Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds

Richard Blundell, Amanda Gosling, Hidehiko Ichimura and Costas Meghir

Main Question

- In both the US and UK, the period between 1978-1999 witnessed:
  1. Large changes in the distribution of observed wages: gender wage gap, skill premium and within group inequality.
  2. Shifts in workforce composition: gender, educational...
- How to recover the latent distribution of wages, using the distribution of observed wages, when the selection rule might be time varying?
- Point-Estimates of Wage Distribution:
  - Methods available to control for selection effects require very strong assumptions, e.g. exclusion restrictions, or normality.
- Using UK data, non-parametrically estimate *bounds* for the distribution of wages.

Worst Case Bounds

- Take W to be the log wage, X is a conditioning vector.
- Employment indicator $E = 1$ if W observed, and $E = 0$ otherwise.
- $P(x) = Prob(E = 1|x)$
- Conditional distribution of wages, given $x: \; F(w|x)$
- Decompose unobserved object of interest $F(w|x)$ into:

$$F(w|x) = F(w|x, E = 1)P(x) + F(w|x, E = 0)[1 - P(x)]$$

- Manski (1994) noted that $0 \leq F(w|x, E = 0) \leq 1 \Rightarrow$

$$F(w|x, E = 1)P(x) \leq F(w|x) \leq F(w|x, E = 1)P(x) + [1 - P(x)]$$

BOUNDS TO QUANTILES

- Let $w^q(x)$ be the q^{th} quantile of $F(w|x)$
- Define the bounds to the conditional quantile as:

$$w^{q(l)}(x) \leq w^q(x) \leq w^{q(u)}(x)$$

where...

- $w^{q(l)}(x)$ is the wage (w) that solves the equation:

$$q = F(w|x, E = 1)P(x) + [1 - P(x)]$$

- $w^{q(u)}(x)$ is the wage (w) that solves the equation:

$$q = F(w|x, E = 1)P(x)$$

Assumption: Positive Selection

1. **Stochastic Dominance:**

$$F(w|x, E=1) \leq F(w|x, E=0), \quad \forall w, \ \forall x$$

   this will imply a higher lower bound:

$$F(w|x, E=1) \leq F(w|x) \leq \ldots$$

2. **Median Restriction:** the median wage offer for those not working is not higher than the median *observed* wage
   (a) The bounds for wages below the median observed wage are still the worst case bounds
   (b) But the lower bound for all wages above the median is lifted to

$$F(w|x, E=1)P(x) + 0.5(1 - P(x)) \leq F(w|x) \leq \ldots$$

Instrumental Variable Z

- **Exclusion Restriction:** W independent of Z conditional on X

$$F(w|x,z) = F(w|x), \quad \forall w, x, z$$

(a) The bounds are given by:

$$\max_z\{F(w|x,z,E=1)P(x,z)\} \leq F(w|x)$$
$$\leq \min_z\{F(w|x,z,E=1)P(x,z) + [1-P(x,z)]\}$$

(b) Intuitively, underlying wage distribution does not move with Z but the *observed* one will (through the participation rate).

$$F(w|x,z,E=1)P(x,z) = F(w|x)\text{Prob}(E=1|W \leq w, x, z)$$

(c) The ”strength” of the instrument comes through its effect on participation.

INSTRUMENTAL VARIABLE Z

- **Monotonicity:** first-order stochastic dominance by the distribution with higher values of Z

$$F(w|x,z') \leq F(w|x,z), \quad \forall w,x,z,z' \text{ with } z < z'$$

  (a) In this case, the bounds tighten through the instrument's effect on the underlying distribution itself.
  (b) In practice, find the tightest bounds over the support of Z, and then integrate Z out.

- Validity Test: nothing guarantees *lower bound* $\leq$ *upper bound* if monotonicity or exclusion restriction assumptions are not true.

Estimation Method

- Data: UK Family Expenditure Survey (1978 to 1999)
- X: gender, education, age and time
    - Compute participation rate for each cell
    - Estimate wage distribution for each cell x_k:

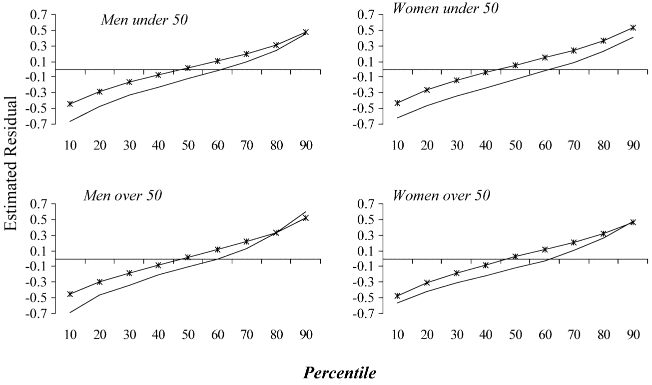
$$\hat{F}(w|E_1 = 1, x_k) = \frac{\sum_{i=1}^{N} \Phi((w - w_i)/(\sigma_w/5))\mathbb{I}_{(E_i=1)}\kappa_k(x_i)}{\sum_{i=1}^{N} \mathbb{I}_{(E_i=1)}\kappa_k(x_i)}$$

$$\kappa_k(x_i) = \mathbb{I}_{(year_i=year_k)}\mathbb{I}_{(ed_i=ed_k)}\mathbb{I}_{(gender_i=gender_k)}\mathbb{I}_{(age_i \in age_k)}$$

- Z: out-of-work income (unemployment benefits)
    - Same as above, but this time, each cell characterized by (gender, education level, age bracket, year bracket, benefits bracket)
- Then... compute bounds accordingly, given assumptions.
- Inequality measure: Interquartile Range

VALIDITY OF RESTRICTIONS

- Positive Selection (Stochastic Dominance and Median Restrictions)
  - Using British Household Panel Survey (1991-2001)
  - Controlling for education and age



(*) Always Work (-) With Spells out of the Labor Market

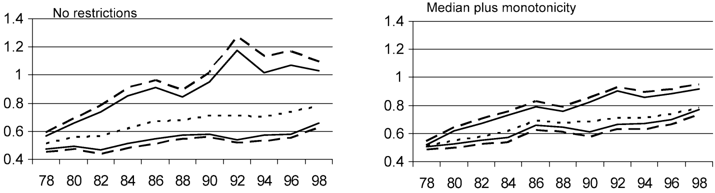
Validity of Restrictions

- Exclusion Restriction: Rejected
  - The upper and lower bounds cross for many groups, when this is imposed
  - Welfare benefits positively related to housing costs in the UK: high wage people → expensive housing → higher benefits.
  - Thus, dependence between wages and benefits.
- Monotonicity: Never Rejected
- Most of the results use combination of

  Monotonicity Assumption and Median Restriction

Overall Inequality

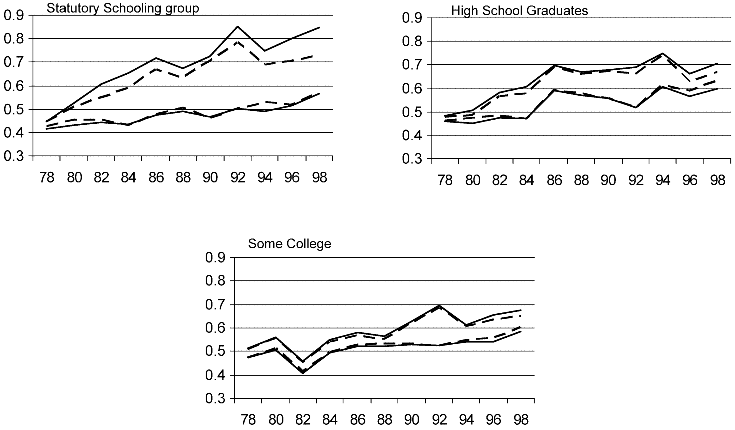
- Focus: $LowerBound_{98} - UpperBound_{78}$ (Solid Lines)



  - No Restrictions: inequality increased by 0.089 log points.
  - Median + Monotonicity: inequality increased by 0.252 log pts.
  - *Observed* inequality (Dotted Line) increased by 0.268 log points.
- *Latent* Inequality increased by at least almost as much as *Observed* inequality.

Increased Within Group Inequality

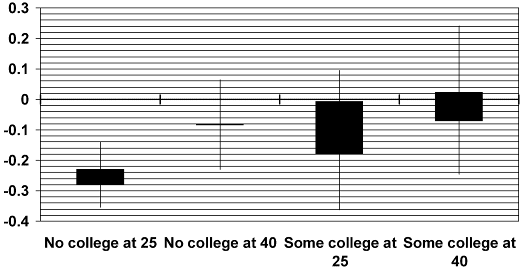
[Solid: Monotonicity] [Dashed: Mono + Median]



Statutory Schooling: 0.076 (left full-time edu at or before 16),
HS: 0.108 (completed edu between ages 17 and 18),
College: 0.129 (completed full time edu after 18)

Gender Wage Differentials

- Monotonicity + Median Restriction
- 2 Groups Only: No College and (Some) College
- Changes in the male/female differential:



  - Only lowest education category showed underlying improvement in the distribution
  - Composition effects actually lead gender gap to close by *less* than it would have otherwise