 determined by quantile $Q(F; 1-\beta)$
  - dispersion parameter α estimated from the data
- The *semi-parametric distribution* :

$$\widetilde{F}(y) = \begin{cases} F(y) & y \leq Q(F;1-\beta) \\ 1-\beta\left(\frac{y}{Q(F;1-\beta)}\right)^{-\alpha} & y > Q(F;1-\beta) \end{cases}$$

- Density
  - $\widetilde{f}(y;\alpha) = \beta\alpha Q(F;1-\beta)^{\alpha} y^{-\alpha-1}$

Semi-parametric distribution

- Pareto model has two parameters:
  - y_0 determined by quantile $Q(F;1-\beta)$
  - dispersion parameter α estimated from the data
- The *semi-parametric distribution* :

$$\widetilde{F}(y) = \begin{cases} F(y) & y \leq Q(F;1-\beta) \\ 1-\beta\left(\frac{y}{Q(F;1-\beta)}\right)^{-\alpha} & y > Q(F;1-\beta) \end{cases}$$

- Density
  - $\widetilde{f}(y;\alpha) = \beta\alpha Q(F;1-\beta)^{\alpha} y^{-\alpha-1}$
  - $\widetilde{f}(y_{1-\beta};\alpha) = \frac{\beta\alpha}{y_{1-\beta}}$

Parade and Lorenz Curves

- Quantile functional $Q(\widetilde{F};q) =$

$$\begin{cases} Q(F;q) & q \leq 1-\beta \\ Q(F;1-\beta)\left(\frac{1-q}{\beta}\right)^{-1/\hat{\alpha}(\widetilde{F})} & q > 1-\beta \end{cases}$$

Parade and Lorenz Curves

- Quantile functional $Q(\widetilde{F};q) =$

$$\begin{cases} Q(F;q) & q \leq 1-\beta \\ Q(F;1-\beta)\left(\frac{1-q}{\beta}\right)^{-1/\hat{\alpha}(\widetilde{F})} & q > 1-\beta \end{cases}$$

- Cumulative-income functional $C(\widetilde{F};q) =$

$$\begin{cases} \int_{\underline{z}}^{Q(F;q)} y\,\mathrm{d}F(y) & q \leq 1-\beta \\ C(\widetilde{F};q1-\beta) + \beta\frac{\hat{\alpha}(\widetilde{F})}{1-\hat{\alpha}(\widetilde{F})}Q(F;1-\beta) & \\ \qquad \times\left[\left(\frac{1-q}{\beta}\right)^{\frac{\hat{\alpha}(\widetilde{F})-1}{\hat{\alpha}(\widetilde{F})}} - 1\right] & q > 1-\beta \end{cases} .$$

Parade and Lorenz Curves

- Quantile functional $Q(\widetilde{F};q) =$

$$\begin{cases} Q(F;q) & q \leq 1-\beta \\ Q(F;1-\beta)\left(\frac{1-q}{\beta}\right)^{-1/\hat{\alpha}(\widetilde{F})} & q > 1-\beta \end{cases}$$

- Cumulative-income functional $C(\widetilde{F};q) =$

$$\begin{cases} \int_{\underline{z}}^{Q(F;q)} y\,\mathrm{d}F(y) & q \leq 1-\beta \\ C(\widetilde{F};q1-\beta) + \beta\frac{\hat{\alpha}(\widetilde{F})}{1-\hat{\alpha}(\widetilde{F})}Q(F;1-\beta) & \\ \quad\quad \times\left[\left(\frac{1-q}{\beta}\right)^{\frac{\hat{\alpha}(\widetilde{F})-1}{\hat{\alpha}(\widetilde{F})}} - 1\right] & q > 1-\beta \end{cases}.$$

- From this we can derive:

Parade and Lorenz Curves

- Quantile functional $Q(\widetilde{F};q) =$

$$\begin{cases} Q(F;q) & q \leq 1-\beta \\ Q(F;1-\beta)\left(\frac{1-q}{\beta}\right)^{-1/\hat{\alpha}(\widetilde{F})} & q > 1-\beta \end{cases}$$

- Cumulative-income functional $C(\widetilde{F};q) =$

$$\begin{cases} \int_{\underline{z}}^{Q(F;q)} y \, \mathrm{d}F(y) & q \leq 1-\beta \\ C(\widetilde{F};q1-\beta) + \beta \frac{\hat{\alpha}(\widetilde{F})}{1-\hat{\alpha}(\widetilde{F})} Q(F;1-\beta) & \\ \qquad \times \left[\left(\frac{1-q}{\beta}\right)^{\frac{\hat{\alpha}(\widetilde{F})-1}{\hat{\alpha}(\widetilde{F})}} - 1\right] & q > 1-\beta \end{cases} .$$

- From this we can derive:
  - mean $\mu(\widetilde{F}) = C(\widetilde{F};q1-\beta) - \beta Q(F;1-\beta)\frac{\hat{\alpha}(\widetilde{F})}{1-\hat{\alpha}(\widetilde{F})}$

Parade and Lorenz Curves

- Quantile functional $Q(\widetilde{F};q) =$

$$\begin{cases} Q(F;q) & q \leq 1-\beta \\ Q(F;1-\beta)\left(\frac{1-q}{\beta}\right)^{-1/\hat{\alpha}(\widetilde{F})} & q > 1-\beta \end{cases}$$

- Cumulative-income functional $C(\widetilde{F};q) =$

$$\begin{cases} \int_{\underline{z}}^{Q(F;q)} y\,\mathrm{d}F(y) & q \leq 1-\beta \\ C(\widetilde{F};q1-\beta) + \beta\frac{\hat{\alpha}(\widetilde{F})}{1-\hat{\alpha}(\widetilde{F})}Q(F;1-\beta) & \\ \quad \times \left[\left(\frac{1-q}{\beta}\right)^{\frac{\hat{\alpha}(\widetilde{F})-1}{\hat{\alpha}(\widetilde{F})}} - 1\right] & q > 1-\beta \end{cases}.$$

- From this we can derive:
  - mean $\mu(\widetilde{F}) = C(\widetilde{F};q1-\beta) - \beta Q(F;1-\beta)\frac{\hat{\alpha}(\widetilde{F})}{1-\hat{\alpha}(\widetilde{F})}$
  - semi-parametric RLC: graph of $L(\widetilde{F};q) = \frac{C(\widetilde{F};q)}{\mu(\widetilde{F})}$

Lorenz Curves (empirical)

- Cumulative-income functional $C(F^{(n)};q) =$

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{\kappa(n,q)} y_{(i)} & q \leq 1-\beta \\ C(F^{(n)};q1-\beta) + \beta\frac{\hat{\alpha}(\widetilde{F})}{1-\hat{\alpha}(\widetilde{F})}Q(F;1-\beta) & \\ \qquad \times\left[\left(\frac{1-q}{\beta}\right)^{\frac{\hat{\alpha}(\widetilde{F})-1}{\hat{\alpha}(\widetilde{F})}} - 1\right] & q > 1-\beta \end{cases}.$$

- $\kappa(n,q)$ is largest integer no greater than $nq - q + 1$

Lorenz Curves (empirical)

- Cumulative-income functional $C(F^{(n)};q) =$

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{\kappa(n,q)} y_{(i)} & q \leq 1-\beta \\ C(F^{(n)};q1-\beta) + \beta\frac{\hat{\alpha}(\widetilde{F})}{1-\hat{\alpha}(\widetilde{F})}Q(F;1-\beta) & \\ \qquad \times\left[\left(\frac{1-q}{\beta}\right)^{\frac{\hat{\alpha}(\widetilde{F})-1}{\hat{\alpha}(\widetilde{F})}} - 1\right] & q > 1-\beta \end{cases}.$$

- $\kappa(n,q)$ is largest integer no greater than $nq-q+1$
- mean $\mu(F^{(n)}) = C(F^{(n)};q1-\beta) - \beta Q(F;1-\beta)\frac{\hat{\alpha}(\widetilde{F})}{1-\hat{\alpha}(\widetilde{F})}$

Lorenz Curves (empirical)

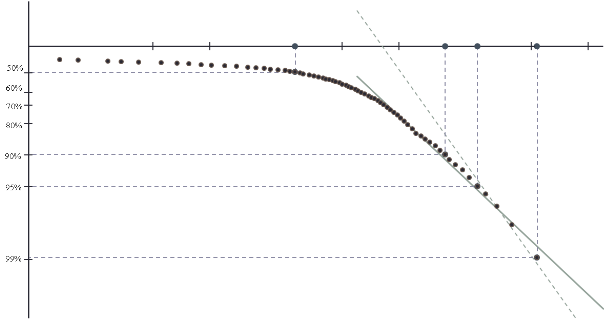
- Cumulative-income functional $C(F^{(n)};q) =$

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{\kappa(n,q)} y_{(i)} & q \le 1-\beta \\ C(F^{(n)};q1-\beta)+\beta\frac{\hat{\alpha}(\widetilde{F})}{1-\hat{\alpha}(\widetilde{F})}Q(F;1-\beta) & \\ \qquad \times\left[\left(\frac{1-q}{\beta}\right)^{\frac{\hat{\alpha}(\widetilde{F})-1}{\hat{\alpha}(\widetilde{F})}}-1\right] & q > 1-\beta \end{cases}.$$

- $\kappa(n,q)$ is largest integer no greater than $nq-q+1$
- mean $\mu(F^{(n)}) = C(F^{(n)};q1-\beta)-\beta Q(F;1-\beta)\frac{\hat{\alpha}(\widetilde{F})}{1-\hat{\alpha}(\widetilde{F})}$
- semi-parametric RLC: graph of $L(F^{(n)};q) = \frac{C(F^{(n)};q)}{\mu(F^{(n)})}$

Semi-parametric: estimation method

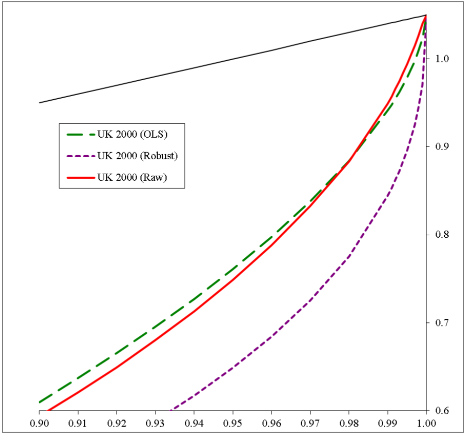
Semi-parametric: estimation method



- Fit Pareto to top ten percent of UK net worth
  - broken line – OLS
  - solid line – OBRE

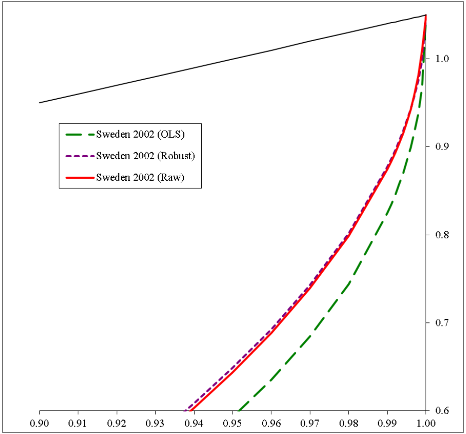
Semi-parametric: UK Lorenz curves

Semi-parametric: UK Lorenz curves



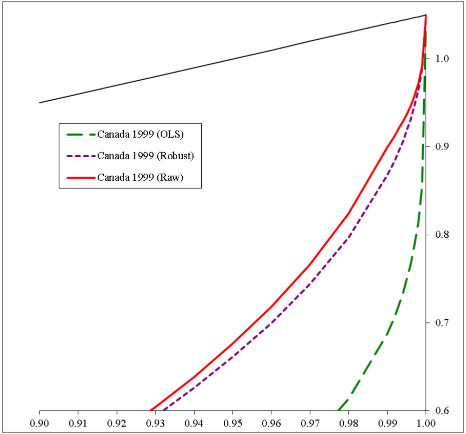
Semi-parametric: Sweden Lorenz curves

Semi-parametric: Sweden Lorenz curves



Semi-parametric: Canada Lorenz curves

Semi-parametric: Canada Lorenz curves



Density estimation, parametric

Density estimation, parametric

- GB2 distribution encompasses all the standard parametric distribution for income distribution

Density estimation, parametric

- GB2 distribution encompasses all the standard parametric distribution for income distribution
- A "good" goodness-of-fit criterion is important:
  - do use the Anderson-Darling statistic, the Cramér-von-Mises statistic
  - do use the Cowell-Davidson-Flachaire measure
  - do not use the χ^2 statistic

Density estimation, semi/non-parametric

- Standard kernel-density methods very sensitive to the choice of the bandwidth

Density estimation, semi/non-parametric

- Standard kernel-density methods very sensitive to the choice of the bandwidth
- Standard approach (the Silverman rule-of-thumb) is known to
  - oversmooth in parts of the distribution where the data are dense
  - undersmooth where the data are sparse

Density estimation, semi/non-parametric

- Standard kernel-density methods very sensitive to the choice of the bandwidth
- Standard approach (the Silverman rule-of-thumb) is known to
  - oversmooth in parts of the distribution where the data are dense
  - undersmooth where the data are sparse
- Standard approach may not be suitable for income distributions
  - typically heavy-tailed

Density estimation, semi/non-parametric

- Standard kernel-density methods very sensitive to the choice of the bandwidth
- Standard approach (the Silverman rule-of-thumb) is known to
  - oversmooth in parts of the distribution where the data are dense
  - undersmooth where the data are sparse
- Standard approach may not be suitable for income distributions
  - typically heavy-tailed
- More appropriate method
  - adaptive kernel
  - mixture model

Welfare measures

- A global approach to the derivation of variance expressions
  - all inequality measures
  - all poverty measures
  - ordinates of Lorenz curves etc

Welfare measures

- A global approach to the derivation of variance expressions
  - all inequality measures
  - all poverty measures
  - ordinates of Lorenz curves etc
- Method uses the Influence Function to provide a shortcut to the formulas

Welfare measures

- A global approach to the derivation of variance expressions
  - all inequality measures
  - all poverty measures
  - ordinates of Lorenz curves etc
- Method uses the Influence Function to provide a shortcut to the formulas
- Necessary to analyse the tails
  - plot of Hill estimators
  - use appropriate methods with heavy-tailed distributions

Distributional comparisons

- Adopt a unified simple approach
  - apply to the variance and covariance formulas
  - makes use of the Influence Function
  - just as with welfare measures

Distributional comparisons

- Adopt a unified simple approach
  - apply to the variance and covariance formulas
  - makes use of the Influence Function
  - just as with welfare measures
- A plot of Lorenz curve differences can provide useful information
  - even where Lorenz curves cross

Distributional comparisons

- Adopt a unified simple approach
  - apply to the variance and covariance formulas
  - makes use of the Influence Function
  - just as with welfare measures
- A plot of Lorenz curve differences can provide useful information
  - even where Lorenz curves cross

Data problems

- Careful modelling is essential to understand what can be done
  - in the case of possible data-contamination
  - in the case of incomplete data

Data problems

- Careful modelling is essential to understand what can be done
  - in the case of possible data-contamination
  - in the case of incomplete data
- Again Influence Function is a valuable tool

Data problems

- Careful modelling is essential to understand what can be done
  - in the case of possible data-contamination
  - in the case of incomplete data
- Again Influence Function is a valuable tool
- Try to"patch" an empirical distribution with a parametric model?
  - useful for the upper tail

Data problems

- Careful modelling is essential to understand what can be done
  - in the case of possible data-contamination
  - in the case of incomplete data
- Again Influence Function is a valuable tool
- Try to"patch" an empirical distribution with a parametric model?
  - useful for the upper tail
- Special attention to the way the parameters of the model are to be estimated