Distributional Change

Winter School, Canazei

Frank A. Cowell

STICERD, London School of Economics

January 2010

The Setting

- *Meaning and motivation*
- *Information and income distribution*
- *Information and entropy*
  - *uncertainty and income distribution*
  - *entropy and inequality*
- *Entropy: "dynamic" aspects*

Motivation

- What do we mean by "distributional change"?
  - welfare analysis mainly about single-distribution comparisons
  - examples: inequality, poverty. polarisation
  - distributional change concerns a class of two-distributional problems
- Why two-distribution problems?
  - mobility
  - effects of taxes and benefits?
  - other economic applications involving reranking
- Is there a unifying approach?
  - welfare economics?
  - statistical tools?
  - a general informational approach?
- Are standard tools appropriate?

This talk

- Informational analysis
  - Background in information theory
  - Information theory extensions
- Connections to income-distribution analysis
  - measurement of mobility
  - goodness-of-fit measures and economics
- Axiomatisation
- Application
  - mobility indices
  - goodness-of-fit criteria
  - evaluation of some standard tools

Literature: Key references

- Cowell, F. A. (1985). Measures of distributional change: An axiomatic approach. *Review of Economic Studies* **52**, 135–151.
- Cowell, F. A., E. Flachaire, and S. Bandyopadhyay (2009). Goodness-of-fit: An economic approach. Distributional Analysis Discussion Paper 101, STICERD, London School of Economics, London WC2A 2AE.
- Fields, G. S. and E. A. Ok (1999). Measuring movement of incomes. *Economica* **66**, 455–472.
- Van Kerm, P. (2004). What lies behind income mobility? Reranking and distributional change in Belgium, Western Germany and the USA. *Economica* **71**, 223–239.

Modelling information

- Entropy is an aggregation of information about a system:
  - the "degree of disorder"
- Consider a discrete set of events $\Theta := \{\theta_1, ..., \theta_n\}$ with probabilities $\{p_1, ..., p_n\}$
- An event $\theta_i \in \Theta$ occurs
  - model information content of this event as $h : \Theta \to \mathbb{R}$
- What properties for the function h?

Valuing information

- The more unlikely is θ_i the more valuable is the information that θ_i occurred
  - $p_i < p_j$ implies $h(p_i) > h(p_j)$
- For independent compound events value of information is additive
  - $h(p_i p_j) = h(p_i) + h(p_j)$
- Additive property is a log Cauchy equation
  - solution $h(p) = C \log(p)$
  - where C is a constant
- Decreasingness property implies C is negative

Aggregating information

- Aggregate information by taking expectation
- Shannon entropy measure for discrete distribution:
  - $-\sum_{i=1}^{n} p_i \log(p_i)$
  - max value is $-\sum_{i=1}^{n} \frac{1}{n} \log\left(\frac{1}{n}\right) = \log(n)$
- But why use this precise formulation?

Entropy: role of axioms

- Role of some axioms is clear
  - decreasingness
  - additivity of independent compound events
- But there is a “hidden” assumption
  - why additivity in aggregation?
  - implied by expectational approach
- Perhaps a more general approach would be worth while
  - to generalise additivity of independent compound events
  - to motivate additive aggregation

Entropy: a general approach

- Model entropy directly as $H^n : \Delta^n \to \mathbb{R}$ and assume:

1. *Continuity*
2. *Symmetry*: $H^n(p_1, p_2, p_3, ...) = H^n(p_2, p_1, p_3, ...) = ...$
3. *Grouping axiom*: $0 < m < n$ and $p := \sum_{j=1}^{m} p_i$:

$$\begin{aligned} H^n(p_1, ..., p_n) &= pH^m\left(\frac{p_1}{p}, ..., \frac{p_m}{p}\right) \\ &\quad + [1-p]\, H^{n-m}\left(\frac{p_{m+1}}{1-p}, ..., \frac{p_n}{1-p}\right) \\ &\quad + H^2(p, 1-p) \end{aligned}$$

4. 0-*irrelevance*: $H^{n+1}(p_1, ..., p_n, 0) = H^n(p_1, ..., p_n)$

- Then you get Shannon entropy

Entropy: extensions (1)

- Extension to continuous distributions
  - probability density f over event space Θ
  - information function $h(f(\theta))$ where h is decreasing
- Entropy as an aggregation of information
  - $H(f) := \mathbf{E}h(f(\theta))$
- Shannon entropy measure: $H(f) = -\int_\Theta \log f(\theta)\, f(\theta) d\theta$
  - $h(f) = -\log f(\theta)$

Entropy: extensions (2)

- A number of information-theoretic arguments for a more general approach
- Focus on the nature of the function h rather than expectational aggregation
- Generalise $h(f) = -\log(f)$ to become $h(f) = \frac{1}{\alpha-1}\left[1 - f^{\alpha-1}\right]$
- Get the α-class entropy $H_\alpha(f) = \frac{1}{\alpha-1}\left[1 - E(f(\theta)^{\alpha-1})\right]$

Entropy and inequality (1)

- Take $x \in X \subseteq \mathbb{R}_+$ where x can be thought of as "income"
- If x has cdf F then income share function is $s(q) := \frac{F^{-1}(q)}{\int_0^1 F^{-1}(t)dt} = \frac{x}{\mu}$
  - population normalised to 1
- Use this to get entropy-based inequality measures
  - apply entropy concept to $s(\cdot)$ rather than $f(\cdot)$
- Theil inequality index $I_1 := \int_0^\infty \frac{x}{\mu} \log\left(\frac{x}{\mu}\right) dF(x)$
  - where $I_1 = -H(s)$

Entropy and inequality (2)

- Generalised entropy index
  $I_\alpha = \int_0^\infty \frac{1}{\alpha(\alpha-1)} \left[\left[\frac{x}{\mu} \right]^\alpha - 1 \right] dF(x)$
  - where $I_\alpha = -\alpha^{-1} H_\alpha(s)$
- Important special cases
  - Theil's second index $I_0 = -\int_0^\infty \log\left(\frac{x}{\mu}\right) dF(x)$
  - also known as MLD index $\int_0^\infty [\log(\mu) - \log(x)]\, dF(x)$
  - Atkinson indices
- Social values emerge implicitly
  - choice of sensitivity index α
  - $\alpha = 1 - \epsilon$ (for $\alpha < 1$) gives Atkinson inequality aversion

Relative entropy (1)

- Use entropy to characterise changes in distributions
  - *relative* entropy
  - (equivalently) *divergence* entropy
- Divergence of f_2 with respect to f_1, is given by $H(f_1,f_2) = \int_\Theta h\left(\frac{f_1}{f_2}\right) f_1 d\theta$
  - "relative entropy": expected information content in f_2 with respect to f_1
  - get standard entropy index as a special case
- Corresponding to α-class entropy get a class of divergence measures:
  - $H_\alpha(f_1,f_2) = \frac{1}{\alpha-1}\int_\Theta \left[1 - f_1\left[\frac{f_1}{f_2}\right]^{\alpha-1}\right] d\theta$

Relative entropy (2)

- Switch from probabilities to income shares:
  - $s_1(q) = \frac{F_1^{-1}(q)}{\int_0^1 F_1^{-1}(t)dt} = \frac{x}{\mu_1}$ and $s_2(q) = \frac{F_2^{-1}(q)}{\int_0^1 F_2^{-1}(t)dt} = \frac{y}{\mu_2}$
- We can also apply relative entropy here:
  - $H_1(s_1, s_2) = -\int_0^1 s_1(q) \log\left(\frac{s_2(q)}{s_1(q)}\right) dq$
- Raises a number of issues:
  - how to interpret?
  - generalisation?
  - axiomatisation?

Literature: Information theory

- Kullback, S. and R. A. Leibler (1951). On information and sufficiency. *Annals of Mathematical Statistics* **22**, 79–86.
- Havrda, J. and F. Charvat (1967). Quantification method in classification processes: concept of structural -entropy. *Kybernetica* **3**, 30–35.
- Renyi, G. (1961). On measures of entropy and information. In Proceedings of the Fourth Berkeley Symposium on Statistics, Volume 1: Probability, pp. 547.561. University of California Press.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal* **106**, 379–423 and 623–656.
- Theil, H. (1967). Economics and Information Theory, Chapter 4, pp. 91–134. Amsterdam: North Holland.

Dynamic aspect: Distributional change

- Revisit analogy between entropy (H) and inequality (I)
- Divergence entropy has a counterpart: measure of distributional change
- Distributional change measure:

$$J_\alpha(\mathbf{x}, \mathbf{y}) := \frac{1}{n\alpha(\alpha-1)} \sum_{i=1}^{n} \left[\left[\frac{x_i}{\mu_1}\right]^{\alpha} \left[\frac{y_i}{\mu_2}\right]^{1-\alpha} - 1 \right]$$

  - $J_\alpha(\mathbf{x}, \mathbf{y}) = -\alpha^{-1} H_\alpha(s_1, s_2)$
  - captures the average distance of an income distribution s_1 from a reference distribution s_2.
- $J_\alpha(\mathbf{x}, \mathbf{y})$ is an aggregate measure of *discrepancy* between two distributions
  - use this to build analytical tools

Axiomatics

- *Basics*
- *Representation of problem*
- *The axioms*
  - *fundamental structure*
  - *income scaling*
- *Characterisation theorems*
- *The index*

Axiomatics: basics

- Purpose:
  - to give meaning to the distributional change problem
  - to avoid concealed arbitrariness
- Principles:
  - parsimony: not impose too much mathematical structure
  - consistency with axiomatisation of other economic problems
- Precedents:
  - inequality
  - welfare
  - poverty
  - mobility

Representation of problem

- $\mathbf{z} := (z_1, z_2, ..., z_n)$
  - z_i is the ordered pair (x_i, y_i)
- Work with vector of discrepancies $(D(z_1), ..., D(z_n))$
  - discrepancy function $D : Z \to \mathbb{R}$
  - $D(z_i)$ strictly increasing in $|x_i - y_i|$
- Two-step approach
  1. characterise a weak ordering $\succeq$ on Z^n, the space of $\mathbf{z}$
  2. use the function representing $\succeq$ to generate index J.

Basic axioms

- **Continuity:** $\succeq$ is continuous on Z^n
- **Monotonicity:** if $\mathbf{z}, \mathbf{z}' \in Z^n$ differ only in their ith component then $D(x_i, y_i) < D(x_i', y_i') \Longleftrightarrow \mathbf{z} \succ \mathbf{z}'$
- **Symmetry:** If $\mathbf{z}'$ is obtained by permuting the components of $\mathbf{z}$: $\mathbf{z} \sim \mathbf{z}'$.
  - we can impose a simultaneous ordering on the x and y components of $\mathbf{z}$
- **Independence:** If $\mathbf{z} \sim \mathbf{z}'$ and $z_i = z_i'$ for some i then $\mathbf{z}(\zeta, i) \sim \mathbf{z}'(\zeta, i)$ for all $\zeta \in [z_{i-1}, z_{i+1}] \cap [z_{i-1}', z_{i+1}']$
  - $\mathbf{z}(\zeta, i)$ means "replace ith component of $\mathbf{z}$ by ζ"
- **Perfect local fit:** Suppose $x_i = y_i$, $x_j = y_j$, $x_i' = x_i + \delta$, $y_i' = y_i + \delta$, $x_j' = x_j - \delta$, $y_j' = y_j - \delta$ and, for all $k \neq i, j$, $x_k' = x_k$, $y_k' = y_k$. Then $\mathbf{z} \sim \mathbf{z}'$

First theorem

- **Theorem:** Given basic axioms $\succeq$ is representable by $\sum_{i=1}^{n} \phi_i (z_i)$
  1. ϕ_i is continuous and strictly decreasing in $|x_i - y_i|$
  2. $\phi_i (x, x) = a_i + b_i x$
- An additivity result
- We can evaluate distributional change focusing on one income-position at a time

Second theorem

- We need one more axiom
- **Income scale irrelevance:** If $\mathbf{z} \sim \mathbf{z}'$ then $t\mathbf{z} \sim t\mathbf{z}'$for all $t > 0$
- **Theorem:** Given conditions of first theorem and scale irrelevance $\succeq$ is representable by $\phi\left(\sum_{i=1}^{n} x_i h_i\left(\frac{x_i}{y_i}\right)\right)$
- The function h_i is essentially arbitrary
  - we need to impose more structure
  - do this in step 2

Income discrepancy and distributional change

- How to compare (x, y) discrepancies in different parts of the income distribution
- From Theorem 2 comparisons in terms of proportional differences: discrepancy should be $D(z_i) = \max\left(\frac{x_i}{y_i}, \frac{y_i}{x_i}\right)$
- **Discrepancy scale irrelevance:** Suppose $\mathbf{z}_0 \sim \mathbf{z}_0'$. Then $\mathbf{z} \sim \mathbf{z}'$ for all $\mathbf{z}, \mathbf{z}'$ such that $D(\mathbf{z}) = tD(\mathbf{z}_0)$ and $D(\mathbf{z}') = tD(\mathbf{z}_0')$
  - suppose vectors $\mathbf{z}_0$ and $\mathbf{z}_0'$ are equivalent under $\succeq$
  - rescale all discrepancies in $\mathbf{z}_0$ and $\mathbf{z}_0'$ by the same factor t
  - resulting pair of vectors $\mathbf{z}$ and $\mathbf{z}'$ will also be equivalent

The index

- **Theorem:** Given discrepancy scale irrelevance $\succeq$ is representable by $\phi\left(\sum_{i=1}^{n} x_i^{\alpha} y_i^{1-\alpha}\right)$
  - $\alpha \neq 1$ is a constant

- Use the "natural" cardinalisation $\sum_{i=1}^{n} x_i^{\alpha} y_i^{1-\alpha}$
- Normalise with reference to case where $x_i = \mu_1$ and $y_i = \mu_2$ for all i
  - observed and modelled distribution exhibit complete equality

- This gives $J_\alpha(\mathbf{x}, \mathbf{y}) := \frac{1}{n\alpha(\alpha-1)} \sum_{i=1}^{n} \left[\left[\frac{x_i}{\mu_1}\right]^{\alpha} \left[\frac{y_i}{\mu_2}\right]^{1-\alpha} - 1 \right]$

Application: mobility

- *Mobility concepts*
- *Mobility modelling*
- *Mobility measures*
- *Distance and mobility: examples*

Mobility concepts

- Variety of approaches
  - ad hoc classification?
  - focus on information theory?
- How does mobility use information?
  - use a pre-grouped scheme?
  - use individual information in relation to others?
  - use distance concept?
- Distance concept
  - not only concerned with how many people move
  - also we want to know how far
  - some of this lost in the transition-matrix approach

Mobility modelling

- Basic information is the temporal pair $z_i = (x_{i,t-1}, x_{i,t})$
- Bivariate distribution
  - distribution function $F(z) = F(x_{t-1}, x_t)$
  - marginal distributions F_{t-1} and F_t give income distribution in each period
- Time-aggregated income
  - derived from z_i using weights w_{t-1}, w_t
  - $\bar{x}_i := w_{t-1}x_{t-1} + w_t x_t$
  - Distribution $\bar{F}$ derived from F

Mobility measures in practice

- Stability indices: $1 - \frac{I(F_{\mathbf{w}})}{w_{t-1}I(F_{t-1})+w_tI(F_t)}$
- Hart (1976): $1 - r(\log x_{t-1}, \log x_t)$
  -
    - where r is the correlation coefficient
- King (1983): $1 - \left[\frac{\int\int\left(x_t e^{\gamma r(F,z)}\right)^k dF(z)}{\mu_k(F_t)}\right]^{\frac{1}{k}}$
  - $k \leq 1,\ k \neq 0,\ \gamma \geq 0$
  - where $r(F;z)$ is a rank indicator: $\mu_1(F_t)^{-1}\left|x_t - Q(F_t;F_{t-1}(x_{t-1}))\right|$
  - $Q(G;q) := \inf\{x : G(x) \geq q\}$
- Fields-Ok (1999): $c\int\int\left|\log x_{t-1} - \log x_t\right| dF(x_{t-1}, x_t)$

Mobility measures: distance

- $M(F) := \phi\left(\int D(z)\, dF(z),\ \boldsymbol{\mu}(F)\right)$
- the function $D : X \times X \to \mathbb{R}$ incorporates the distance concept
- If D is homothetic, the measure takes the form:
  - $\frac{1}{\alpha^2-\alpha}\int\int \left[\left[\frac{x_{t-1}}{\mu(F_{t-1})}\right]^{1-\alpha} \left[\frac{x_t}{\mu(F_t)}\right]^{\alpha} - 1\right] dF(x_{t-1}, x_t)$
  - α is a sensitivity parameter

Mobility: example

- Van Kerm's comparison of mobility in Europe and USA
- Uses trimmed panel data for each case
- Compares 1985, 1997

	Belgium	*Germany*	*USA*
Shorrocks (1978)	0.150	**0.161**	0.137
Hart (1976)	0.584	**0.630**	0.544
King (1983)	0.263	0.300	**0.375**
Fields–Ok (1999)	0.335	0.392	**0.523**
Fields–Ok (1996)	0.37	0.399	**0.534**

Literature: Mobility

- Fields, G. S. and E. A. Ok (1996). The meaning and measurement of income mobility. *Journal of Economic Theory* **71**, 349–377.
- Fields, G. S. and E. A. Ok (1999). The measurement of income mobility: An Introduction to the literature. In J. Silber (ed.), *Handbook of Income Inequality Measurement*. Dewenter: Kluwer.
- King, M. A. (1983). An index of inequality: with applications to horizontal equity and social mobility. *Econometrica* **51**, 99-116.
- Shorrocks, A. F.. Income inequality and income mobility. Journal of Economic Theory, 46, 376–93.
- Van de gaer, D., E. Schokkaert, and M. Martinez (2001). Three meanings of intergenerational mobility. *Economica* **68**, 519–537.

Application: Goodness-of-Fit

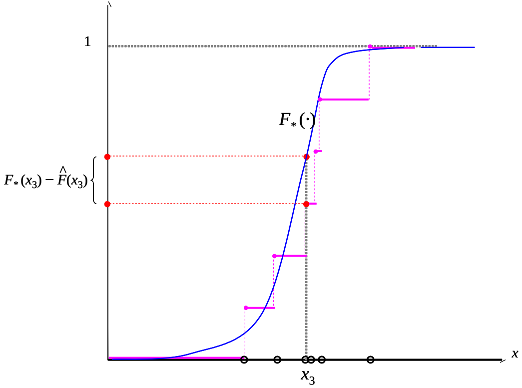
- *GoF: general approach*
- *Evaluation of fit*
  - *The EDF approach*
  - *Quantile approach*
- *Aggregating discrepancy*
- *Implementation*

EDF and Model

- GoF problem requires representation of the facts and the model used to represent them
- x_i: sample observations
  - $i = 1, ..., n$
  - x_i is a scalar
- **E**mpirical **D**istribution **F**unction
  - $\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \iota(x_i \leq x)$
  - $\iota(S)$ means "statement S is true"
- Modelled distribution $F_*(x_i)$
  - could be continuous or discrete

The EDF approach

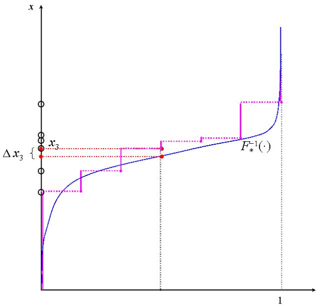
- What underlies the standard statistical approach
- Plot $\hat{F}(x)$ and $F_*(x)$ against x



- For each x_i evaluate distance between F values

Quantile approach

- A kind of "dual" approach
- Compute the quantiles $F_*^{-1}(q)$ where $q = \frac{i}{n+1}$
- Transpose previous diagram: plot quantiles against q



- For each q evaluate distance between incomes on vertical axis

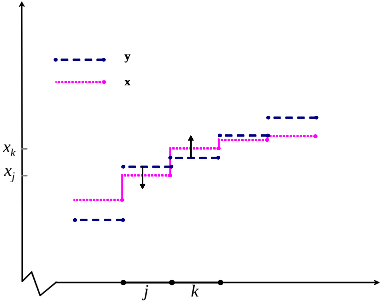
Aggregating discrepancy (1)

- Suppose distributions are discrete point masses
  - one observes $\mathbf{x}$
  - proposed distribution is $\mathbf{y} = \mathbf{x} + \Delta\mathbf{x}$
- Consider three methods of evaluating overall discrepancy

1. **Welfare loss:** $W(\mathbf{y}) - W(\mathbf{x}) \simeq \sum_i \frac{\partial W(\mathbf{x})}{\partial x_i} \Delta x_i$
   - if W is ordinal this is not a well-defined loss function
   - also, can find $\Delta\mathbf{x} \neq 0$ such that expression is zero
2. **Inequality change:** $I(\mathbf{y}) - I(\mathbf{x}) \simeq \sum_i \frac{\partial I(\mathbf{x})}{\partial x_i} \Delta x_i$
   - same basic objections as for welfare loss
3. **Distributional change:** examine more closely $\longrightarrow$

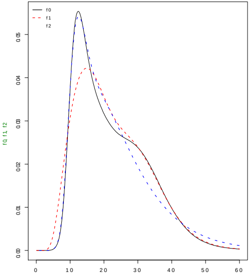
Aggregating discrepancy (2)

- Consider a change in **y**: $y'_k = y_k + \delta, y'_j = y_j - \delta$
  - does the change $\mathbf{y} \to \mathbf{y}'$ move one closer to x?
- Ineq difference: up/down as $y_i \gtrless y_j$, *irrespective of* **x**
- Distributional change measure: up/down as $\frac{y_k}{y_j} \gtrless \frac{x_k}{x_j}$



- Use this to formalise (1) aggregate discrepancy, (2) GoF

Simulation



GE inequality

	f_0	f_1	f_2
I_0	0.10440	0.11389	0.12064
I_1	0.10135	0.10649	0.12177

- f_0, f_1, f_2 each formed from a mixture of three lognormals
- f_1, f_0 similar in high incomes; f_2 , f_0 similar in low incomes
- In inequality terms f_1 is "closer" than f_2 to the reference distribution f_0

Results for traditional GoF criteria

- Simulate f_0, f_1 and f_2 using 10,000
- Compute Chi-squared criterion (χ^2)
- Also Cramér-von Mises (ω^2)
-

	f_1	f_2
χ^2	0.058679	0.048541
ω^2	3.556511	2.421263

- f_2 is "closer" than f_1 to the reference distribution f_0!

Results for the J index

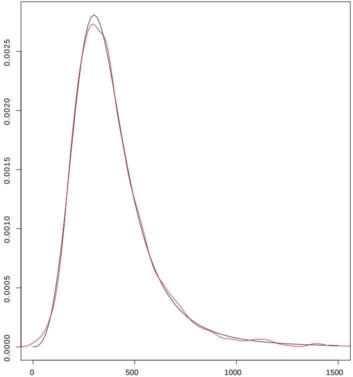
- Now compute $J_\alpha \times 10^2$ for a variety of α-values
-

α	f_1	f_2
−1.0	0.079	0.191
−0.5	0.076	0.195
0.0	0.0742	0.1989
0.5	0.0720	0.2028
1.0	0.0699	0.2070

- f_1 is “closer” than f_2 to the reference distribution f_0 (as with inequality)
- The higher is α, the closer is the approximation of f_1 to f_0 and the worse is that of f_2

Application to UK income data

- HBAI data. n: 3858, mean: 398.28, sd: 253.75
- Fit Singh-Maddala $F_{\text{SM}}(y; a, b, c) = 1 - \frac{1}{\left[1+ay^b\right]^c}$
- MLE $\left(\hat{a}, \hat{b}, \hat{c}\right) = \left(5.75554E^{-10},\ 3.6303,\ 1.0106\right)$
- $\hat{F}$ (in red) and $F_{\text{SM}}(y; \hat{a}, \hat{b}, \hat{c})$



Application: results for traditional GoF measures

		p (%)
χ^2	54.4417	32.33
ω^2	0.04766	29.43

- p values computed using bootstrap
- Singh-Maddala distribution is satisfactory (high p)
- Applies for both traditional GoF criteria, χ^2 and ω^2

Application: results for J

α	$J_\alpha \times 10^2$	p (%)	α	$J_\alpha \times 10^2$	p (%)
-5	2.0313	0.00	0.2	0.1288	5.41
-2	0.1480	1.90	0.5	0.1312	6.01
-1	0.1276	3.80	0.7	0.1332	7.21
-0.7	0.1263	4.00	1	0.1366	6.71
-0.5	0.1261	5.41	2	0.1519	8.31
-0.2	0.1267	5.11	5	0.2394	10.01
0	0.1276	5.31			

- J_α criterion reveals a richer story
- p-values rise with α
  - accept F_{SM} as suitable for $\hat{F}$ if J assigns higher weight to discrepancies at high incomes
  - but for a "bottom-sensitive" GoF criterion ($\alpha < 1$) F_{SM} regarded as unsatisfactory

Conclusions

- What type of mobility index?
  - borrowed from stats?
  - borrowed from inequality?
  - distance approach
- Why do economists want to use GoF criteria?
  - evaluate suitability of a statistical models
  - want a criterion based on economic principles
- J_α indices form a *class* of GoF criteria
  - calibrate to suit the nature of the economic problem under consideration
  - in which part of the distribution do you want the GoF criterion to be sensitive?
- The choice of a fit criterion really matters
  - off-the-shelf tools can be misleading
  - J_α answers accord with common sense
  - α crucial to understanding whether model "fits"