Applied Microeconometrics

Maximilian Kasy

7) Distributional Effects, quantile regression

(cf. “Mostly Harmless Econometrics,” chapter 7)

Sir Francis Galton (*Natural Inheritance*, 1889):

"It is difficult to understand why statisticians commonly limit their inquiries to Averages, and do not revel in more comprehensive views. Their souls seem as dull to the charm of variety as that of the native of one of our flat English counties, whose retrospect of Switzerland was that, if its mountains could be thrown into its lakes, two nuisances would be got rid of at once."

Distributional Effects

- Most empirical research on treatment effects focuses on the estimation of differences in mean outcomes
- But methods exists for estimating the impact of a treatment on the entire distribution of outcomes:
  - Does the intervention increase inequality?
  - Does the intervention affect the distribution at all?
  - Stochastic dominance?
- Methods for estimating distributional effects:
  - Experiments: Compare the distributions of Y_0 and Y_1
  - Selection on observables: Quantile Regression
  - Experiments with Non-compliance: Instrumental Variable Quantile Regression

Distributional Effects

- In an experiment with perfect compliance: $Y_1, Y_0 \perp\!\!\!\perp D$.
- To evaluate distributional effects in a randomized experiment, we can compare the distribution of the outcome for treated and untreated:

$$\begin{aligned} F_{Y_1}(y) &= \Pr(Y_1 \leq y) = \Pr(Y_1 \leq y | D = 1) = \Pr(Y \leq y | D = 1) \\ &= F_{Y|D=1}(y). \end{aligned}$$

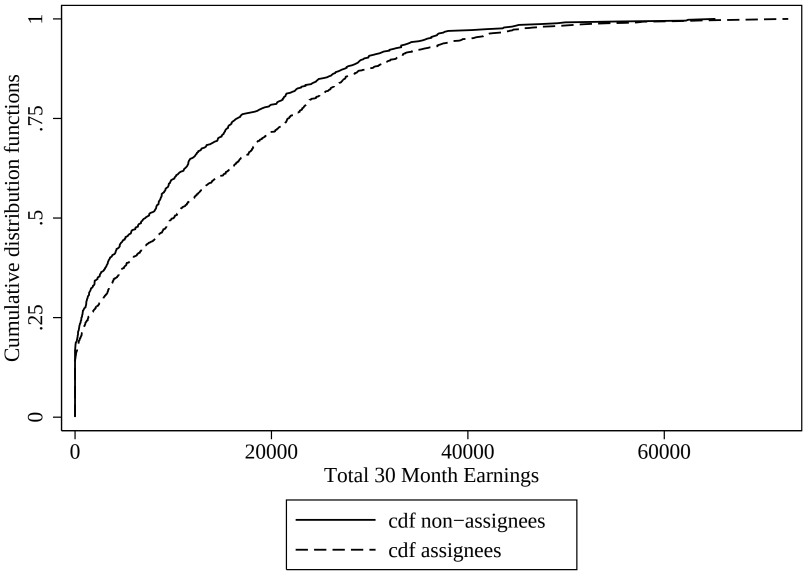
  Similarly,

$$F_{Y_0}(y) = F_{Y|D=0}(y).$$

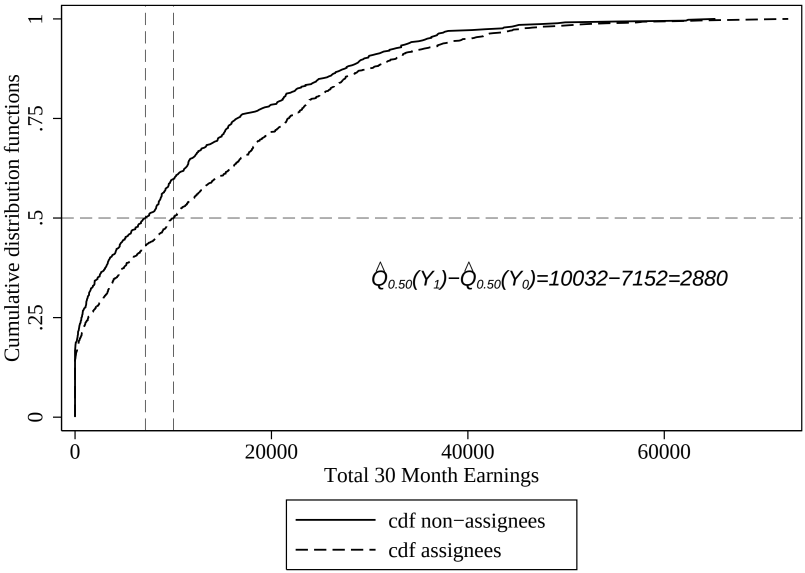
- We can use estimators:

$$\widehat{F}_{Y_1}(y) = \frac{1}{N_1} \sum_{D_i=1} 1\{Y_i \leq y\}, \qquad \widehat{F}_{Y_0}(y) = \frac{1}{N_0} \sum_{D_i=0} 1\{Y_i \leq y\}.$$

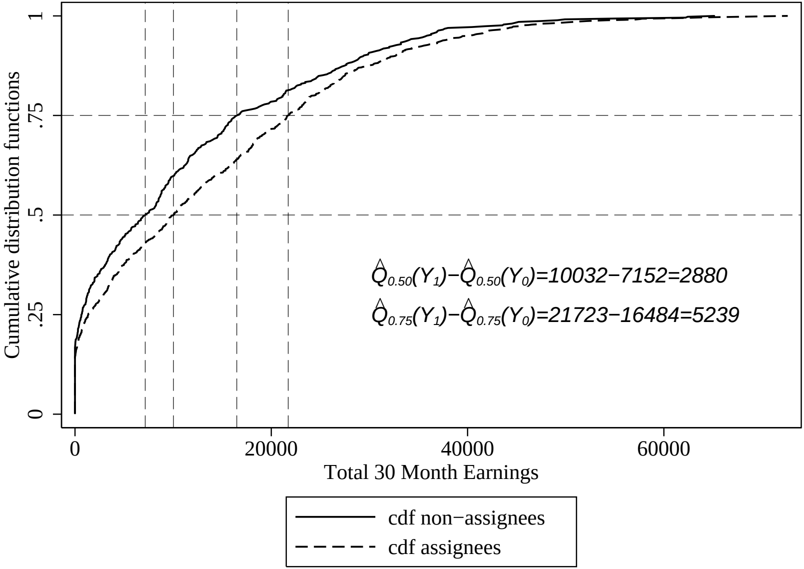
Adult Women in JTPA National Study (other services)



Adult Women in JTPA National Study (other services)



Adult Women in JTPA National Study (other services)



The "Social Welfare Function"

- Sometimes we want to compare the entire distributions, F_{Y_0} and F_{Y_1}
- "Social Welfare Function":

$$W(u, F) = \int_0^\infty u(y) dF(y),$$

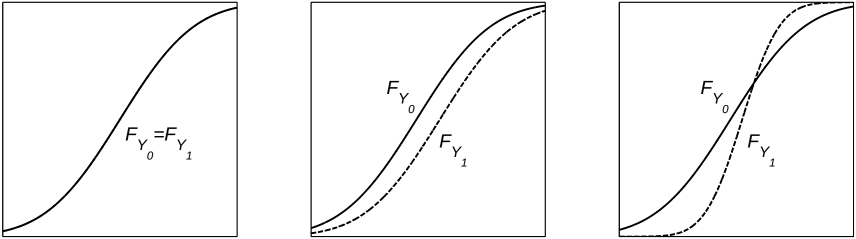
  where y is income, $u(y)$ is utility, and F is the income distribution.
- We want to know if

$$W(u, F_{Y_1}) \geq W(u, F_{Y_0}).$$

- The problem is that $u(y)$ is typically left unspecified
- However, we usually assume
  1. $u' \geq 0$ (utility is increasing in income)
  2. $u'' \leq 0$ (utility is concave $\Rightarrow$ preference for redistribution)

Distributional Effects

Equality (EQ) | First order stochastic dominance (FSD) | Second order stochastic dominance (FSD)



$F_{Y_1}(y) = F_{Y_0}(y)$ $\quad$ $F_{Y_1}(y) \leq F_{Y_0}(y)$ $\quad$ $\int_{-\infty}^{y} F_{Y_1}(x)dx \leq \int_{-\infty}^{y} F_{Y_0}(x)dx$

1. *EQ implies* $W(u, F_{Y_1}) = W(u, F_{Y_0})$ *for all* u
2. *FSD implies* $W(u, F_{Y_1}) \geq W(u, F_{Y_0})$ *for* $u' \geq 0$
3. *SSD implies* $W(u, F_{Y_1}) \geq W(u, F_{Y_0})$ *for* $u' \geq 0$ *and* $u'' \leq 0$

Kolmogorov-Smirnov Test (EQ)

- Suppose that we have data from a randomized experiment. How can we test the null hypothesis $H_0 : F_{Y_1} = F_{Y_0}$?
- Kolmogorov-Smirnov Statistic:

$$T_{\text{eq}} = \left(\frac{N_1 N_0}{N} \right)^{1/2} \sup_y |\widehat{F}_{Y_1}(y) - \widehat{F}_{Y_0}(y)|$$

- If Y is continuous, then the distribution of T_{eq} under H_0 is known
- If Y is not continuous (e.g., positive probability at $Y = 0$), we can use a bootstrap test:
  1. Compute T_{eq} in the original sample
  2. Resample N_1 "treated" and N_0 "non-treated" from the pooled the samples of treated and non-treated. (In this way, we impose the null hypothesis that the distribution of Y is the same for the two groups.) Compute $T_{\text{eq},b}$ for these two samples.
  3. Repeat step 2 many (B) times.
  4. Calculate the p-value as:

$$p\text{-value} = \frac{1}{B} \sum_{b=1}^{B} 1\{T_{\text{eq},b} > T_{\text{eq}}\}$$

Kolmogorov-Smirnov Test (FSD and SSD)

- The Kolmogorov-Smirnov bootstrap test of EQ can be easily adapted for FSD and SSD, just by changing the test statistics
- For first order stochastic dominance:

$$T_{\mathrm{fsd}} = \left(\frac{N_1 N_0}{N}\right)^{1/2} \sup_y \left(\widehat{F}_{Y_1}(y) - \widehat{F}_{Y_0}(y)\right)$$

- For second order stochastic dominance:

$$T_{\mathrm{ssd}} = \left(\frac{N_1 N_0}{N}\right)^{1/2} \sup_y \int_{-\infty}^{y} (\widehat{F}_{Y_1}(x) - \widehat{F}_{Y_0}(x))dx$$

Conditioning on Covariates: Quantile Regression

Identification Assumption

1. *Assume that the θ-quantile of the distribution of Y given D and X is linear:*
$$Q_\theta(Y|D,X) = \alpha_\theta D + X'\beta_\theta.$$
2. *D is randomized or there is selection on observables*

Identification Result

$$(\alpha_\theta, \beta_\theta) = \mathrm{argmin}_{(\alpha,\beta)} E[\rho_\theta(Y - \alpha D - X'\beta)]$$

where $\rho_\theta(\lambda) = (\theta - 1\{\lambda < 0\})\lambda$ identifies α_θ, the effect of the treatment on the θ-quantile of the conditional distribution of the outcome variable:

$$\begin{aligned}
\alpha_\theta &= Q_\theta(Y|D=1,X) - Q_\theta(Y|D=0,X) \\
&= Q_\theta(Y_1|D=1,X) - Q_\theta(Y_0|D=0,X) \\
&= Q_\theta(Y_1|X) - Q_\theta(Y_0|X).
\end{aligned}$$

Conditioning on Covariates: Quantile Regression

Identification Assumption

1. *Assume that the θ-quantile of the distribution of Y given D and X is linear:*
$$Q_\theta(Y|D,X) = \alpha_\theta D + X'\beta_\theta.$$
2. *D is randomized or there is selection on observables*

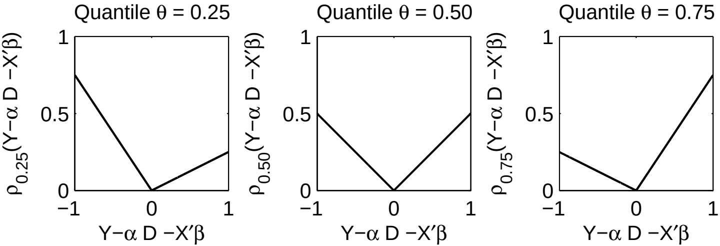
Estimator

The quantile regression estimator (Koenker and Bassett (1978)) is the sample analog:

$$(\widehat{\alpha}_\theta, \widehat{\beta}_\theta) = \mathrm{argmin}_{(\alpha,\beta)} \frac{1}{N}\sum_{i=1}^{N} \rho_\theta(Y_i - \alpha D_i - X_i'\beta)$$

Conditioning on Covariates: Quantile Regression

Recall:



For example:

$$(\widehat{\alpha}_{0.5}, \widehat{\beta}_{0.5}) = \operatorname{argmin}_{(\alpha,\beta)} \frac{1}{N} \sum_{i=1}^{N} |Y_i - \alpha D_i - X_i'\beta|$$

Kolmogorov-Smirnov Tests with Instrumental Variables

Identification Assumption

1. *Independence:* $(Y_0, Y_1, D_0, D_1) \perp\!\!\!\perp Z$
2. *First Stage:* $0 < P(Z = 1) < 1$ *and* $P(D_1 = 1) > P(D_0 = 1)$
3. *Monotonicity:* $D_1 \geq D_0 = 1$

Identification Result

For any function $h(\cdot)$ *(*$E|h(Y)| < \infty$*),*

$$\frac{E[h(Y)D|Z=1] - E[h(Y)D|Z=0]}{E[D|Z=1] - E[D|Z=0]} = E[h(Y_1)|D_1 > D_0],$$

$$\frac{E[h(Y)(1-D)|Z=1] - E[h(Y)(1-D)|Z=0]}{E[(1-D)|Z=1] - E[(1-D)|Z=0]} = E[h(Y_0)|D_1 > D_0].$$

Kolmogorov-Smirnov Tests with Instrumental Variables

Identification Result

Let

$$F_{Y_1|D_1>D_0}(y) = E[1\{Y_1 \le y\}|D_1 > D_0],$$
$$F_{Y_0|D_1>D_0}(y) = E[1\{Y_0 \le y\}|D_1 > D_0].$$

Apply result in previous slide with $h(Y) = 1\{Y \le y\}$ *to obtain:*

$$F_{Y_1|D_1>D_0}(y) = \frac{E[1\{Y \le y\}D|Z=1] - E[1\{Y \le y\}D|Z=0]}{E[D|Z=1] - E[D|Z=0]},$$

$$F_{Y_0|D_1>D_0}(y) = \frac{E[1\{Y \le y\}(1-D)|Z=1] - E[1\{Y \le y\}(1-D)|Z=0]}{E[(1-D)|Z=1] - E[(1-D)|Z=0]}.$$

- Sample counterparts can be used to estimate $F_{Y_1|D_1>D_0}(y)$ and $F_{Y_0|D_1>D_0}(y)$
- Tells us how treatment affects different parts of the outcome distribution for compliers
- Bootstrap tests for inference

Earning for Veterans and Non-veterans

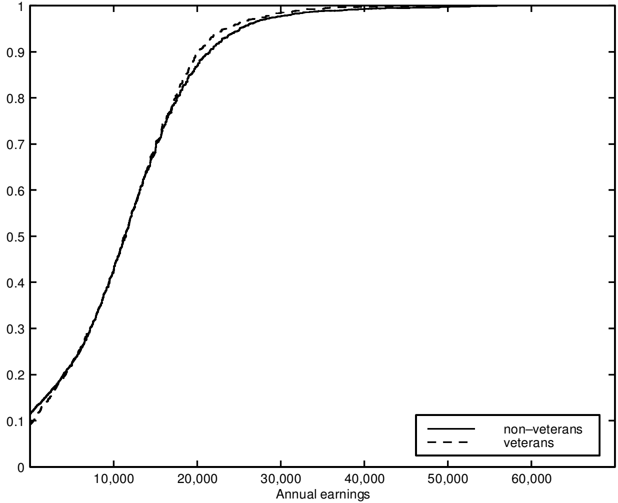


Figure 1. Empirical Distributions of Earnings for Veterans and Nonveterans.

Earning for Veterans and Non-veterans (Compliers)

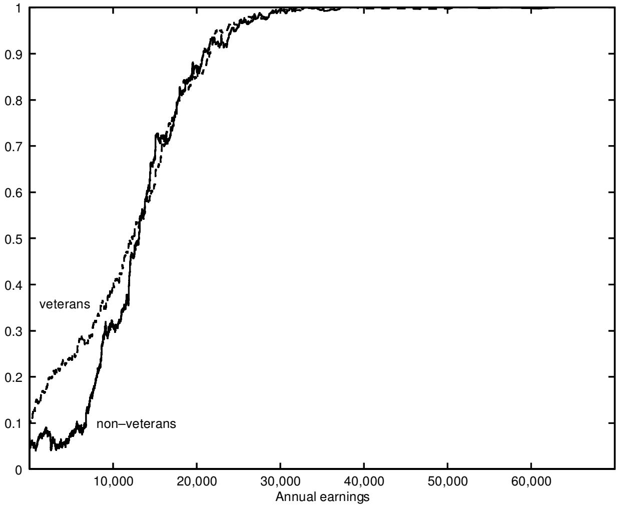


Figure 2. Estimated Distributions of Potential Earnings for Compliers.

Distributional Tests

Table 1. Tests on Distributional Effects of Veteran Status on Civilian Earnings, p-values

Outcome variable	*Equality in distributions*	*First-order stochastic dominance*	*Second-order stochastic dominance*
Annual earnings	.1245	.6260	.7415
Weekly wages	.2330	.6490	.7530

Quantile Regression with Instrumental Variables

Identification Assumption

1. *Conditional Independence of the Instrument:* $(Y_0, Y_1, D_0, D_1) \perp\!\!\!\perp Z | X$
2. *First Stage:* $0 < P(Z = 1|X) < 1$ *and* $P(D_1 = 1|X) > P(D_0 = 1|X)$
3. *Monotonicity:* $P(D_1 \geq D_0|X) = 1$

Estimate quantile regression for compliers:

$$Q_\theta(Y|D, X, D_1 > D_0) = \alpha_\theta D + X'\beta_\theta$$

Estimator

Using κ:

$$(\widehat{\alpha}_\theta, \widehat{\beta}_\theta) = \mathrm{argmin}_{(\alpha,\beta)} \sum_{i=1}^{N} \widehat{\kappa}_i \cdot \rho_\theta(Y_i - \alpha D_i - X_i'\beta).$$

JTPA: Quantile Regression

QUANTILE REGRESSION AND OLS ESTIMATES

Dependent Variable: 30-month Earnings

		Quantile				
	OLS	0.15	0.25	0.50	0.75	0.85
A. Men						
Training	3,754	1,187	2,510	4,420	4,678	4,806
	(536)	(205)	(356)	(651)	(937)	(1,055)
% Impact of Training	21.2	135.6	75.2	34.5	17.2	13.4
High school or GED	4,015	339	1,280	3,665	6,045	6,224
	(571)	(186)	(305)	(618)	(1,029)	(1,170)
Black	−2,354	−134	−500	−2,084	−3,576	−3,609
	(626)	(194)	(324)	(684)	(1,087)	(1,331)
Hispanic	251	91	278	925	−877	−85
	(883)	(315)	(512)	(1,066)	(1,769)	(2,047)
Married	6,546	587	1,964	7,113	10,073	11,062
	(629)	(222)	(427)	(839)	(1,046)	(1,093)
Worked less than 13	−6,582	−1,090	−3,097	−7,610	−9,834	−9,951
weeks in past year	(566)	(190)	(339)	(665)	(1,000)	(1,099)
Constant	9,811	−216	365	6,110	14,874	21,527
	(1,541)	(468)	(765)	(1,403)	(2,134)	(3,896)
B. Women						
Training	2,215	367	1,013	2,707	2,729	2,058
	(334)	(105)	(170)	(425)	(578)	(657)
% Impact of Training	18.5	60.8	44.4	32.3	14.5	8.09
High school or GED	3,442	166	681	2,514	5,778	6,373
	(341)	(99)	(156)	(396)	(606)	(762)
Black	−544	22	−60	−129	−866	−1,446
	(397)	(115)	(188)	(451)	(679)	(869)
Hispanic	−1,151	−31	−222	−995	−1,620	−1,503
	(488)	(130)	(194)	(546)	(911)	(992)
Married	−667	−213	−392	−758	−1,048	−902
	(436)	(127)	(209)	(522)	(785)	(970)
Worked less than 13	−5,313	−1,050	−3,240	−6,872	−7,670	−6,470
weeks in past year	(370)	(137)	(289)	(522)	(672)	(787)
AFDC	−3,009	−398	−1,047	−3,389	−4,334	−3,875
	(378)	(107)	(174)	(468)	(737)	(834)
Constant	10,361	649	2,633	8,417	16,498	20,689
	(815)	(255)	(490)	(966)	(1,554)	(1,232)

JTPA: Quantile Regression with IV

QUANTILE TREATMENT EFFECTS AND 2SLS ESTIMATES

Dependent Variable: 30-month Earnings

		Quantile				
	2SLS	0.15	0.25	0.50	0.75	0.85
A. Men						
Training	1,593	121	702	1,544	3,131	3,378
	(895)	(475)	(670)	(1,073)	(1,376)	(1,811)
% Impact of Training	8.55	5.19	12.0	9.64	10.7	9.02
High school or GED	4,075	714	1,752	4,024	5,392	5,954
	(573)	(429)	(644)	(940)	(1,441)	(1,783)
Black	−2,349	−171	−377	−2,656	−4,182	−3,523
	(625)	(439)	(626)	(1,136)	(1,587)	(1,867)
Hispanic	335	328	1,476	1,499	379	1,023
	(888)	(757)	(1,128)	(1,390)	(2,294)	(2,427)
Married	6,647	1,564	3,190	7,683	9,509	10,185
	(627)	(596)	(865)	(1,202)	(1,430)	(1,525)
Worked less than 13	−6,575	−1,932	−4,195	−7,009	−9,289	−9,078
weeks in past year	(567)	(442)	(664)	(1,040)	(1,420)	(1,596)
Constant	10,641	−134	1,049	7,689	14,901	22,412
	(1,569)	(1,116)	(1,655)	(2,361)	(3,292)	(7,655)
B. Women						
Training	1,780	324	680	1,742	1,984	1,900
	(532)	(175)	(282)	(645)	(945)	(997)
% Impact of Training	14.6	35.5	23.1	18.4	10.1	7.39
High school or GED	3,470	262	768	2,955	5,518	5,905
	(342)	(178)	(274)	(643)	(930)	(1026)
Black	−554	0	−123	−401	−1,423	−2,119
	(397)	(204)	(318)	(724)	(949)	(1,196)
Hispanic	−1,145	−73	−138	−1,256	−1,762	−1,707
	(488)	(217)	(315)	(854)	(1,188)	(1,172)
Married	−652	−233	−532	−796	38	−109
	(437)	(221)	(352)	(846)	(1,069)	(1,147)
Worked less than 13	−5,329	−1,320	−3,516	−6,524	−6,608	−5,698
weeks in past year	(370)	(254)	(430)	(781)	(931)	(969)
AFDC	−2,997	−406	−1,240	−3,298	−3,790	−2,888
	(378)	(189)	(301)	(743)	(1,014)	(1,083)
Constant	10,538	984	3,541	9,928	15,345	20,520
	(828)	(547)	(837)	(1,696)	(2,387)	(1,687)