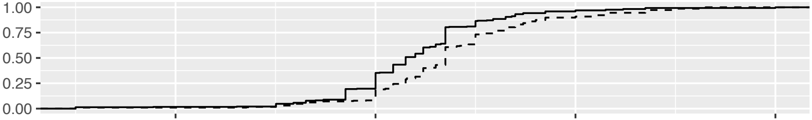
Lecture 18: Distributional Effects

POL-GA 1251
Quantitative Political Analysis II
Prof. Cyrus Samii
NYU Politics

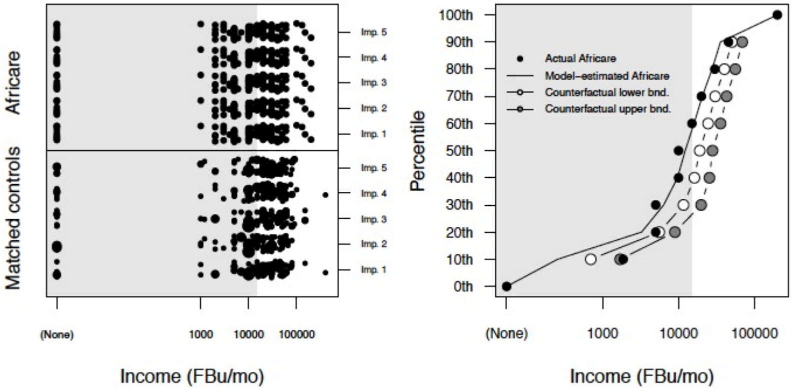
April 11, 2022

Overview

- Focus up to now on *average* causal effects (or conditional average effects).
- Substantive and technical motivation:
  - Linear social welfare functions allowing for transfers.
  - Allows for reliable inference.
- Other kinds of effects may be more informative.
- Today we discuss *distributional* effects: effects defined in terms of arbitrary features of marginal potential outcome distributions: F_{Y_1} vs. F_{Y_0}.



Motivating Example



Source: Gilligan et al. 2012

(The five replicates on left are multiple-imputation completed replicates of the sample.)

Approaches to distributional effects

We will review two regression-based approaches (MHE, Ch. 7; Chernozhukov et al. 2013):

- Quantile regression.
- Distribution regression.

CDFs and quantiles

- Cumulative distribution functions (CDFs) characterize the entirety of random variable distributions.
- Quantiles are numbers on scale of Y that map back from values for the CDF of Y.

CDFs and quantiles

- Cumulative distribution functions (CDFs) characterize the entirety of random variable distributions.
- Quantiles are numbers on scale of Y that map back from values for the CDF of Y.
- "quantile at .9 point in CDF of Y" = "9th decile" value = "90% percentile" value = value of Y such that 90% of observations are less than Y.

CDFs and quantiles

- Cumulative distribution functions (CDFs) characterize the entirety of random variable distributions.
- Quantiles are numbers on scale of Y that map back from values for the CDF of Y.
- "quantile at .9 point in CDF of Y" = "9th decile" value = "90% percentile" value = value of Y such that 90% of observations are less than Y.
- CDF at y: $F_Y(y) = \Pr(Y \leq y) = \mathrm{E}[I(Y \leq y)]$.
- Quantile function: quantile at τ in CDF of Y is

$$Q_\tau(Y) = F_Y^{-1}(\tau) = \inf\{y : F_Y(y) \geq \tau\}.$$

CDFs and quantiles

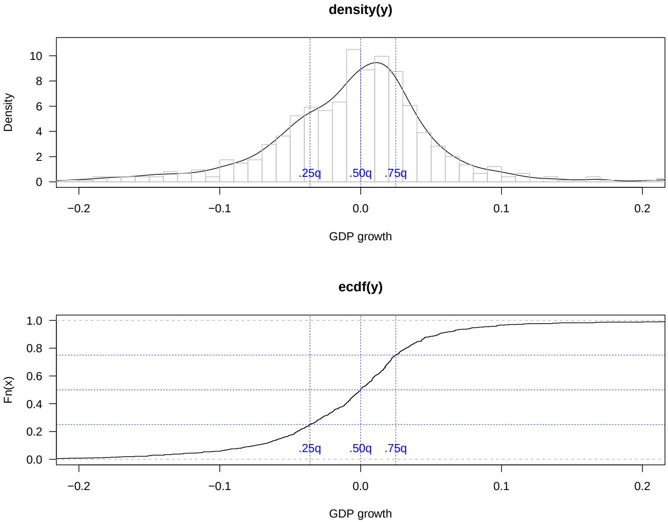
- Cumulative distribution functions (CDFs) characterize the entirety of random variable distributions.
- Quantiles are numbers on scale of Y that map back from values for the CDF of Y.
- "quantile at .9 point in CDF of Y" = "9th decile" value = "90% percentile" value = value of Y such that 90% of observations are less than Y.
- CDF at y: $F_Y(y) = \Pr(Y \leq y) = \mathrm{E}\left[I(Y \leq y)\right]$.
- Quantile function: quantile at τ in CDF of Y is

$$Q_\tau(Y) = F_Y^{-1}(\tau) = \inf\{y : F_Y(y) \geq \tau\}.$$

- Conditional CDF:
  $F_{Y|X}(y|x) = \Pr(Y \leq y|X = x) = \mathrm{E}\left[I(Y \leq y)|X = x\right]$.
- Conditional quantile: for units with $X = x$, quantile at τ in conditional CDF of Y is

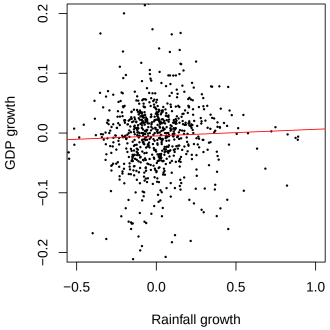
$$Q_\tau(Y|X = x) = F_{Y|X}^{-1}(\tau|X = x) = \inf\{y : F_{Y|X}(y|X = x) \geq \tau\}.$$

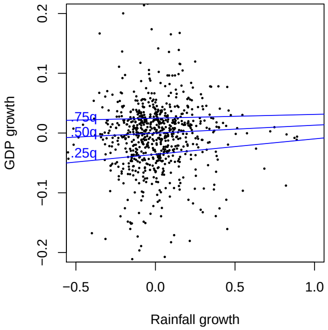
CDFs and quantiles



(Source: Miguel et al., 2004)

Conditional quantiles





(Source: Miguel et al., 2004)

Quantile regression model

- So far, we have focused on conditional expectations functions:

$$\mathrm{E}\left[Y_i|X_i\right] = \int_y y dF_{Y|X}(y|X_i)$$

which solve the mean-squared error prediction problem:

$$\mathrm{E}\left[Y_i|X_i\right] = \arg\min_{m(X_i)} \mathrm{E}\left[(Y_i - m(X_i))^2\right].$$

Quantile regression model

- So far, we have focused on conditional expectations functions:

$$\mathrm{E}\left[Y_i|X_i\right] = \int_y y dF_{Y|X}(y|X_i)$$

  which solve the mean-squared error prediction problem:

$$\mathrm{E}\left[Y_i|X_i\right] = \arg\min_{m(X_i)} \mathrm{E}\left[(Y_i - m(X_i))^2\right].$$

- Linear regression is a minimum expected loss problem, with loss being mean square error.
- Sample analogue:

$$\hat{\mathrm{E}}\left[Y_i|X_i\right] = \arg\min_{m(X_i)} \frac{1}{N}\sum_{i=1}^{N}(Y_i - m(X_i))^2$$

- Linear specification:

$$\hat{\beta}_{OLS} = \arg\min_{\beta} \frac{1}{N}\sum_{i=1}^{N}(Y_i - X_i'\beta)^2$$

Quantile regression model

- We can also define a conditional quantile function for τ in conditional CDF of Y:

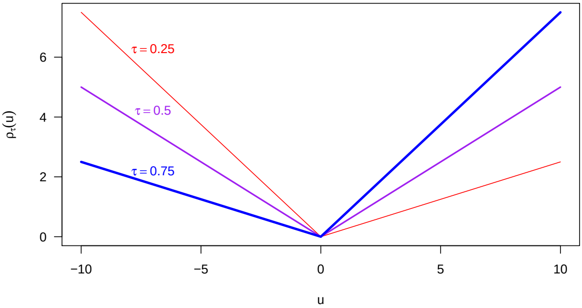
$$Q_\tau(Y_i|X_i) = F^{-1}_{Y|X}(\tau|X_i)$$

  which can be shown to solve another minimization problem:

$$Q_\tau(Y_i|X_i) = \arg\min_{q(X_i)} \mathrm{E}\left[\rho_\tau(Y_i - q(X_i))\right]$$

  with $\rho_\tau(u) = u(\tau - 1(u \leq 0))$, the “check” function.
- Minimize expected loss, with loss given by mean “check loss”.

Quantile regression model



- For median ($\tau = .5$), minimizing wrt $\rho_{.5}(u)$ amounts to minimizing expected absolute deviations:

$$\rho_{.5}(u) = u(.5 - 1(u \leq 0)) = \begin{cases} .5u & \text{if } u > 0 \\ -.5u & \text{if } u \leq 0 \end{cases}$$

- 0.75 quantile: penalize deviations from $q(X_i)$ to right more.
- 0.25 quantile: penalize deviations from $q(X_i)$ to left more.

Quantile regression model

- Sample analogue (assuming equal probability sample):

$$\hat{Q}_\tau(Y_i|X_i) = \arg\min_{q(X_i)} \frac{1}{N}\sum_{i=1}^{N}\rho_\tau(Y_i - q(X_i))$$

(easy to see how weighting could be accommodated)

Quantile regression model

- Sample analogue (assuming equal probability sample):

$$\hat{Q}_\tau(Y_i|X_i) = \arg\min_{q(X_i)} \frac{1}{N}\sum_{i=1}^{N}\rho_\tau(Y_i - q(X_i))$$

  (easy to see how weighting could be accommodated)

- Linear specification:

$$\hat{\beta}_\tau = \arg\min_{\beta_\tau} \frac{1}{N}\sum_{i=1}^{N}\rho_\tau(Y_i - X_i'\beta_\tau)$$

- $x'\hat{\beta}_\tau$ predicts quantile corresponding to τ in conditional CDF of Y for units with $X_i = x$.

Quantile regression model

- Sample analogue (assuming equal probability sample):

$$\hat{Q}_\tau(Y_i|X_i) = \arg\min_{q(X_i)} \frac{1}{N}\sum_{i=1}^{N} \rho_\tau(Y_i - q(X_i))$$

  (easy to see how weighting could be accommodated)
- Linear specification:

$$\hat{\beta}_\tau = \arg\min_{\beta_\tau} \frac{1}{N}\sum_{i=1}^{N} \rho_\tau(Y_i - X_i'\beta_\tau)$$

- $x'\hat{\beta}_\tau$ predicts quantile corresponding to τ in conditional CDF of Y for units with $X_i = x$.
- Trace out entire distribution of Y_i using $\tau = .1, .2, .3, \ldots, .9$.

Quantile regression model

- Sample analogue (assuming equal probability sample):

$$\hat{Q}_\tau(Y_i|X_i) = \arg\min_{q(X_i)} \frac{1}{N}\sum_{i=1}^{N}\rho_\tau(Y_i - q(X_i))$$

(easy to see how weighting could be accommodated)

- Linear specification:

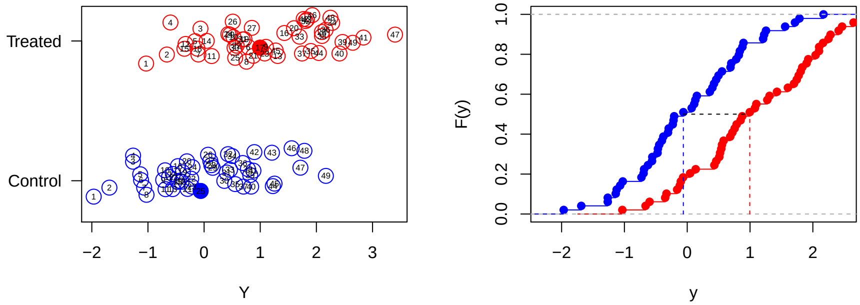
$$\hat{\beta}_\tau = \arg\min_{\beta_\tau} \frac{1}{N}\sum_{i=1}^{N}\rho_\tau(Y_i - X_i'\beta_\tau)$$

- $x'\hat{\beta}_\tau$ predicts quantile corresponding to τ in conditional CDF of Y for units with $X_i = x$.
- Trace out entire distribution of Y_i using $\tau = .1, .2, .3, \ldots, .9$.
- No closed form solution and $Q_\tau(.)$ not everywhere differentiable, so estimates found via linear programming (e.g. simplex) methods (cf. Koenker, 2000).

Quantile regression model

- $\hat{\beta}_\tau$ asymptotically normal under a range of conditions (Mosteller, 1946).
- Standard errors and confidence intervals are typically obtained via bootstrap.

Quantile regression model



- When estimating causal effects, the quantile regression coefficient estimates changes in the *overall outcome distribution*.
- Not necessarily “effect for unit at τ quantile.”
- Latter only holds under rank invariance.

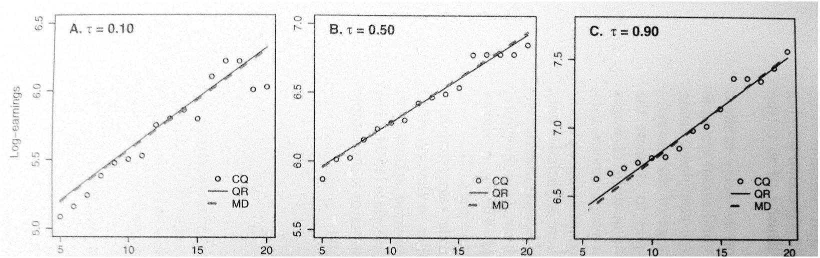
Quantile regression model

As was the case with OLS, we have (MHE p. 272):

> *If $Q_\tau(Y_i|X_i)$ is in fact linear, the quantile regression minimand will find it....As it turns out, however,, the assumption of a linear CQF is unnecessary: quantile regression is useful [for describing conditional quantiles over X_i] whether or not we believe this.*

Of course if the predictors are dummy variables, the point is moot. MHE, p. 277-28 explain that the quantile regression solution minimizes something that comes very close to the expected squared prediction error for the true quantiles:

Quantile regression model



(MHE, p. 280. *X*-axis is years of schooling. Data are from 1980 census.)

An application

Let's return to the Gilligan et al. application:

An application

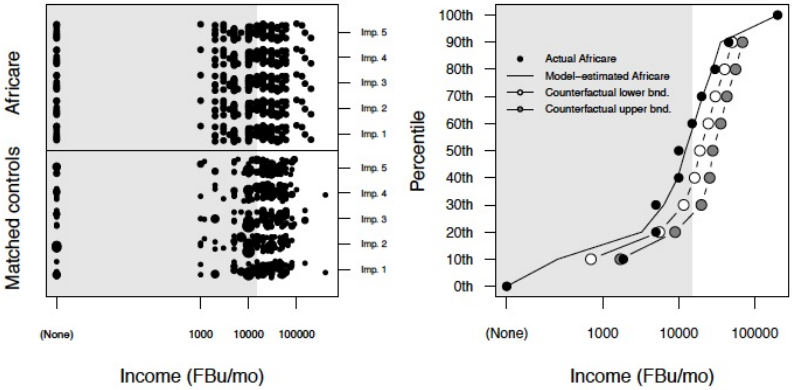
Table 5: **Estimates from quantile regressions on log(income/month+1)**

	No adjustment			Matching & regression			Matching, regression, & het. exposure adjustment		
	coef.	s.e.	*p*-val.	coef.	s.e.	*p*-val.	coef.	s.e.	*p*-val.
Decile 1	-0.92	2.89	0.75	-1.01	2.24	0.65	-1.90	4.22	0.65
Decile 2	-0.69	0.24	0.00	-0.54	0.98	0.59	-1.01	1.85	0.59
Decile 3	-0.69	0.16	0.00	-0.60	0.31	0.06	-1.14	0.58	0.06
Decile 4	-0.41	0.17	0.02	-0.51	0.28	0.07	-0.97	0.52	0.07
Decile 5	-0.64	0.17	0.00	-0.44	0.24	0.07	-0.83	0.45	0.07
Decile 6	-0.60	0.21	0.01	-0.43	0.18	0.02	-0.80	0.34	0.02
Decile 7	-0.41	0.14	0.00	-0.39	0.18	0.03	-0.74	0.33	0.03
Decile 8	-0.51	0.18	0.00	-0.37	0.22	0.11	-0.70	0.42	0.11
Decile 9	-0.25	0.17	0.14	-0.35	0.21	0.10	-0.66	0.40	0.10
N from imputed datasets	371,371,371,371,371			177,177,181,178,177			177,177,181,178,177		

Standard errors computed using robust inverted rank test intervals. Regressions on matched data include the covariates in Table 4. The coefficients on these covariates are not displayed to save space.

Program does more to shift incomes at lower end of distribution. Therefore, program reduces inequality by bringing up lower end of distribution more than it increases upper end. This implies that the program results in a reduction in the variance of incomes.
(Gilligan et al. 2012)

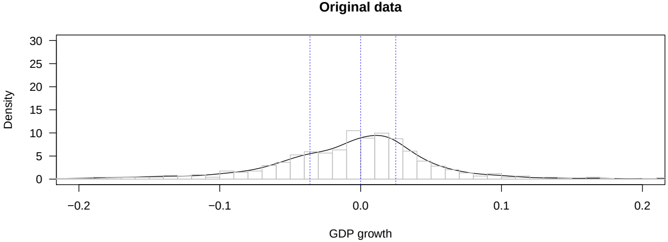
An application

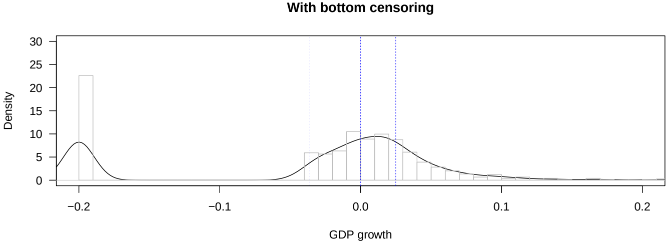


Reduction in inequality and variance evident in predicted CDFs being *steeper* in the treated relative to the control (Africare) condition. (Gilligan et al. 2012)

Censored data

Quantiles are insensitive to bottom or top censoring:





Censored data

Applying quantile regression to censored data:

- Suppose observed Y_i are top censored such that,

$$Y_{i,obs} = Y_i \cdot 1(Y_i < c) + c \cdot 1(Y_i \geq c)$$

  (e.g., censored duration data or top-coded income data).
- While we know quantiles for the unconditional distribution of Y_i (up to c), we may not know which *conditional* quantiles are pushed above the censoring point.
- Of course, if less than c proportion of data are censored conditional on all values of X_i, you can estimate conditional quantile function up to the $1 - p$ quantile.
- Otherwise you need to modify the estimation algorithm.

Censored data

- What we do is to define a linear conditional quantile function that *itself* is censored at c:

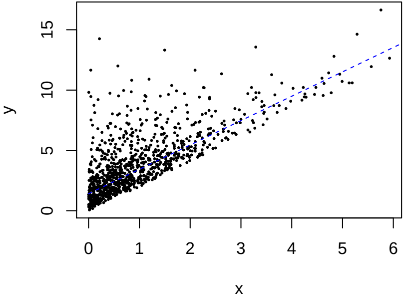
$$Q_{\tau,c}(Y_i|X_i) = \min(c, X_i'\beta_\tau^c).$$

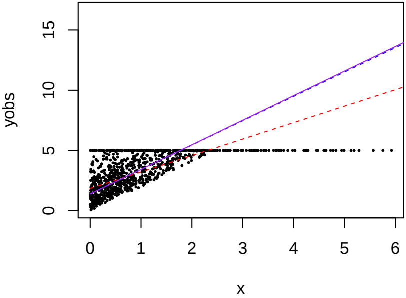
and estimate β_τ^c from,

$$\beta_\tau^c = \arg\min_b \mathrm{E}\left[1(X_i'\beta_\tau^c < c) \cdot \rho_\tau(Y_i - X_i'b)\right]$$

- The recursiveness of the expression suggests an iterated fitting algorithm (Buchinsky 1994).
- Generalizations to random censoring (Koenker 2008).

Censored data





Quantile IV

Recall LATE theorem set up:

- "Compliers" randomly induced by Z_i into treatment, D_i.
- Thus:

$$(Y_{1i}, Y_{0i}) \perp\!\!\!\perp D_i | X_i, \underbrace{D_{1i} > D_{0i}}_{\text{complier}}.$$

- Would be nice if we could estimate treatment effects using thesubsample of compilers.
- We don't know who is a complier, so we use indirect method (2SLS).

Quantile IV

- Quantile regression function for compliers would be,

$$Q_\tau(Y_i|X_i, D_i, D_{1i} > D_{0i}) = \alpha_\tau^c D_i + X_i'\beta_\tau$$

  where α_τ^c is the complier ATE.
- Model implies

$$Q_\tau(Y_{1i}|X_i, D_{1i} > D_{0i}) - Q_\tau(Y_{0i}|X_i, D_{1i} > D_{0i}) = \alpha_\tau^c$$

Quantile IV

- Estimation uses Abadie's κ weighting:

$$
\begin{aligned}
(\alpha_\tau^c, \beta_\tau^c) &= \arg\min_{a,b} \mathrm{E}\left[\rho_\tau(Y_i - aD_i - X_i'b) | D_{1i} > D_{0i}\right] \\
&= \arg\min_{a,b} \mathrm{E}\left[\kappa_i \rho_\tau(Y_i - aD_i - X_i'b)\right],
\end{aligned}
$$

  with

$$
\kappa_i = 1 - \frac{D_i(1-Z_i)}{1 - \Pr[Z_i = 1|X_i]} - \frac{(1-D_i)Z_i}{\Pr[Z_i = 1|X_i]}
$$

- Implement in Stata or R:
  - Estimate κ_i's.
  - Fit outcome quantile regression with weights equal to κ_i.
  - Bootstrap.

Distribution regression

Chernozhukov et al. (2013) discuss distribution regression as alternative to quantile regression:

Distribution regression

Chernozhukov et al. (2013) discuss distribution regression as alternative to quantile regression:

- Recall $F(y|x) = \Pr(Y \leq y|X = x) = \mathrm{E}\left[I(Y \leq y)|X = x\right]$.

Distribution regression

Chernozhukov et al. (2013) discuss distribution regression as alternative to quantile regression:

- Recall $F(y|x) = \Pr(Y \leq y|X = x) = \mathrm{E}[I(Y \leq y)|X = x]$.
- Estimate effects on distn. at y by modeling $\mathrm{E}[I(Y \leq y)|X = x]$.

Distribution regression

Chernozhukov et al. (2013) discuss distribution regression as alternative to quantile regression:

- Recall $F(y|x) = \Pr(Y \leq y|X = x) = \mathrm{E}[I(Y \leq y)|X = x]$.
- Estimate effects on distn. at y by modeling $\mathrm{E}[I(Y \leq y)|X = x]$.
- With binary $D = 0, 1$ could use, $\mathrm{E}[I(Y \leq y)|D] = \beta_0 + \beta_1 D$.

Distribution regression

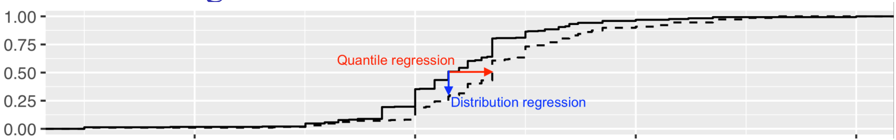
Chernozhukov et al. (2013) discuss distribution regression as alternative to quantile regression:

- Recall $F(y|x) = \Pr(Y \leq y|X = x) = \mathrm{E}[I(Y \leq y)|X = x]$.
- Estimate effects on distn. at y by modeling $\mathrm{E}[I(Y \leq y)|X = x]$.
- With binary $D = 0, 1$ could use, $\mathrm{E}[I(Y \leq y)|D] = \beta_0 + \beta_1 D$.
- To estimate effects on the whole distribution, simply define a series of evaluation values $y_1, \ldots, y_P$, and estimate a series of effects:

$$\mathrm{E}[I(Y \leq y_1)|D] = \beta_0^{(1)} + \beta_1^{(1)} D + \varepsilon^{(1)}$$
$$\vdots$$
$$\mathrm{E}[I(Y \leq y_P)|D] = \beta_0^{(P)} + \beta_1^{(P)} D + \varepsilon^{(P)}$$

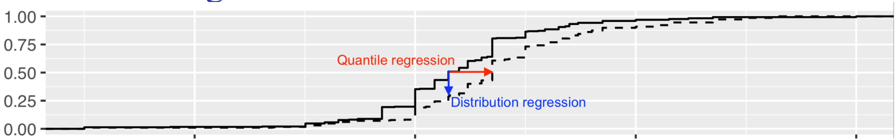
- With continuous predictors, may want to use a limited dependent variable model for the distribution regression (e.g, logit — more on that later in the semester).

Distribution regression



- $Q_{.5}(Y|D) = \alpha_0 + \alpha_1 D$: how does the median change when treated?
- $F(y|D) = \beta_0 + \beta_1 D$: how does $CDF(y)$ change when treated?
- (Maybe more intuitive is $1 - F(y|D) = \mathrm{E}[I(Y > y)|X = x]$.)

Distribution regression



- $Q_{.5}(Y|D) = \alpha_0 + \alpha_1 D$: how does the median change when treated?
- $F(y|D) = \beta_0 + \beta_1 D$: how does $CDF(y)$ change when treated?
- (Maybe more intuitive is $1 - F(y|D) = \mathrm{E}[I(Y > y)|X = x]$.)
- Either approach can be used to model distributional effects.
- With saturated specifications, the two coincide (Chernozhukov et al. 2013), but not otherwise.
- DR can be more robust when outcomes are clumpy or mixed continuous-discrete.
- Straightforward to use various identification strategies to estimated DR.
- Specification test for distribution models like QR and DR: Rothe and Wied (2013, JASA).

Distribution regression

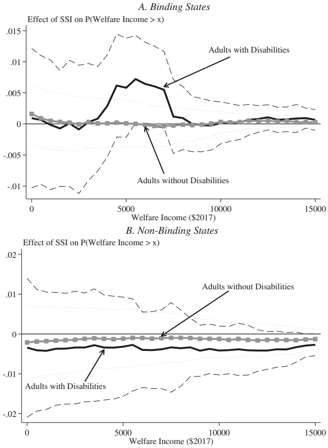


Fig. 6. The effect of SSI on the distribution of welfare income for adults with and without disabilities, 1970 and 1980 censuses. Notes: This figure plots coefficient estimates from difference-in-differences specifications like Eq. (1) but with only two time periods: the 1970 and 1980 censuses. Each point comes from a regression whose outcome is the share of adults, with disabilities (panel A) or without disabilities (panel B), in state *s* in year *t* who report welfare income greater than *x*, indicated on the horizontal axis. The disability distinction comes from the self-reported work-limiting-disability question. We find no effect of SSI on the probability of reporting a disability, which supports stratifying by disability status. Source: Ruggles et al. (2010) and DHEW (1972).

(Goodman-Bacon & Schmidt 2020)

Remarks

- Methods discussed today allow us to go beyond average treatment effects to consider effects on variance, levels of inequality, quantile effects, and other features of distributions.
- E.g., treatments may leave averages untouched but result in changes in inequality levels. OLS may not pick that up; distributional methods will.
- These methods are also helpful for irregular outcomes (e.g., truncated/censored).
- Estimation and inference methods are straightforward in R and Stata.