Summer Workshop on Inequality and Poverty Measurement:

2021 LIS Virtual Summer Workshop

Distribution regression & Inequality and poverty decomposition methods

Philippe Van Kerm

e-Belval
July 6, 2021

Faculty of Language and Literature, Humanities, Arts and Education

LISER
Luxembourg Institute of Socio-Economic Research

uni.lu
Université du Luxembourg

Material

https://dropit.uni.lu/invitations?share=4ecbb458fa62c5086dcc

Quick poll and Q&A

What are you interested in? What would you like covered?

`http://www.sli.do/`

Event code: #874114

Or go directly to `https://app.sli.do/event/mzlxfin9`

Who am I?

Economist, Professor of Social Inequality and Social Policy Analysis at UL and LISER

Research on various dimensions of social inequalities and the econometrics of income distribution

philippe.vankerm@liser.lu —@uni.lu

https://ideas.repec.org/e/pva19.html

http://prophil.vankerm.net

http://www.liser.lu

http://wwwen.uni.lu

Course objectives

1. Review practitioners' toolbox for income distribution analysis
2. Focus on tools to help "explain" change over time or differences across populations
   » some well-known tools (index decomposition methods)
   » some less well-known tools ('distribution regression' approaches)
3. Implementation in LIS ...
4. ... with Stata

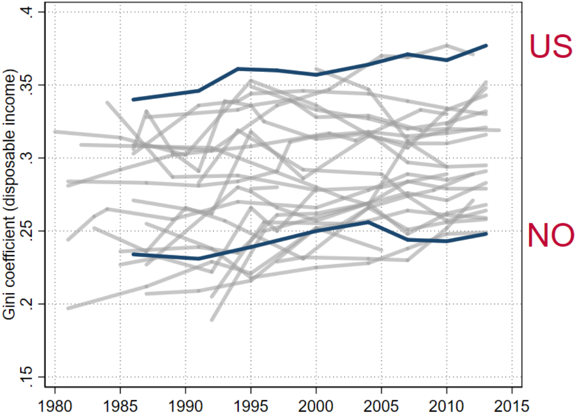
"Learn how to fish" approach:
cover methods and tools—"how?" rather than "what?"

Analysis of distributional change

- Large literature on income distribution *trends* (over time): e.g., *what drives increases in inequality in country X?* (e.g., Hyslop and Maré, 2005, Daly and Valletta, 2006, Fiorio, 2011, Biewen and Juhasz, 2012, Larrimore, 2014, Belfield et al., 2017)
- Fairly large literature on *differences* in income distributions aross groups: e.g., by gender or ethnicity (e.g., Butcher and DiNardo, 2002, Arulampalam et al., 2007)
- Smaller literature on cross-national *differences* in income distributions: *why is there more inequality in country A than in country B?* (e.g., Bourguignon et al., 2008, Sologon et al., 2021)

Gini coefficients in rich countries since 1980 (LIS)

(disposable income)



Source: LIS Key Figures -- http://www.lisdatacenter.org/

What is driving distribution differences?

The smoking guns

- Demography: population composition (ageing, immigration), household formation (declining hh size, fertility, assortative mating)
- Employment, human capital and labour market structure (female LFP, industrial change)
- LM returns and wages, self-employment
- Non-labour (market) incomes (capital income)
- Taxes and benefit policies

What is driving distribution differences?

Deeper layers

- Preferences for redistribution?
- Social norms?
- Institutions?
- Economic structures?
- Path-dependence and adaptation?
- ...

Two main approaches: the ‘classics’ and the ‘moderns’

The ‘classics’

- exploit decomposition properties of inequality or poverty measures
- by subgroup: population partition and contribution of ‘between groups’ and ‘within groups’ inequality and population shares

$$I = \Phi\left(\left\{I^k, s_k, \mu_k\right\}_{k=1}^{K}; I^B\right)$$

- by income source: contribution of ‘source inequality’, size of source and correlation among sources

$$I = \phi\left(\left\{I^k, s_k, \rho_k\right\}_{k=1}^{K}\right)$$

- nicely additive only for particular measures
- contribution to change over time in those indicators is easily assessed from there (see, e.g. Mookherjee and Shorrocks, 1982, Jenkins, 1996, Belfield et al., 2017)
- difficult to handle multiple explanatory factors

Two main approaches: the ‘classics’ and the ‘moderns’

The ‘moderns’

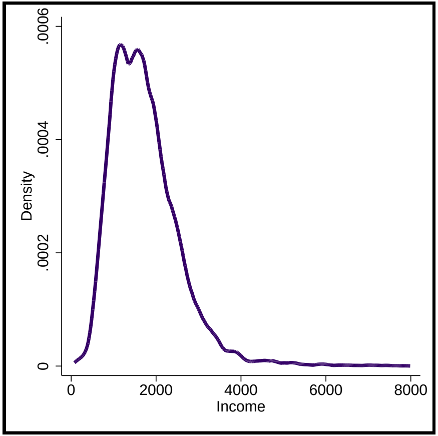
- “generalized Oaxaca-Blinder” decompositions of distribution difference
- based on a (semi-)parametric statistical representation of the distributions of interest
  - reweighting techniques, distribution regression, and some ad hoc simulations
  - not of particular functionals thereof—not ‘index specific’
- construction of ‘what if’/simulated/counterfactual distributions
  - “what if such or such factor had not changed?”
- combination of factors related to differences in sources and population subgroups

Basic distribution analysis toolkit: Quick reminder

Decomposition approaches

Modelling distributions: reweighting and distribution regression methods

Kernel density estimates

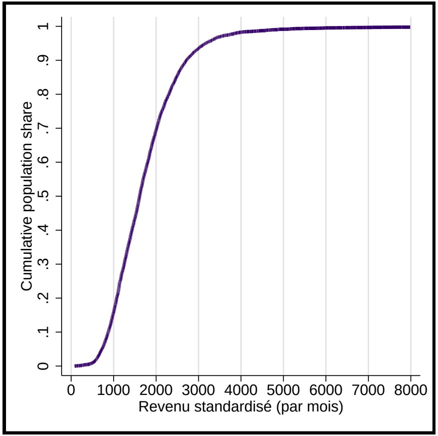


A ‘continuous’ version of the histogram: the density function

For kernel density estimation, can think of histogram with moving window

$\hat{f}(y) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{y_i - y}{h}\right)$
(where K is a kernel function and h a bandwidth)

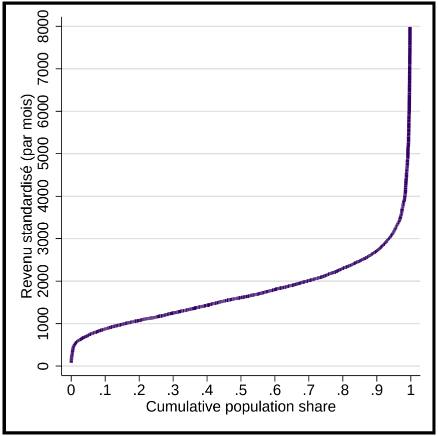
The cumulative distribution function (CDF)



$$F(y) = \Pr(Y \leqslant y)$$

$$\hat{F}(y) = \frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i \mathbf{I}(y_i \leqslant y)$$

The quantile function



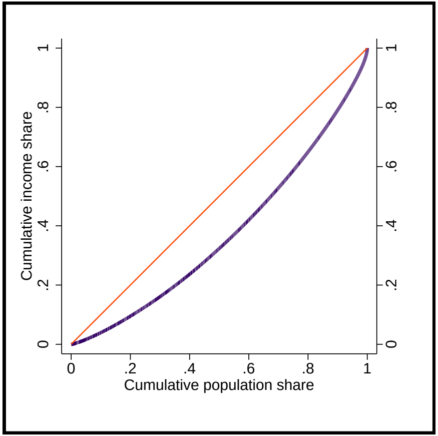
The CDF reversed:

$Q(p) = F^{-1}(y)$

$Q(p) = \inf\{y : p \leqslant F(y)\}$

cf. “Pen’s parade of giants and dwarves”

The Lorenz curve



If data are ordered by income:

$y_1 \leqslant y_2 \leqslant \ldots \leqslant y_N$ then

$L(p) = \frac{\sum_{i=1}^{Np} y_i}{\sum_{i=1}^{N} y_i}$ with

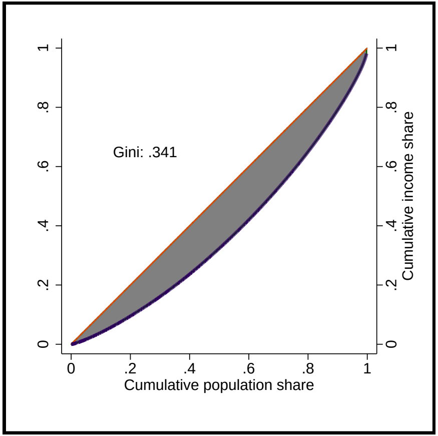
$0 \leqslant p \leqslant 1$

Also,

$L(p) = \frac{1}{\mu} \int_0^{F^{-1}(p)} x f(x) dx$

$L'(p) = \frac{Q(p)}{\mu}$

Lorenz curve and the Gini coefficient



The Gini coefficient is equal to twice the area between the 45 degree line and the Lorenz curve:

$$G = 1 - 2\int L(p)\,dp$$

Equivalently:

$$G = 1 - \frac{1}{N}\sum_{i=1}^{N} 2(1 - F(y_i))\frac{y_i}{\mu}$$

Or:

$$G = \frac{2Cov(y, F(y_i))}{\mu}$$

Or:

$$G = \frac{1}{2N^2\mu}\sum_{i=1}^{N}\sum_{j=1}^{N}|x_i - x_j|$$

Other common summary statistics

- Variance, σ^2: $\frac{1}{N-1}\sum_{i=1}^{N}(y_i-\mu)^2$
- Coefficient of variation, CV: $\frac{\sqrt{\sigma^2}}{\mu}$
- Inter-quintile ratio, P80/P20 (or inter-quartile P75/P25): $\frac{Q(0.80)}{Q(0.20)}$
- Income share ratios S80/S20: ratio between cumulative income of the richest 20% to cumulative income of poorest 20%; $\frac{(1-L(0.80))}{L(0.20)}$
- Top income shares (top 10%, 5%, 1%, 0.1%, ...)

Theil, mean log deviation and generalized entropy measures

Another popular family of inequality measures is the generalized entropy

$$\text{Theil} = \text{GE}_1 = \mathrm{E}\left(\frac{y}{\mu}\ln\frac{y}{\mu}\right) = \frac{1}{\mu}\mathrm{E}\left(y\ln y\right) - \ln\mu$$

$$\text{MLD} = \text{GE}_0 = \mathrm{E}\left(\ln\frac{\mu}{y}\right) = \ln\mu - \mathrm{E}\left(\ln y\right)$$

$$\text{GE}_\alpha = \frac{1}{\alpha(\alpha-1)}\left(\frac{\mathrm{E}\left(y^\alpha\right)}{\mu^\alpha} - 1\right)$$

Note that $\text{GE}_2 = \frac{1}{2}(\text{CoV})^2$

Responsive to low income with low α and to high incomes with high α

Poverty

- Focus on lower tail of the distribution
- examine a "censored" distribution of income shortfalls below a poverty threshold:

$$g_i = \max\left(\frac{z - y_i}{z}, 0\right)$$

- Tools reviewed for inequality and welfare carry through to analysis of censored distribution (with obvious adaptation)
- The 'Foster-Greer-Thorbecke' (FGT) index (Foster et al., 1984)

$$P_\alpha^{\mathrm{FGT}} = \frac{1}{N}\sum_{i=1}^{N} g_i^\alpha$$

with headcount ratio as special case $\alpha = 0$,

$$H = \frac{1}{N}\sum_{i=1}^{N} \mathbf{I}(y_i \leqslant z) = F(z)$$

Basic distribution analysis toolkit: Quick reminder

Decomposition approaches

- Decomposition by income sources
- Decompositions by population characteristics

Modelling distributions: reweighting and distribution regression methods

Outline

Basic distribution analysis toolkit: Quick reminder

Decomposition approaches

- Decomposition by income sources
- Decompositions by population characteristics

Modelling distributions: reweighting and distribution regression methods

- Reweighting methods
- Distribution regression methods
  - Modelling conditional distributions
  - Simulating counterfactual distributions

Inequality contribution of income sources

Household income is the sum of market income (incl. employment and self-employment incomes of all members, capital income), public transfers, pensions minus taxes and social security contributions

Net worth is the sum of various assets types minus liabilities

What is the contribution of different components to inequality?

Does capital income increase inequality? Do transfers reduce it? By how much? What is the contribution of housing wealth? etc.

Decomposition by factor components

- The Shorrocks (1982*a*,*b*) rule

$$S_j = \frac{\mathrm{Cov}_j}{V} = \rho_j \frac{\mu_j}{\mu} \frac{CV_j}{CV}$$

  where ρ_j is the correlation between total incomes and income source j
- Valid for ‘any’ inequality measure (choice of decomposition rule independent on index)

Decomposition of Gini coefficient

- Estimate the contributions to Gini coefficient (using concentration coefficient)
- Lerman and Yitzhaki (1985)

$$G = \sum_{j=1}^{K} \frac{\mu_j}{\mu} C_j$$

where C_j is the Concentration coefficient of source j on between total incomes and income source j

- Concentration coefficient of source j can itself be expressed as product of Gini of source j and 'Gini correlation' between source j and total income

$$C_j = G_j \rho_j^G$$

(Gini correlation is correlation between a household's income from source j and their *rank* in the distribution of *total* income)

Examining contribution to change

- Analysis of change:

$$\Delta G = \sum_{j=1}^{K} s_j^0 \Delta C_j + \sum_{j=1}^{K} \Delta s_j C_j^1$$

  with $s_j^t = \frac{\mu_j^t}{\mu^t}$ the share of source j at time t and $\Delta C_j = \Delta G_j \rho_j^0 + G_j^1 \Delta \rho_j$
- beware of index number issue

Alternative approaches: ‘shutting off’ sources

Alternative methods applicable to any generic measure consist in ‘shutting off’ sources (or inequality in given sources)

- Set all $y_{ij} = 0$ for source j and recalculate inequality measures of interest
- Similarly, could set all $y_{ij} = \bar{y_j}$ and recalculate inequality measures of interest
- Impact of source j given by $I(Y) - I(Y \setminus \{y_j\})$
- not a decomposition! Sum over j does not lead to $I(Y)$

$\Longrightarrow$ Shapley decomposition

Shapley decomposition (Chantreuil and Trannoy, 2013, Shorrocks, 2013, Chantreuil et al., 2019)

In fact, effect of shutting off source j can be calculated for any other reference value for other sources (turned on or off)

Shapley value averages impact of shutting off source j over all possible sequences

- 2^J possible combinations of sources (‘coalitions’; e.g. with three sources: (L,K,T), (L,K), (K,T), (L,T), (L), (K), (T), ())
- Four possible marginal values, e.g., for source L: $I(L, K, T) - I(K, T)$, $I(L, K) - I(K)$, $I(L, T) - I(T)$, $I(L) - I()$
- Beware: *weighted* average so that each ‘coalition size’ receives equal weight

$$C_j = \sum_{S \subset J, j \in S} \frac{!(s-1)!(J-s)}{!J} (I(S) - I(S \setminus y_j))$$

- exact decomposition: $I = \sum_{j \in J} C_j$ and no dependence on sequence

Outline

- Basic distribution analysis toolkit: Quick reminder
- Decomposition approaches
  - Decomposition by income sources
  - Decompositions by population characteristics
- Modelling distributions: reweighting and distribution regression methods
  - Reweighting methods
  - Distribution regression methods
    - Modelling conditional distributions
    - Simulating counterfactual distributions

Between- and within-group inequality

How do income differences between and within groups combine to shape the overall level of inequality? (Shorrocks, 1984)

- Partition the population into groups, e.g.,
  - urban vs. rural or other geography
  - age groups
  - gender, industries, household types, etc.
- Within-group inequality is as described so far, but assessed in a subgroup
- Between-group inequality
  - inequality if individuals have the average income in their group

Between- and within-group inequality

- In general cannot write $I(Y)$ as additive function of *between group* inequality and *within group* inequality
- ... except for a particular family of inequality measures, the Generalized Entropy family:

$$\mathrm{GE}(\alpha) = \mathrm{GE}_B(\alpha) + \sum_{k=1}^{K} v_k^{\alpha} s_k^{1-\alpha} \mathrm{GE}_k(\alpha)$$

  where v_k and s_k are subgroup k shares of total income and population respectively, $\mathrm{GE}_k(\alpha)$ is inequality in subgroup k and $\mathrm{GE}_B(\alpha)$ is between group inequality obtained by assuming everyone in a group obtains the group mean income

  » $\mathrm{GE}(\alpha) = \frac{1}{\alpha-\alpha^2}\left(1 - \int (\frac{x}{\mu})^{\alpha} f(x)dx\right)$

Gini coefficient

- The Gini coefficient is *not* exactly additively decomposable

$$\mathsf{GINI} = \mathsf{GINI}^{B}(Y^{*}) + \mathsf{GINI}^{W} + R$$

  with

$$\mathsf{GINI}^{W} = \sum_{k=1}^{K} \nu_k s_k \mathsf{GINI}^{(k)}$$

- $\mathsf{GINI}^{B}(Y^{*})$ is defined as for the GE decomposition as the inequality if all agents received their subgroup mean income.
- The term $R(Y)$ depends the degree of overlap between the income range of the different subgroups. (Reflects difference between ranks in own group distribution and in overall distribution.)

Poverty decompositions

FGT indices of poverty are, of course, easy to decompose

$$\mathsf{FGT}(\alpha) = \sum_{k=1}^{K} s_k \mathsf{FGT}_k(\alpha)$$

(assuming a common poverty line)

Multivariate approaches

- In principle, we can combine multiple variables, but cells quickly become too small (need relatively large number of observations to estimate inequality measures)
- Smoothing techniques? (to estimate subgroup distributions or subgroup inequality measures)—interpretation of 'within inequality' becomes somewhat fuzzy
- Morduch and Sicular (2002)'s simple OLS strategy (see below)
- Explicit models for conditional distributions and estimation of "partial" effects

Regression components as income sources Morduch and Sicular (2002), Fields (2003), Cowell and Fiorio (2011)

– Step 1: regress income on variables of interest

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + ... + e$$

– Step 2: apply inequality decomposition by source where the sources are $\beta_j x_j$ and e
– Contribution of component depends on β and variation in x
– See Cowell and Fiorio (2011) for the links between this and the between-group decomposition
– But often most of the contribution is in the residual e...
– ... sensitive to specification (and interpretation gets complicated with interaction terms)

Some Stata (user-written) routines

`ineqfac`	Ineq decomp by source (Shorrocks rule)
`sgini`, `descogini`	Gini decomp by source
`ineqdeco`	GE decomp by subgroup
`povdeco`	FGT decomp by subgroup
`ineqrbd`, `gfields`	Morduch-Sicular regression-based decomp
`anogi`	Gini decomp by subgroup
`msdeco`	Mookherjee-Shorrocks contribution of groups to change

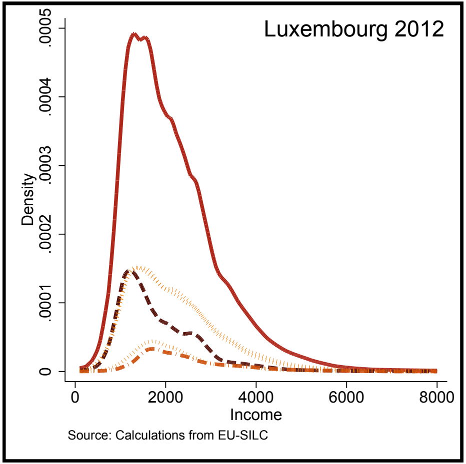
Basic distribution analysis toolkit: Quick reminder

Decomposition approaches

Modelling distributions: reweighting and distribution regression methods

- Reweighting methods
- Distribution regression methods

Examine PDFs directly (Jenkins and Van Kerm, 2005)



Often useful to examine PDFs (or CDFs directly)

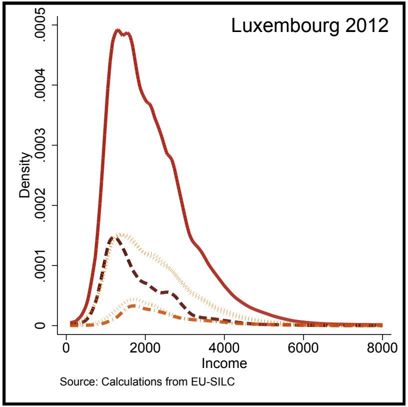
$$f(y) = \sum_{k=1}^{K} s_k f^{(k)}(y)$$

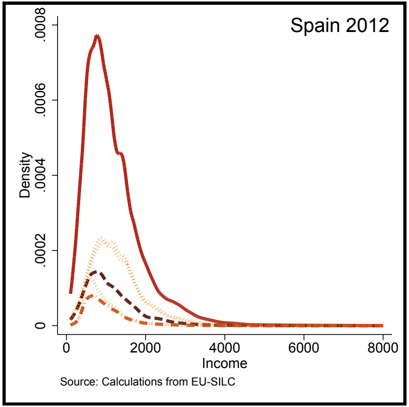
$$F(y) = \sum_{k=1}^{K} s_k F^{(k)}(y)$$

where $f^{(k)}$ and $F^{(k)}$ are the PDF and the CDF of incomes in group m respectively.

Use logarithmic scale for income (to visualize *relative* differences)

Contribution to differences





Decomposition

Difference between two PDFs

$$\Delta^{(1,2)} f(y) \equiv f^2(y) - f^1(y)$$

The contribution of differences in population composition to $\Delta^{(1,2)} f(y)$ can be trivially expressed as

$$\begin{aligned}
\Delta^{(1,2)} f(y) &\equiv f^2(y) - f^1(y) \\
&= \sum_{k=1}^{K} \left(s_k^2 f_k^2(y) - s_k^1 f_k^1(y) \right) \\
&= \underbrace{\left(\sum_{k=1}^{K} \left(s_k^2 - s_k^1 \right) f_k^1(y) \right)}_{\text{composition}} + \underbrace{\left(\sum_{k=1}^{K} s_k^2 \left(f_k^2(y) - f_k^1(y) \right) \right)}_{\text{subgroup distribution}}
\end{aligned}$$

– Each of the terms in the decomposition can be easily calculated (provided K small enough relative to sample size)

Many or continouous variables?

Estimation of the components unproblematic when the set is small (and discrete)

With large dimension and/or continuous X, estimation of both the shares and the densities becomes difficult. Solution: reweighting and 'distribution regression'

Express $f^{(m)}(y)$ in terms of conditional distributions given covariates X

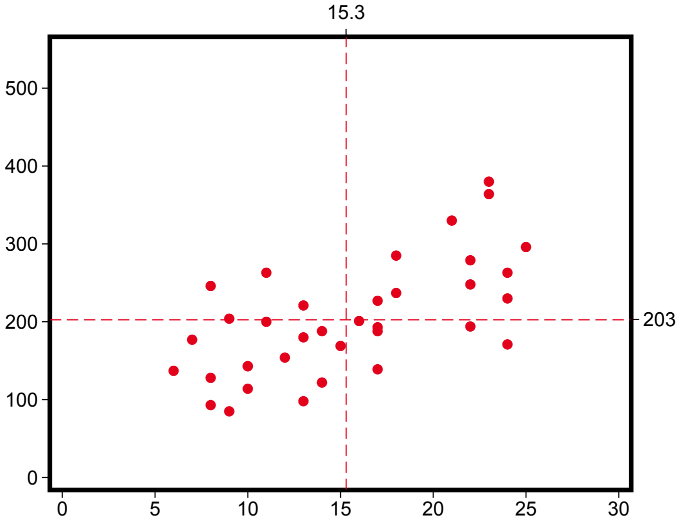
$$f^{(m)}(y) = \int_{\Omega_X} f^{(m)}(y|X)g^{(m)}(X)dX$$

How much of $\Delta f^{(1,2)}(y)$ (for any y) is due to differences in $f^{(m)}(y|X)$ and how much is due to differences in $g^{(m)}(X)$

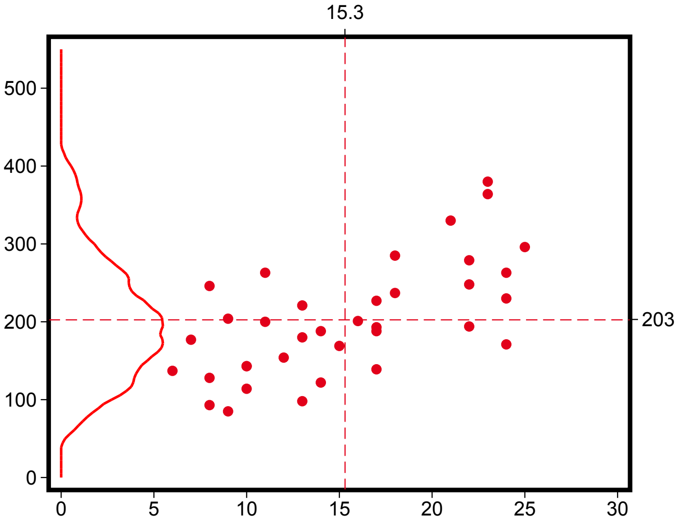
$$\begin{aligned}\Delta f^{(1,2)}(y) &= \underbrace{\left(\int_{\Omega_X} f^{(2)}(y|X)g^{(2)}(X)dX - \int_{\Omega_X} f^{(2)}(y|X)g^{(1)}(X)dX\right)}_{\text{composition differences}} \\ &+ \underbrace{\left(\int_{\Omega_X} f^{(2)}(y|X)g^{(1)}(X)dX - \int_{\Omega_X} f^{(1)}(y|X)g^{(1)}(X)dX\right)}_{\text{conditional distribution differences}}\end{aligned}$$

Outline

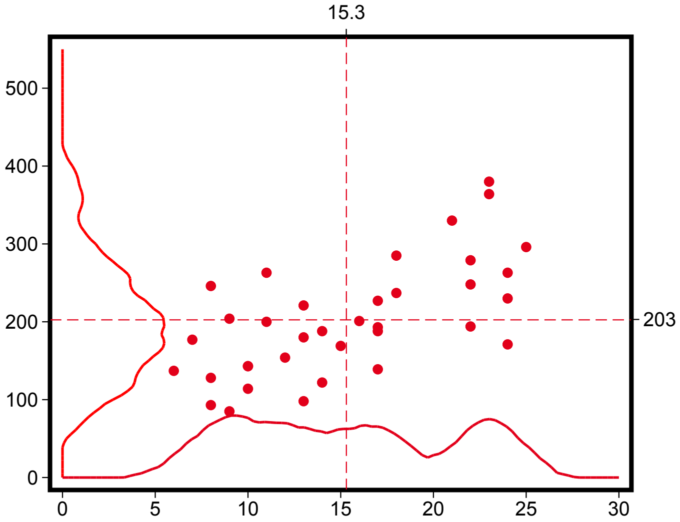
Easy on a diagram



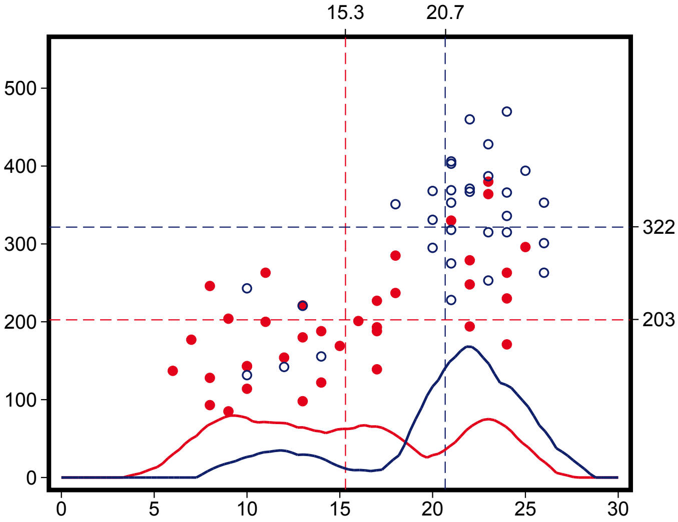
Easy on a diagram



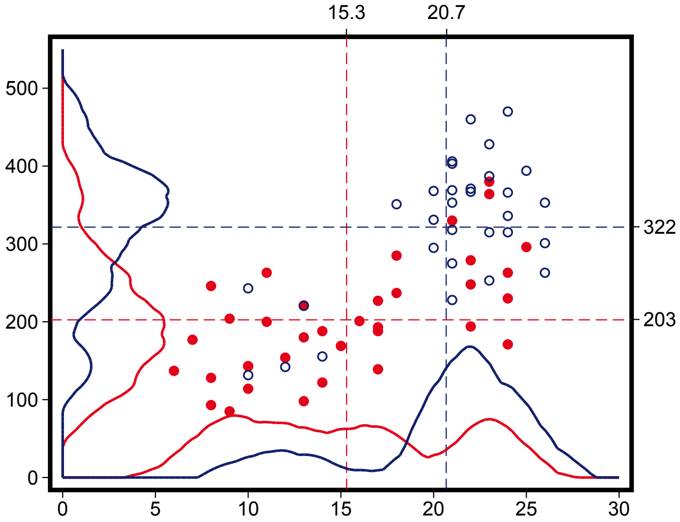
Easy on a diagram



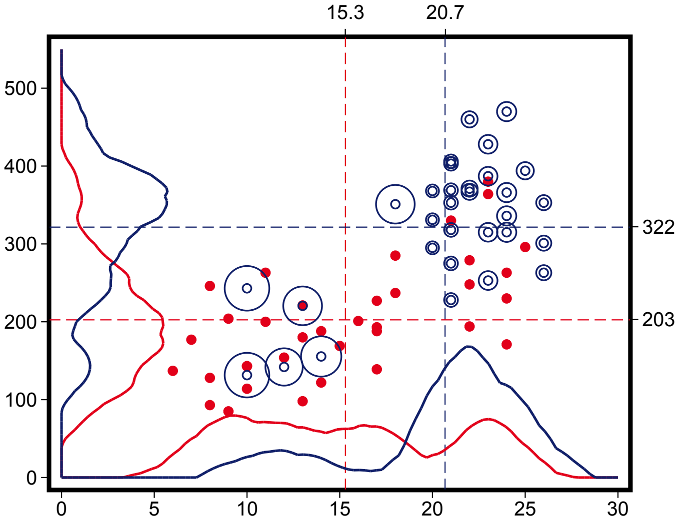
Easy on a diagram



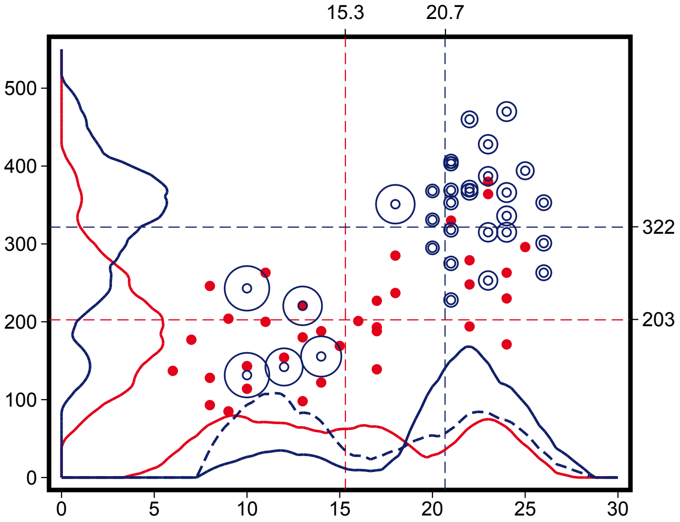
Easy on a diagram



Easy on a diagram



Easy on a diagram



Reweighting principle

Define the ‘reweighting function’ as ratio of density of covariates

$$\psi^{(1,2)}(X) = \frac{g^{(1)}(X)}{g^{(2)}(X)}$$

Inserting $\psi^{(1,2)}(X)$ in the PDF decomposition gives

$$\begin{aligned}
\Delta f^{(1,2)}(y) &= \underbrace{\left(\int_{\Omega_X} f^{(2)}(y|X)g^{(2)}(X)dX - \int_{\Omega_X} \psi^{(1,2)}(X)f^{(2)}(y|X)g^{(2)}(X)dX\right)}_{\text{composition differences}} \\
&+ \underbrace{\left(\int_{\Omega_X} \psi^{(1,2)}(X)f^{(2)}(y|X)g^{(2)}(X)dX - \int_{\Omega_X} f^{(1)}(y|X)g^{(1)}(X)dX\right)}_{\text{conditional distribution differences}}
\end{aligned}$$

Estimation of components (ctd.)

Weighted kernel density estimators:

$$f^{(m)}(y) = \int_{\Omega_X} f^{(m)}(y|X) g^{(m)}(X) dX = \frac{1}{N^{(m)}} \sum_{i \in S^{(m)}} \frac{1}{h} K\left(\frac{y - y_i}{h}\right)$$

and

$$\int_{\Omega_X} \psi^{(1,2)}(X) f^{(2)}(y|X) g^{(2)}(X) dX = \frac{1}{N^{(2)}} \sum_{i \in S^{(2)}} \frac{\psi^{(1,2)}(X_i)}{h} K\left(\frac{y - y_i}{h}\right)$$

No estimation of conditional distributions anymore!

Estimation of components (ctd.)

Call upon Bayes' rule and binary choice models

Express reweighting function, using Bayes' rule, as

$$\psi^{(1,2)}(X) = \frac{\Pr[m = 1|X]}{\Pr[m = 2|X]} \times \frac{\Pr[m = 2]}{\Pr[m = 1]}$$

where $\Pr[m = i|X]$ is probability that a randomly selected agent with characteristics X belongs to group i. $\Pr[m = i]$ is probability that any randomly selected agent belong to group i.

Whereas $g^{(m)}(X)$ is multivariate, the four terms in $\psi^{(1,2)}(X)$ are easy to compute.

Di Nardo et al. (1996) specify simple single index model to estimate $\Pr[m = i|X]$ from the pooled population (logit or probit). Barsky et al. (2002), Firpo and Pinto (2016) adopt a more flexible non-parametric approach (cf. estimation of propensity score in matching methods (Hirano et al., 2003))

Common support restriction

Problem if some covariate combinations do not exist in one of the two populations.

- Case 1: $g^{(1)}(X) = 0$ for some X, then $\psi^{(1,2)}(X) = 0$
  - No big deal—that subgroup is ignored
- Case 2: $g^{(2)}(X) = 0$ for some X, then $\psi^{(1,2)}(X) = \infty$!
  - We cannot ‘standardize’ population 2 to population 1 characteristics because no matching observations

$\Longrightarrow$ Take reference population as the ‘most compact’ or take alternative reference population (pooled population)

(milder version of the problem when $g^{(2)}(X) \approx 0$ and reweighting factor gets very large: estimation unstable)

Separating different covariates

Say, $X = \{Z, W\}$

We have

$$g^{(m)}(W, Z) = g^{(m)}(W|Z) \times h^{(m)}(Z)$$

and

$$f^{(m)}(y) = \int_{\Omega_Z} \int_{\Omega_W} f^{(m)}(y|W, Z) g^{(m)}(W|Z) h^{(m)}(Z) dW dZ.$$

A sequence of counterfactual distributions can now be constructed:

$$\int_{\Omega_Z} \int_{\Omega_W} f^{(2)}(y|W, Z) g^{(2)}(W|Z) h^{(1)}(Z) dW dZ$$

and

$$\int_{\Omega_Z} \int_{\Omega_W} f^{(2)}(y|W, Z) g^{(1)}(W|Z) h^{(1)}(Z) dW dZ$$

Separating different covariates

Define

$$\begin{aligned}\psi_{\mathrm{I}}^{(1,2)}(Z) &= \frac{h^{(1)}(Z)}{h^{(2)}(Z)} \\ \psi_{\mathrm{II}}^{(1,2)}(W,Z) &= \frac{g^{(1)}(W|Z)}{g^{(2)}(W|Z)}\end{aligned}$$

So $\psi^{(1,2)}(W,Z) = \psi_{\mathrm{I}}^{(1,2)}(Z) \times \psi_{\mathrm{II}}^{(1,2)}(W,Z)$.

Direct estimation of the distributions $h^{(m)}(Z)$ is only needed for constructing $\psi_{\mathrm{I}}^{(1,2)}(Z)$ but given Z is a single covariate, this does not pose any problem.

If W contains a large number of covariates, $\psi_{\mathrm{II}}^{(1,2)}(W,Z)$ can be estimated by applying the Bayes' rule as above.

Sequence issue

Shapley

One problem of this technique is the ‘index number problem’ – the position at which the subsets of covariates are neutralized in the sequence of elimination influences their estimated contribution.

Additionally, one could take the conditional outcome distribution of group 1 as reference rather than group 2

Shapley value approach: estimate all possible elimination sequences and take the average across sequence (see, e.g., Cobb-Clark and Hildebrand, 2006).

CDFs, quantiles and other functionals

Reweighting approach applies similarly for

- CDFs: replace (weighted) kernel density estimation by (weighted) empirical CDF estimators (Bover, 2010)
- quantiles: weighted quantiles (Firpo, 2007)
- direct functionals of interest such as inequality measures (Biewen, 2001, Firpo and Pinto, 2016)

Outline

Basic distribution analysis toolkit: Quick reminder

Decomposition approaches

- Decomposition by income sources
- Decompositions by population characteristics

Modelling distributions: reweighting and distribution regression methods

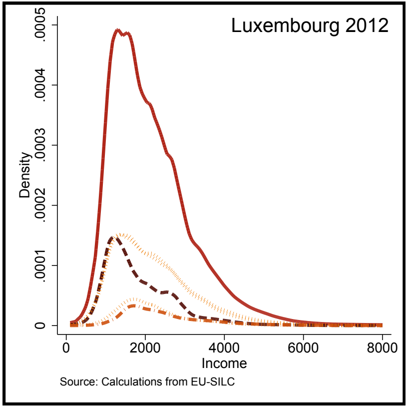
- Reweighting methods
- Distribution regression methods
  - Modelling conditional distributions
  - Simulating counterfactual distributions

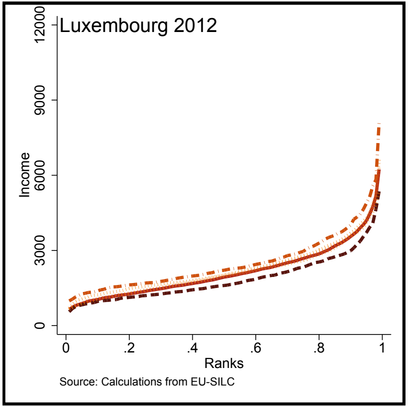
Distinct questions

Methods address two related but distinct questions:

1. How much does X contribute to $\upsilon(F)$?
   - How much can a change in some element in X affect $\upsilon(F)$? ('policy effects')
   - How much do differences in X account for differences in $\upsilon(F)$ between A and B (across time, countries, gender, race, ...)?
2. How does $\upsilon(F_x)$ vary with X?
   That is, calculate and summarize $\upsilon(F_x)$ (remember $\dim(X) > 1$), 'partial effects')
   - EOp, Intergenerational mob, Educ choices, Income risk and vulnerability, wage gap and glass ceilings. etc.

Distinct questions





Distribution regression

Part I:

Conditional distribution models (‘distribution regression’)

Part II:

Simulating unconditional (counterfactual) distributions

Array of estimators

Many estimators available:

- quantile regression (Koenker and Bassett, 1978)
- distribution regression (Foresi and Peracchi, 1995)
- parametric income distribution models (Biewen and Jenkins, 2005, Van Kerm et al., 2017)
- also: duration models (Donald et al., 2000), ordered probit model (Fortin and Lemieux, 1998)

Quantile regression

Distribution regression

Parametric models

Linear quantile regression model

Assume a particular relationship (linear) between conditional quantile and x:

$$Q_\tau(y|x) = x\beta_\tau$$

(Or equivalently $y_i = x_i\beta_\tau + u_i$ where $F^{-1}_{u_i|x_i}(\tau) = 0$)

$$\hat{\beta}_\tau = \arg\min_{\beta} \sum_i \rho_\tau(y_i - x_i\beta)$$

(Koenker and Bassett, 1978)

▸ Check function

Estimate of the conditional quantile (given linear model):

$$\hat{Q}_\tau(y|x) = x\hat{\beta}_\tau$$

$\hat{\beta}_\tau$ can be interpreted as the marginal change in the τ conditional quantile for a marginal change in x

(Stata: `qreg`)

Illustrative examples

Quantile regressions (Lux 2012)

```
.1 Quantile regression                              Number of obs =     16,096
  Raw sum of deviations 43979.03 (about 7.2287302)
  Min sum of deviations 39977.57                    Pseudo R2     =     0.0910
```

lninc	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
femmain	-.0605848	.0129394	-4.68	0.000	-.0859475	-.0352222
agemain	.0087514	.0009275	9.44	0.000	.0069334	.0105693
intover62	.0094736	.0027045	3.50	0.000	.0041724	.0147748
shatwork	.4142671	.0226747	18.27	0.000	.3698221	.4587121
intshatover62	.0257374	.1639491	0.16	0.875	-.2956211	.347096
nadu2	-.0049892	.0103614	-0.48	0.630	-.0252987	.0153203
nkid06	-.0340307	.011162	-3.05	0.002	-.0559094	-.012152
nkid712	-.0239635	.0064344	-3.72	0.000	-.0365756	-.0113514
nkid1318	-.1378561	.0075209	-18.33	0.000	-.152598	-.1231143
nkid19plus	-.0299346	.0184035	-1.63	0.104	-.0660074	.0061382
_cons	6.691165	.054638	122.46	0.000	6.584068	6.798261

Illustrative examples

Quantile regressions (Lux 2012)

```
Median regression                                   Number of obs =     16,096
  Raw sum of deviations 101339.5 (about 7.8499236)
  Min sum of deviations 89961.48                    Pseudo R2     =     0.1123

------------------------------------------------------------------------------
             |               Robust
       lninc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     femmain |  -.0775493   .0105686    -7.34   0.000    -.0982649   -.0568336
     agemain |   .0115235   .0007851    14.68   0.000     .0099846    .0130624
   intover62 |  -.0043322    .001896    -2.28   0.022    -.0080486   -.0006158
    shatwork |   .5130259   .0172372    29.76   0.000     .4792391    .5468128
intshatover62 |  -.0336135   .0349685    -0.96   0.336    -.1021557    .0349287
       nadu2 |  -.0199042   .0057565    -3.46   0.001    -.0311876   -.0086208
      nkid06 |  -.0856363   .0097761    -8.76   0.000    -.1047986    -.066474
     nkid712 |  -.0869693   .0068856   -12.63   0.000    -.1004659   -.0734728
    nkid1318 |  -.2117496   .0075151   -28.18   0.000      -.22648   -.1970191
  nkid19plus |  -.0657151   .0100478    -6.54   0.000      -.08541   -.0460202
       _cons |    7.15211   .0452135   158.19   0.000     7.063486    7.240733
------------------------------------------------------------------------------
```

Illustrative examples

Quantile regressions (Lux 2012)

```
.9 Quantile regression                              Number of obs =     16,096
  Raw sum of deviations 44819.03 (about 8.4579811)
  Min sum of deviations 40827.28                    Pseudo R2     =     0.0891

------------------------------------------------------------------------------
             |              Robust
       lninc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     femmain |  -.0387195   .0181955    -2.13   0.033    -.0743846   -.0030544
     agemain |   .0143392   .0011628    12.33   0.000       .01206    .0166183
  intover62 |  -.0085015   .0031314    -2.71   0.007    -.0146394   -.0023636
    shatwork |   .4959385   .0250408    19.81   0.000     .4468556    .5450213
intshatover62 |   .0174295   .1048977     0.17   0.868    -.1881817    .2230408
       nadu2 |  -.0403751   .0108402    -3.72   0.000    -.0616231   -.0191271
      nkid06 |   -.039867   .0124805    -3.19   0.001    -.0643302   -.0154038
     nkid712 |  -.0838864   .0174193    -4.82   0.000    -.1180302   -.0497426
    nkid1318 |  -.1070223   .0126726    -8.45   0.000    -.1318619   -.0821827
  nkid19plus |  -.1775758   .0202811    -8.76   0.000     -.217329   -.1378225
       _cons |   7.582603   .0677037   112.00   0.000     7.449897     7.71531
------------------------------------------------------------------------------
```

Recovering $\upsilon(F_x)$

Estimation of $\hat{Q}_\tau(y|x)$ for a continuum of τ in $(0, 1)$ provides a model for the entire conditional quantile function (quantile process) of Y given X

But we are interested in $\upsilon(F_x)$ not (necessarily) in the quantiles!

After estimation of the quantile process $(0, 1)$, estimation of the distributional statistic conditional on X is straightfoward:

- The set of predicted conditional quantile values $\{x_i\hat{\beta}_\theta\}_{\theta\in(0,1)}$ is a pseudo-random draw from F_x (if grid for θ is equally-spaced) (Autor et al., 2005)
- So a simple estimator for υ from unit-record data can be used to estimate $\upsilon(F_{X_i})$

Quantile regression

Distribution regression

Parametric models

'Distribution regression'

$F_x(y) = \Pr\{y_i \leqslant y|x\}$ is a binary choice model once y is fixed (dependent variable is $1(y_i < y)$)

‘Distribution regression’

$F_x(y) = \Pr\{y_i \leqslant y|x\}$ is a binary choice model once y is fixed (dependent variable is $1(y_i < y)$)

Idea is to estimate $F_x(y)$ on a grid of values for y spanning the domain of definition of Y by running repeated standard binary choice models, e.g. a logit:

$$\begin{aligned} F_x(y) &= \Pr\{y_i \leqslant y|x\} \\ &= \Lambda(x\beta_y) \\ &= \frac{\exp(x\beta_y)}{1+\exp(x\beta_y)} \end{aligned}$$

or a probit $F_x(y) = \Phi(x\beta_y)$ or else ...
(see Foresi and Peracchi, 1995, Chernozhukov et al., 2013)

‘Distribution regression’

- Estimation of these models is well-known and straightforward! (`probit`, `logit`)
- Faster to run than quantile regression
- Evidence that provides better fit than quantile regression (Rothe and Wied, 2013, Van Kerm et al., 2017)
- See Chernozhukov et al. (2013) on inference

‘Distribution regression’

Drawback: Conditional statistic $\upsilon(F_x)$ often less easy to recover from the $\hat{F}_X$ predictions than with quantile regression

- invert the predicted F_x to obtain predicted quantiles
- proceed as with quantiles predicted from quantile regression (see above)

Quantile regression

Distribution regression

Parametric models

Parametric distribution fitting

Assume that the conditional distribution has a particular parametric form: e.g., (log-)normal (2 parameters – quite restictive), Fisk (2 params), Gamma (2 params), Singh-Maddala (3 param.), Dagum (3 param.), GB2 (4 param.), ... or any other distribution that is likely to fit the data at hand (think domain of definition, fatness of tails, modality)

Let parameters (say vector θ) depend on x in a particular fashion, typically linearly (up to some transformation), e.g., $\theta_1 = \exp(x\beta_1)$, $\theta_2 = \exp(x\beta_2)$ and $\theta_3 = x\beta_3$

This gives a fully specified parametric model which can be estimated using maximum likelihood.

Illustrative examples

Singh-Maddala distribution (Lux 2012)

```
ML fit of Singh-Maddala distribution                 Number of obs   =      16096
                                                     Wald chi2(10)   =      28.27
Log pseudolikelihood =  -9759.637                    Prob > chi2     =     0.0016

                                 (Std. Err. adjusted for 6,003 clusters in hid)
------------------------------------------------------------------------------
             |               Robust
       lninc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
a            |
     femmain |  -.3505394   1.245992    -0.28   0.778    -2.792638    2.091559
     agemain |  -.1820589   .0621474    -2.93   0.003    -.3038656   -.0602522
   intover62 |    .636284   .1487384     4.28   0.000      .344762    .9278059
    shatwork |  -.9185511   1.401675    -0.66   0.512    -3.665784    1.828681
intshatover62 |  -1.448951    4.25992    -0.34   0.734    -9.798241    6.900338
       nadu2 |   1.718954   1.139223     1.51   0.131    -.5138812    3.951789
      nkid06 |  -1.205291   .9535327    -1.26   0.206     -3.07418    .6635993
     nkid712 |    .038569   1.149245     0.03   0.973     -2.21391    2.291048
    nkid1318 |  -1.329007   .9144699    -1.45   0.146    -3.121335     .463321
  nkid19plus |   2.513926   1.901001     1.32   0.186    -1.211968     6.23982
       _cons |   35.14723   3.993827     8.80   0.000     27.31948    42.97499
-------------+----------------------------------------------------------------
b            |
     femmain |  -.0668657   .0215139    -3.11   0.002    -.1090322   -.0246991
     agemain |   .0117232   .0013603     8.62   0.000     .0090571    .0143893
   intover62 |  -.0034883    .002821    -1.24   0.216    -.0090174    .0020407
    shatwork |   .5130456   .0324419    15.81   0.000     .4494606    .5766305
intshatover62 |  -.0074776   .0842774    -0.09   0.929    -.1726582     .157703
       nadu2 |  -.0177483   .0170645    -1.04   0.298    -.0511941    .0156975
      nkid06 |  -.0718088   .0213111    -3.37   0.001    -.1135779   -.0300398
     nkid712 |  -.0753621   .0200344    -3.76   0.000    -.1146289   -.0360953
    nkid1318 |  -.1868884    .020133    -9.28   0.000    -.2263484   -.1474284
  nkid19plus |  -.0711948   .0255365    -2.79   0.005    -.1212454   -.0211441
       _cons |   7.190304   .0928204    77.46   0.000      7.00838    7.372229
-------------+----------------------------------------------------------------
q            |
       _cons |   1.209398   .1588487     7.61   0.000     .8980604    1.520736
------------------------------------------------------------------------------
```

Illustrative examples

Singh-Maddala distribution (Lux 2012)

```
Expression    : predict(equation(b))*((1-.1)^(-1/predict(equation(q))) - 1)^(1/predict(equation(a)))
dy/dx w.r.t. : femmain agemain intover62 shatwork intshatover62 nadu2 nkid06 nkid712 nkid1318 nkid19plus

------------------------------------------------------------------------------
             |      dy/dx
-------------+----------------------------------------------------------------
     femmain |  -.0687356
     agemain |   .0071228
   intover62 |   .0096611
    shatwork |    .454373
intshatover62 |  -.0362167
       nadu2 |   .0184259
      nkid06 |  -.0905907
     nkid712 |  -.0686936
    nkid1318 |  -.1991831
  nkid19plus |  -.0147566
------------------------------------------------------------------------------
```

Illustrative examples

Singh-Maddala distribution (Lux 2012)

```
Expression   : predict(equation(b))*((1-.5)^(-1/predict(equation(q))) - 1)^(1/predict(equation(a)))
dy/dx w.r.t. : femmain agemain intover62 shatwork intshatover62 nadu2 nkid06 nkid712 nkid1318 nkid19plus

------------------------------------------------------------------------------
             |      dy/dx
-------------+----------------------------------------------------------------
     femmain |   -.067103
     agemain |   .0111973
   intover62 |  -.0019753
    shatwork |    .506458
intshatover62 |  -.0107894
       nadu2 |  -.0135886
      nkid06 |  -.0739952
     nkid712 |   -.074619
    nkid1318 |  -.1883656
  nkid19plus |  -.0647193
------------------------------------------------------------------------------
```

Illustrative examples

Singh-Maddala distribution (Lux 2012)

```
Expression   : predict(equation(b))*((1-.9)^(-1/predict(equation(q))) - 1)^(1/predict(equation(a)))
dy/dx w.r.t. : femmain agemain intover62 shatwork intshatover62 nadu2 nkid06 nkid712 nkid1318 nkid19plus

------------------------------------------------------------------------------
              |      dy/dx
--------------+---------------------------------------------------------------
      femmain |  -.0649957
      agemain |    .015526
    intover62 |   -.014494
     shatwork |   .5599016
intshatover62 |   .0166446
        nadu2 |  -.0479881
       nkid06 |  -.0557414
      nkid712 |  -.0806103
     nkid1318 |  -.1757376
   nkid19plus |  -.1181774
------------------------------------------------------------------------------
```

Outline

Part I:

Conditional distribution models ('distribution regression')

Part II:

Simulating unconditional (counterfactual) distributions

Inference via counterfactual distributions (Chernozhukov et al., 2013)

Analysis via counterfactual distributions is typically three-stage:

- model and estimate conditional distribution functions $F_x(y)$ (or conditional quantile functions)
- recover prediction for F by averaging over covariate distribution: $F(y) = \int F_x(y) h(x) dx$
  - in a sample: $\hat{F}(y) = \frac{1}{N} \sum_{i=1}^{N} \hat{F}_{x_i}(y)$
- Generalized Oaxaca-Blinder: simulate counterfactual distributions $\tilde{F}$ by manipulating
  - the conditional distribution functions: $\tilde{F}(y) = \int G_x(y) h(x) dx$:
  - the covariate distributions: $\tilde{F}(y) = \int F_x(y) g(x) dx$:
  - (typical analysis swaps either component across, say countries, gender, etc.)

Counterfactual distributional statistics

Simulation consists in generating a simulated sample from F on the basis of conditional quantile estimates.

Machado and Mata (2005) algorithm:

- pick a random value $\theta \in (0, 1)$ and calculate conditional quantile regression for the θ-th quantile
- select a random observation x_j from the sample and calculate predicted value $Q_{x_j} = x_j \hat{\beta}_\theta$
- repeat steps above B times to generate a simulated sample from F based on the conditional quantile model

- $\upsilon(F)$ calculated with standard formulae from the simulated sample

Decomposition of quantile differences

Machado-Mata very computationally intensive, especially since large B required for accurate estimation of $\upsilon(F)$.

Simplified version (Autor et al., 2005, Melly, 2005):

- estimate uniform (equally-spaced) sequence of conditional quantile predictions for each observations—pseudo-random sample from the conditional distribution F_x
- stack vectors of predictions for all observations into one long vector V—pseudo-random sample from the unconditional distribution!

Note: can obtain the conditional quantile predictions by ‘distribution regression’ or ...

Some Stata (user-written) routines

`teffects ipw`, `dfl`	Reweighting approaches
`counterfactual`, `drprocess`...	Quantile process counterfactual estimation
`drprocess`, `drpredict`	
`smfit`, `dagumfit`...	Fitting parametric distribution models
`fiskfit`, `gb2fit`	with covariates

To wrap up

Extensions of ‘Oaxaca-Blinder’ techniques to income *distribution* and welfare measures useful for cross-country and inter-temporal analysis of inequality—and this is what LIS/LWS is made for!

- ‘Classic’ decompositions: easy but for specific index numbers only (and trends may be sensitive to index used)
- ‘Modern’ (Generalized Oaxaca-Blinder) decompositions: relatively easy (albeit computational intensive)
- Descriptive methods? Causality depends on the design of the data generating process (not the methods)!
- Practical nuisance 1: path-dependence is pervasive (Shapley value – sensitivity analysis)
- Practical nuisance 2: bootstrap inference may be required (and computationally intensive)
- Practical nuisance 3: sensitivity to extreme data (beware of GE(2) and CV)!

References I

Arulampalam, W., Booth, A. L. and Bryan, M. L. (2007), ‘Is there a glass ceiling over Europe? Exploring the gender pay gap across the wage distribution’, *Industrial and Labor Relations Review* **60**(2), 163–186.

Autor, D. H., Katz, L. F. and Kearney, M. S. (2005), Rising wage inequality: The role of composition and prices, NBER Working Paper 11628, National Bureau of Economic Research, Cambridge MA, USA.

Barsky, R., Bound, J., Charles, K. K. and Lupton, J. P. (2002), ‘Accounting for the black-white wealth gap: A nonparametric approach’, *Journal of the American Statistical Association* **97**(459), 663–673.

Belfield, C., Blundell, R., Cribb, J., Hood, A. and Joyce, R. (2017), ‘Two decades of income inequality in Britain: The role of wages, household earnings and redistribution’, *Economica* **84**(334), 157–179.

Biewen, M. (2001), ‘Measuring the effects of socio-economic variables on the income distribution: An application to the East German transition process’, *Review of Economics and Statistics* **83**(1), 185–190.

Biewen, M. and Jenkins, S. P. (2005), ‘A framework for the decomposition of poverty differences with an application to poverty differences between countries’, *Empirical Economics* **30**(2), 331–358.
URL: *http://dx.doi.org/10.1007/s00181-004-0229-1*

Biewen, M. and Juhasz, A. (2012), ‘Understanding rising income inequality in Germany, 1999/2000-2005/2006’, *Review of Income and Wealth* **58**(4), 622–647.
URL: *http://dx.doi.org/10.1111/j.1475-4991.2012.00514.x*

References II

Bourguignon, F., Ferreira, F. H. and Leite, P. G. (2008), 'Beyond Oaxaca-Blinder: Accounting for differences in household income distributions', *Journal of Economic Inequality* **6**(2), 117–148.

Bover, O. (2010), 'Wealth inequality and household structure: U.S. versus Spain', *Review of Income and Wealth* **56**(2), 259–290.
URL: *http://ideas.repec.org/a/bla/revinw/v56y2010i2p259-290.html*

Butcher, K. F. and DiNardo, J. (2002), 'The immigrant and native-born wage distributions: Evidence from United States censuses', *Industrial and Labor Relations Review* **56**(1), 97–121.
URL: *http://ideas.repec.org/a/ilr/articl/v56y2002i1p97-121.html*

Chantreuil, F., Courtin, S., Fourrey, K. and Lebon, I. (2019), 'A note on the decomposability of inequality measures', *Social Choice and Welfare* **53**(2), 283–298.

Chantreuil, F. and Trannoy, A. (2013), 'Inequality decomposition values: the trade-off between marginality and efficiency', *Journal of Economic Inequality* **11**(1), 83–98.

Chernozhukov, V., Fernandez-Val, I. and Melly, B. (2013), 'Inference on counterfactual distributions', *Econometrica* **81**(6), 2205–2268.
URL: *http://dx.doi.org/10.3982/ECTA10582*

Cobb-Clark, D. A. and Hildebrand, V. A. (2006), 'The wealth of Mexican Americans', *Journal of Human Resources* **41**(4), 841–868.

References III

Cowell, F. and Fiorio, C. (2011), 'Inequality decompositions—a reconciliation', *Journal of Economic Inequality* **9**(4), 509–528.

Daly, M. C. and Valletta, R. G. (2006), 'Inequality and poverty in United States: The effects of rising dispersion of men's earnings and changing family behaviour', *Economica* **73**(289), 75–98.
URL: *http://www.jstor.org/stable/3549189*

Di Nardo, J., Fortin, N. M. and Lemieux, T. (1996), 'Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach', *Econometrica* **64**(5), 1001–1044.

Donald, S. G., Green, D. A. and Paarsch, H. J. (2000), 'Differences in wage distributions between Canada and the United States: An application of a flexible estimator of distribution functions in the presence of covariates', *Review of Economic Studies* **67**(4), 609–633.

Fields, G. S. (2003), 'Accounting for income inequality and its change: A new method, with application to the distribution of earnings in the United States', *Research in Labor Economics* **22**, 1–38.

Fiorio, C. V. (2011), 'Understanding Italian inequality trends', *Oxford Bulletin of Economics and Statistics* **73**(2), 255–275.

Firpo, S. (2007), 'Efficient semiparametric estimation of quantile treatment effects', *Econometrica* **75**(1), 259–276.
URL: *http://www.jstor.org/stable/4123114*

References IV

Firpo, S. and Pinto, C. (2016), 'Identification and estimation of distributional impacts of interventions using changes in inequality measures', *Journal of Applied Econometrics* **31**(3), 457–486.
URL: *http://dx.doi.org/10.1002/jae.2448*

Foresi, S. and Peracchi, F. (1995), 'The conditional distribution of excess returns: An empirical analysis', *Journal of the American Statistical Association* **90**(430), 451–466.

Fortin, N. M. and Lemieux, T. (1998), 'Rank regressions, wage distributions, and the gender gap', *The Journal of Human Resources* **33**(3), 610–643.
URL: *http://www.jstor.org/stable/146335*

Foster, J., Greer, J. and Thorbecke, E. (1984), 'A class of decomposable poverty measures', *Econometrica* **52**(3), 761–766.
URL: *http://www.jstor.org/stable/1913475*

Hirano, K., Imbens, G. W. and Ridder, G. (2003), 'Efficient estimation of average treatment effects using the estimated propensity score', *Econometrica* **71**(4), 1161–1189.

Hyslop, D. R. and Maré, D. C. (2005), 'Understanding New Zealand's changing income distribution, 1983–1998: A semi-parametric analysis', *Economica* **72**(287), 469–495.

Jenkins, S. P. (1996), 'Recent trends in the UK income distribution: What happened and why?', *Oxford Review of Economic Policy* **12**(1), 29–46.

References V

Jenkins, S. P. and Van Kerm, P. (2005), ‘Accounting for income distribution trends: A density function decomposition approach’, *Journal of Economic Inequality* **3**(1), 43–61.
URL: *http://dx.doi.org/10.1007/s10888-004-8309-1*

Koenker, R. and Bassett, G. (1978), ‘Regression quantiles’, *Econometrica* **46**(1), 33–50.
URL: *http://www.jstor.org/stable/1913643*

Larrimore, J. (2014), ‘Accounting for United States household income inequality trends: The changing importance of household structure and male and female labor earnings inequality’, *Review of Income and Wealth* **60**(4), 683–701.

Lerman, R. I. and Yitzhaki, S. (1985), ‘Income inequality effects by income source: A new approach and applications to the United States’, *Review of Economics and Statistics* **67**(1), 151–156.

Machado, J. A. F. and Mata, J. (2005), ‘Counterfactual decomposition of changes in wage distributions using quantile regression’, *Journal of Applied Econometrics* **20**, 445–465.

Melly, B. (2005), ‘Decomposition of differences in distribution using quantile regression’, *Labour economics* **12**(4), 577–590.
URL: *http://www.sciencedirect.com/science/article/pii/S0927537105000382*

Mookherjee, D. and Shorrocks, A. F. (1982), ‘A decomposition analysis of the trend in UK income inequality’, *Economic Journal* **92**(368), 886–902.

References VI

Morduch, J. and Sicular, T. (2002), 'Rethinking inequality decomposition, with evidence from rural China', *Economic Journal* **112**(476), 93–106.

Rothe, C. and Wied, D. (2013), 'Misspecification testing in a class of conditional distributional models', *Journal of the American Statistical Association* **108**(501), 314–324.

Shorrocks, A. F. (1982*a*), 'Inequality decomposition by factor components', *Econometrica* **50**(1), 193–211.
URL: *http://www.jstor.org/stable/1912537*

Shorrocks, A. F. (1982*b*), 'On the distance between income distributions', *Econometrica* **50**(5), 1337–1339.
URL: *http://www.jstor.org/stable/1911879*

Shorrocks, A. F. (1984), 'Inequality decomposition by population subgroups', *Econometrica* **52**(6), 1369–1385.

Shorrocks, A. F. (2013), 'Decomposition procedures for distributional analysis: a unified framework based on the Shapley value', *Journal of Economic Inequality* **11**(1), 99–126.
URL: *http://dx.doi.org/10.1007/s10888-011-9214-z*

Sologon, D. M., Van Kerm, P., Li, J. and O'Donoghue, C. (2021), 'Accounting for differences in income inequality across countries: tax-benefit policy, labour market structure, returns and demographics', *Journal of Economic Inequality* **19**(1), 13–43.

Van Kerm, P., Yu, S. and Choe, C. (2017), 'Decomposing quantile wage gaps: a conditional likelihood approach', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **65**(4), 507–527.
URL: *http://onlinelibrary.wiley.com/doi/10.1111/rssc.12137/pdf*