Testing the LATE Assumptions

Francis DiTraglia

Oxford Economics Summer School 2022

Review of "Textbook" Instrumental Variables (IV) Model

Observed

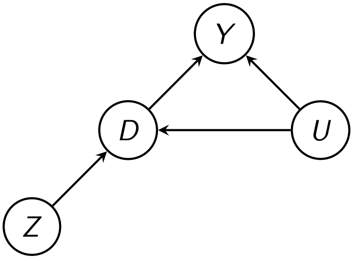
- Y = Outcome (Wage)
- D = Treatment (Attend Uni)
- Z = IV (Live Nearby)

Unobserved

- U = Confounders (Ability)

Assumptions

- Model: $Y = \alpha + \beta D + U$
- Relevance: $\text{Cov}(Z, D) \neq 0$
- Exogeneity: $\text{Cov}(Z, U) = 0$



A Relevant, Exogenous Instrument Identifies β

Assumptions

$$Y = \alpha + \beta D + U, \quad \mathrm{Cov}(Z, D) \neq 0, \quad \mathrm{Cov}(Z, U) = 0$$

OLS

$$\beta_{\mathrm{OLS}} = \frac{\mathrm{Cov}(D, Y)}{\mathrm{Var}(D)} = \frac{\beta \mathrm{Var}(D) + \mathrm{Cov}(D, U)}{\mathrm{Var}(D)} + \beta + \frac{\mathrm{Cov}(D, U)}{\mathrm{Var}(D)} \neq \beta$$

IV

$$\beta_{\mathrm{IV}} = \frac{\mathrm{Cov}(Z, Y)}{\mathrm{Cov}(Z, D)} = \frac{\beta \mathrm{Cov}(Z, D) + \mathrm{Cov}(Z, U)}{\mathrm{Cov}(Z, D)} = \beta + \frac{\mathrm{Cov}(Z, U)}{\mathrm{Cov}(Z, D)} = \beta$$

```r
library(mvtnorm)
library(tidyverse)
set.seed(587103)
n <- 1e4
sims <- rmvnorm(n, sigma = matrix(c(1, 0.5,
                                    0.5, 1), 2, 2, byrow = TRUE))
U <- sims[,1]
V <- sims[,2]
Z <- rbinom(n, size = 1, prob = 0.3)
D <- -0.5 + 0.3 * Z + V
beta <- -0.3
Y <- 1 + beta * D + U
c(OLS = cov(D, Y) / var(D), IV = cov(Z, Y) / cov(Z, D),
  truth = beta) %>% round(2)
```

```
##    OLS     IV truth
##   0.20 -0.28 -0.30
```

Which assumptions are testable in the textbook IV model?

Instrument Relevance

- Since D and Z are observed, directly estimate $\text{Cov}(D, Z)$.
- Beware of weak instruments!

Instrument Exogeneity

- Since U is unobserved, can't directly estimate $\text{Cov}(Z, U)$.
- Could we use the IV residuals?

```r
library(AER) # for ivreg()
library(broom) # for tidy()
beta <- 0
Y <- 1 + beta * D - Z + U  # Endogenous instrument!
iv_results <- ivreg(Y ~ D | Z)
tidy(iv_results) %>% knitr::kable(digits = 2)
```

term	estimate	std.error	statistic	p.value
(Intercept)	-0.62	0.12	-5.05	0
D	-3.30	0.29	-11.56	0

```r
cov(residuals(iv_results), Z)
```

```
## [1] 1.604262e-14
```

Z Is Uncorrelated with the IV Residuals *By Construction*

- Let U be the **structural error** and V be the **IV residual**: $V \equiv Y - \alpha_{IV} - \beta_{IV} D$.

$$\beta_{IV} = \frac{\text{Cov}(Z,Y)}{\text{Cov}(Z,D)} = \beta + \frac{\text{Cov}(Z,U)}{\text{Cov}(Z,D)}, \quad \alpha_{IV} = \mathbb{E}(Y) - \beta_{IV}\mathbb{E}(D).$$

- $V = U \iff Z$ is exogenous: the only way to obtain $\beta_{IV} = \beta$ and $\alpha_{IV} = \alpha$.

$$\begin{aligned}\text{Cov}(Z,V) &= \text{Cov}(Z, Y - \alpha_{IV} - \beta_{IV} D) = \text{Cov}(Z,Y) - \beta_{IV}\text{Cov}(Z,D)\\ &= \text{Cov}(Z,Y) - \frac{\text{Cov}(Z,Y)}{\text{Cov}(Z,D)}\text{Cov}(Z,D) = 0.\end{aligned}$$

- $\text{Cov}(Z,V) = 0$ *by construction* even when $\text{Cov}(Z,U) \neq 0$

Multiple Instruments and Over-identification

Assumptions

- $Y = \alpha + \beta D + U$
- $\text{Cov}(Z_1, D) \neq 0, \ \text{Cov}(Z_2, D) \neq 0$
- $\text{Cov}(Z_1, U) = \text{Cov}(Z_2, U) = 0$

Implications

- Both IVs identify *same* effect: β
- If not, **at least one is endogenous**

Over-identifying Restrictions Test

- Test of null that all MCs identify same parameters.

$$\beta_{IV}^{(1)} \equiv \frac{\text{Cov}(Z_1, Y)}{\text{Cov}(Z_1, D)} = \beta + \frac{\text{Cov}(Z_1, U)}{\text{Cov}(Z_1, D)}$$

$$\beta_{IV}^{(2)} \equiv \frac{\text{Cov}(Z_2, Y)}{\text{Cov}(Z_2, D)} = \beta + \frac{\text{Cov}(Z_2, U)}{\text{Cov}(Z_2, D)}$$

$$\beta_{IV}^{(1)} - \beta_{IV}^{(2)} = \frac{\text{Cov}(Z_1, U)}{\text{Cov}(Z_1, D)} - \frac{\text{Cov}(Z_2, U)}{\text{Cov}(Z_2, D)}$$

Beyond the Textbook IV Model

Heterogenous Treatment Effects

- $Y = \alpha + \beta D + U$ implies that everyone has the same treatment effect: β.
- In reality, treatment effects differ across people.

Overidentifying restrictions?

- Out the window! Different instruments may identify different causal parameters.

Local Average Treatment Effects (LATE) Model

- What does IV tell us when treatment effects are heterogeneous?

Review of the LATE Model

- Suppose that both D and Z are binary

$$\beta_{\text{IV}} \equiv \frac{\text{Cov}(Z,Y)}{\text{Cov}(Z,D)} = \frac{\frac{\text{Cov}(Y,Z)}{\text{Var}(Z)}}{\frac{\text{Cov}(D,Z)}{\text{Var}(Z)}} = \frac{\mathbb{E}[Y|Z=1]-\mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1]-\mathbb{E}[D|Z=0]} \equiv \text{Wald Estimand}$$

Intent-to-treat Effect: $\mathbb{E}[Y|Z=1]-\mathbb{E}[Y|Z=0]$ (ITT)

- E.g. randomized experiment with treatment **offer** Z and treatment **take-up** D
- **Non-compliance** / randomized encouragement design: D may not equal Z
- In this setting the ITT is the ATE of *offering* treatment.

The Wald Estimand

- ITT is "diluted" by people who are offered $(Z=1)$ but do not take up $(D=0)$
- Divide ATE of offer on outcome $Z \to Y$ by that of offer on take-up $Z \to D$.
- **Under what assumptions does this give us a meaningful causal quantity?**

Decomposing the ITT Effect

- Example: moving to opportunity (MTO) experiment randomly offered housing vouchers to encourage families to move to a more affluent neighborhood.
- 50% of offered families ($Z = 1$) moved; 20% of non-offered families ($Z = 0$) moved

$$Y = (1 - D)Y_0 + DY_1, \quad p_z \equiv \mathbb{P}(D = 1 | Z = z)$$

- $\mathbb{E}[Y|Z = 1]$ is a *mixture* of Y_0 and Y_1 for different types of families:

$$\mathbb{E}[Y|Z = 1] = \underbrace{(1 - p_1)}_{\approx 0.5} \mathbb{E}[Y_0|Z = 1, D = 0] + \underbrace{p_1}_{\approx 0.5} \mathbb{E}[Y_1|Z = 1, D = 1]$$

- $\mathbb{E}[Y|Z = 0]$ is a *mixture* of Y_0 and Y_1 for different types of families:

$$\mathbb{E}[Y|Z = 0] = \underbrace{(1 - p_0)}_{\approx 0.8} \mathbb{E}[Y_0|Z = 0, D = 0] + \underbrace{p_0}_{\approx 0.2} \mathbb{E}[Y_1|Z = 0, D = 1]$$

Compliance "Types" in the LATE Model

- Catalogue all possible treatment take-up "decision rules"

$$\begin{aligned}
\text{Never-taker:} \quad & T = n \quad \Longleftrightarrow \quad D(Z) = 0 \\
\text{Always-taker:} \quad & T = a \quad \Longleftrightarrow \quad D(Z) = 1 \\
\text{Complier:} \quad & T = c \quad \Longleftrightarrow \quad D(Z) = Z \\
\text{Defier:} \quad & T = d \quad \Longleftrightarrow \quad D(Z) = (1 - Z).
\end{aligned}$$

In the MTO Example

- Never-takers: families that refuse to move with or without a voucher
- Always-takers: families that will move with or without a voucher
- Compliers are families that will **only move if given a voucher**
- Defiers are families that will only move if **not** given a voucher

Assumption 1 - Unconfounded Type

For all compliance types $t \in \{a, c, n, d\}$

$$\mathbb{P}(T = t) = \mathbb{P}(T = t|Z = 0) = \mathbb{P}(T = t|Z = 1).$$

Assumption 2 - No Defiers: $\mathbb{P}(T = d) = 0$

Assumption 3 - Mean Exclusion Restriction

For all compliance types $t \in \{a, c, n, d\}$

$$\mathbb{E}[Y_0|Z = 0, T = t] = \mathbb{E}[Y_0|Z = 1, T = t] = \mathbb{E}[Y_0|T = t]$$
$$\mathbb{E}[Y_1|Z = 0, T = t] = \mathbb{E}[Y_1|Z = 1, T = t] = \mathbb{E}[Y_0|T = t]$$

Assumption 4 - Existence of Compliers: $\mathbb{P}(T = c) > 0$

Lemma 1: Assumptions 1–2 $\implies$

$$
\begin{aligned}
\mathbb{P}(D = 1|Z = 1) &= \mathbb{P}(T = a) + \mathbb{P}(T = c) \\
\mathbb{P}(D = 0|Z = 1) &= \mathbb{P}(T = n) \\
\mathbb{P}(D = 1|Z = 0) &= \mathbb{P}(T = a) \\
\mathbb{P}(D = 0|Z = 0) &= \mathbb{P}(T = c) + \mathbb{P}(T = n)
\end{aligned}
$$

Lemma 2: Assumptions 1–3 $\implies$

$$
\begin{aligned}
\mathbb{E}\left[Y|D = 1, Z = 1\right] &= \frac{\mathbb{P}(T = a)\mathbb{E}\left[Y_1|T = a\right] + \mathbb{P}(T = c)\mathbb{E}\left[Y_1|T = c\right]}{\mathbb{P}(T = a) + \mathbb{P}(T = c)} \\
\mathbb{E}\left[Y|D = 0, Z = 1\right] &= \mathbb{E}\left[Y_0|T = n\right] \\
\mathbb{E}\left[Y|D = 1, Z = 0\right] &= \mathbb{E}\left[Y_1|T = a\right] \\
\mathbb{E}\left[Y|D = 0, Z = 0\right] &= \frac{\mathbb{P}(T = n)\mathbb{E}\left[Y_0|T = n\right] + \mathbb{P}(T = c)\mathbb{E}\left[Y_0|T = c\right]}{\mathbb{P}(T = n) + \mathbb{P}(T = c)}
\end{aligned}
$$

The LATE Theorem: Wald = ATE for Compliers

Theorem: Assumptions 1–4 $\implies$

$$\frac{\mathbb{E}(Y|Z=1) - \mathbb{E}(Y|Z=0)}{\mathbb{E}(D|Z=1) - \mathbb{E}(D|Z=0)} = \mathbb{E}\left[Y_1 - Y_0 | T = c\right]$$

Proof

- Algebra and of Iterated Expectations, using the two lemmas. (See lecture notes)

MTO Example

- ITT is the average treatment effect of *offering* a housing voucher.
- Wald = LATE is the average treatment effect of *moving to opportunity* for families who can be induced to move by a housing voucher.
- Different IV $\implies$ different compliers $\implies$ different LATE. It's a **local** effect!

Are the LATE Assumptions Testable?

LATE Assumptions

1. Unconfounded Type
2. No Defiers
3. Mean Exclusion Restriction
4. Existence of Compliers

At Least One is Testable!

- Assumptions 1–3 $\implies \mathbb{P}(T = c) = \mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0]$
- Thus, Assumption 4 is just **instrument relevance**, hence testable.
- What about the others?

Even Nobel Laureates Make Mistakes

Angrist & Imbens (1994)

Part (i) is similar to an exclusion restriction in a regression model. It is not testable and has to be considered on a case by case basis.

Pearl (1995)

exogeneity . . . can be given an empirical test. The test is not guaranteed to detect all violations of exogeneity, but it can, in certain circumstances, screen out very bad would-be instruments.

This Lecture

- Testable implications LATE assumptions from above: Huber & Mellace (2015)

Closely-related Work

- Kitagawa (2015)
- Mourifié & Wan (2017)

Huber & Mellace (2015) – The Big Picture

- Assumptions 1–3 imply four inequalities: $\theta_1 \leq 0$, $\theta_2 \leq 0$, $\theta_3 \leq 0$, $\theta_4 \leq 0$
- $\theta \equiv (\theta_1, \theta_2, \theta_3, \theta_4)$ depend only on (Y, D, Z); we'll define them shortly.
- If any element of θ is *positive* at least one of Assumptions 1–3 must be false.
- In practice: compare estimate $\widehat{\theta}$ to appropriate standard errors.
- Not all violations of the LATE assumptions lead to a positive value for θ
- **Necessary but not sufficient** for validity of LATE assumptions.
- The four inequalities come in *pairs*. We'll look at each pair in turn.

First Pair of Inequalities

Define: $F_{11}(y) \equiv \mathbb{P}(Y \leq y | D = 1, Z = 1)$ and

$$y_q \equiv F_{11}^{-1}(q), \quad y_{1-q} \equiv F_{11}^{-1}(1-q), \quad q \equiv \frac{\mathbb{P}(D=1|Z=0)}{\mathbb{P}(D=1|Z=1)}$$

Under Assumptions 1–3:

$$\mathbb{E}(Y|D=1, Z=1, Y \leq y_q) \leq \mathbb{E}(Y|D=1, Z=0) \leq \mathbb{E}(Y|D=1, Z=1, Y \geq y_{1-q})$$

Key Points

- Lemma 1 $\implies \mathbb{E}(Y|D=1, Z=0) = \mathbb{E}(Y_1|T=a)$ so now we have two *partial identification bounds* as well.
- Why care? **Overidentifying Restrictions**
- At most one of the pair can be violated.

Second Pair of Inequalities

Define $F_{00}(y) \equiv \mathbb{P}(Y \leq y|D=0, Z=0)$ and

$$y_r \equiv F_{00}^{-1}(r), \quad y_{1-r} \equiv F_{00}^{-1}(1-r), \quad r \equiv \frac{\mathbb{P}(D=0|Z=1)}{\mathbb{P}(D=0|Z=0)}.$$

Under Assumptions 1–3:

$$\mathbb{E}(Y|D=0, Z=0, Y \leq y_r) \leq \mathbb{E}(Y|D=0, Z=1) \leq \mathbb{E}(Y|D=0, Z=0, Y \geq y_{1-r})$$

Key Points

- Lemma 1 $\implies \mathbb{E}(Y|D=0, Z=1) = \mathbb{E}(Y_1|T=n)$ so now we have two *partial identification bounds* as well.
- Why care? **Overidentifying Restrictions**
- At most one of the pair can be violated.

Theorem: Assumptions 1–3 $\Longrightarrow$

$$\begin{bmatrix}\theta_1\\ \\ \theta_2\\ \\ \theta_3\\ \\ \theta_4\end{bmatrix} \equiv \begin{bmatrix}\mathbb{E}(Y|D=1,Z=1,Y\leq y_q)-\mathbb{E}(Y|D=1,Z=0)\\ \\ \mathbb{E}(Y|D=1,Z=0)-\mathbb{E}(Y|D=1,Z=1,Y\geq y_{1-q})\\ \\ \mathbb{E}(Y|D=0,Z=0,Y\leq y_r)-\mathbb{E}(Y|D=0,Z=1)\\ \\ \mathbb{E}(Y|D=0,Z=1)-\mathbb{E}(Y|D=0,Z=0,Y\geq y_{1-r})\end{bmatrix} \leq \begin{bmatrix}0\\ \\ 0\\ \\ 0\\ \\ 0\end{bmatrix}$$

Where: $F_{11}(y)\equiv\mathbb{P}(Y\leq y|D=1,Z=1)$, $F_{00}(y)\equiv\mathbb{P}(Y\leq y|D=0,Z=0)$, and

$$y_q\equiv F_{11}^{-1}(q),\quad y_{1-q}\equiv F_{11}^{-1}(1-q),\quad q\equiv\frac{\mathbb{P}(D=1|Z=0)}{\mathbb{P}(D=1|Z=1)}$$

$$y_r\equiv F_{00}^{-1}(r),\quad y_{1-r}\equiv F_{00}^{-1}(1-r),\quad r\equiv\frac{\mathbb{P}(D=0|Z=1)}{\mathbb{P}(D=0|Z=0)}.$$

By Lemma 1, Assumptions 1–2 $\implies$

$$q \equiv \frac{\mathbb{P}(D=1|Z=0)}{\mathbb{P}(D=1|Z=1)} = \frac{\mathbb{P}(T=a)}{\mathbb{P}(T=a)+\mathbb{P}(T=c)}.$$

- $q =$ share of always-takers, $(1-q) =$ share of compliers *among* $(D=1, Z=1)$

Re-write Expression for $\mathbb{E}(Y|D=1, Z=1)$ from Lemma 2

$$\mathbb{E}\left(Y|D=1, Z=1\right) = (1-q)\mathbb{E}\left(Y_1|T=c\right) + q\mathbb{E}\left(Y_1|T=a\right).$$

- **Mixture** of Y_1 for compliers and always-takers
- LHS conditions on Z but RHS doesn't: mean exclusion assumption.

Infeasible Test of Mean Exclusion

- Suppose that we knew *who* was an always-taker among those with $(D=1, Z=1)$
- Compare $\mathbb{E}(Y|Z=0, D=1) = \mathbb{E}(Y_1|Z=0, T=a)$ to $\mathbb{E}(Y_1|Z=1, T=a)$.

We don't know who is an always-taker among $(D = 1, Z = 1)$.

What do we know?

- $q \times 100\%$ of those with $(D = 1, Z = 1)$ are always-takers.
- $(1 - q) \times 100\%$ of those with $(D = 1, Z = 1)$ are compliers.
- This turns out to be enough to obtain bounds.

Shorthand

$$
\begin{aligned}
F(y) &\equiv \mathbb{P}(Y \leq y | D = 1, Z = 1) = \mathbb{P}(Y_1 \leq y | T \in \{a, c\}, Z = 1) \\
G(y) &\equiv \mathbb{P}(Y_1 \leq y | T = c, Z = 1) \\
H(y) &\equiv \mathbb{P}(Y_1 \leq y | T = a, Z = 1).
\end{aligned}
$$

An Abstract Probability Puzzle

(Law of Total Probability + Assumptions 1–2) $\implies$

$$
\begin{aligned}
F(y) &= \mathbb{P}(T = c | T \in \{a, c\}, Z = 1) G(y) + \mathbb{P}(T = a | T \in \{a, c\}, Z = 1) H(y) \\
&= \mathbb{P}(T = a | T \in \{a, c\}) G(y) + \mathbb{P}(T = c | T \in \{a, c\}) H(y) \\
&= \frac{\mathbb{P}(T = c)}{\mathbb{P}(T \in \{a, c\})} G(y) + \frac{\mathbb{P}(T = a)}{\mathbb{P}(T \in \{a, c\})} H(y)
\end{aligned}
$$

Using our expression for q from two slides back

$$F(y) = (1 - q) G(y) + q H(y)$$

- We know q, $(1 - q)$ and F.
- Don't know G or H but mean of H equals $\mathbb{E}(Y_1 | T = a)$ under mean exclusion.
- What all possible values for the mean of H given knowledge of F and q?

Bound the mean of H given $F = (1-q)G + qH$ with q and F known.

Step 1 - Solve for H

$$H(y) = \left(\frac{1}{q}\right) F(y) - \left(\frac{1-q}{q}\right) G(y).$$

Step 2 - Bound H

- We know nothing about G, but it must like between 0 and 1 to be a valid CDF.
- Substitute $G(y) = 0$ and $G(y) = 1$ into the expression for H

$$\frac{F(y)}{q} - \frac{1-q}{q} \le H(y) \le \frac{F(y)}{q}.$$

- But H is a CDF too! Make sure that it lies between 0 and 1

$$\max\left\{0, \frac{F(y)}{q} - \frac{1-q}{q}\right\} \le H(y) \le \min\left\{1, \frac{F(y)}{q}\right\}$$

More Shorthand

$$\overline{H}(y) \equiv \max\left\{0, \frac{F(y)}{q} - \frac{1-q}{q}\right\} \leq H(y) \leq \min\left\{1, \frac{F(y)}{q}\right\} \equiv \underline{H}(y)$$

Both $\overline{H}$ and $\underline{H}$ are CDFs

- Non-decreasing since $F(y)$ is non-decreasing.
- Bounded between 0 and 1
- Approach 0 as $y \to -\infty$
- Approach 1 as $y \to +\infty$

First-order Stochastic Dominance

- F_1 **stochastically dominates** $F_2 \iff F_1(y) \leq F_2(y) \iff F_1^{-1}(y) \geq F_2^{-1}(y)\ \forall y$
- $\overline{H}$ stochastically dominates H and H stochastically dominates $\underline{H}$.

First-Order Stochastic Dominance $\implies$ Inequality for Means

- $\overline{H}$ stochastically dominates H and H stochastically dominates $\underline{H}$.

$$\underbrace{\int_{\mathbb{R}} y\underline{H}(dy)}_{\underline{\mu}} \leq \underbrace{\int_{\mathbb{R}} yH(dy)}_{\mu} \leq \underbrace{\int_{\mathbb{R}} y\overline{H}(dy)}_{\overline{\mu}}.$$

- The mean μ of H, a distribution we don't know, must lie between the means $\underline{\mu}$ and $\overline{\mu}$ of $\underline{H}$ and $\overline{H}$, two distributions we know!
- Given our definitions of $\underline{H}$, H and $\overline{H}$, the preceding is *identical* to

$$\mathbb{E}(Y|D=1, Z=1, Y \leq y_q) \leq \mathbb{E}(Y_1|T=z) \leq \mathbb{E}(Y|D=1, Z=1, Y \geq y_{1-q})$$

- Simplify slightly: suppose F is continuous, strictly increasing with density f
- Next slide: work out the densities $\underline{h}$ and $\overline{h}$ that correspond to $\underline{H}$ and $\overline{H}$.

Deriving $\overline{h}(y)$

- $\overline{H} \equiv \max\left\{0, \frac{F(y)-(1-q)}{q}\right\}$ equals zero until $F(y) > (1-q)$
- Solving: $\overline{H}$ is strictly increasing for $y > F^{-1}(1-q)$
- Differentiating: $\frac{d}{dy}[F(y) - (1-q)]/q = f(y)/q$ and therefore

$$\overline{h}(y) = 1\left\{y > F^{-1}(1-q)\right\}\frac{f(y)}{q}$$

Deriving $\underline{h}(y)$

- $\underline{H} \equiv \min\left\{1, \frac{F(y)}{q}\right\}$ equals $F(y)/q$ until $F(y) = q$

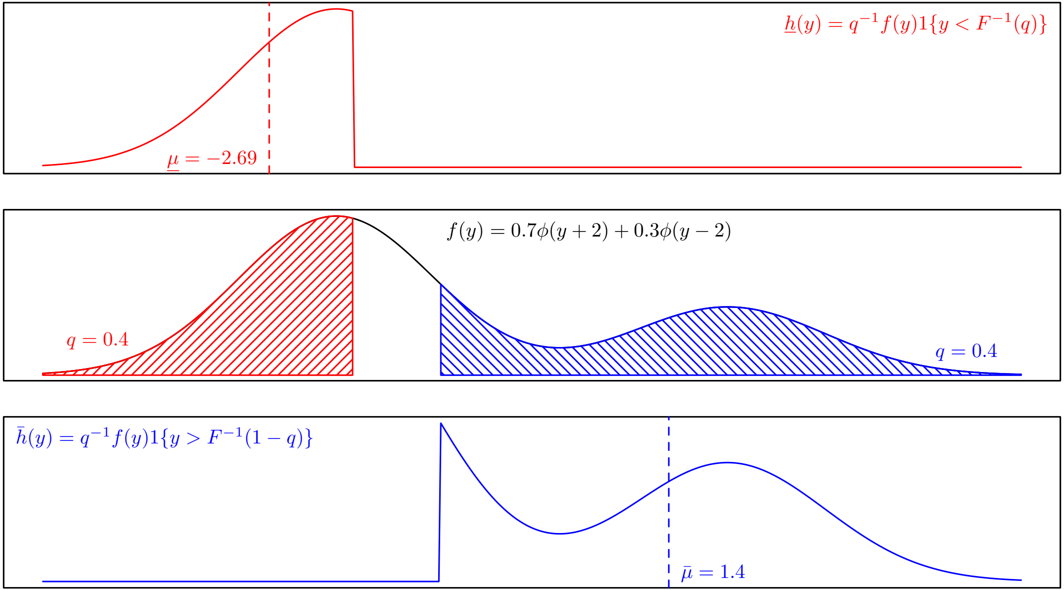
$$\underline{h}(y) = 1\left\{y < F^{-1}(q)\right\}\frac{f(y)}{q}.$$

$\overline{h}$ and $\underline{h}$ are Truncated versions of f

- $\overline{h}(y) = 1\left\{y > F^{-1}(1-q)\right\}\dfrac{f(y)}{q}$ is the **top** $q \times 100\%$ of f rescaled
- $\underline{h}(y) = 1\left\{y < F^{-1}(q)\right\}\dfrac{f(y)}{q}$ is the **bottom** $q \times 100\%$ of f, rescaled

$$\underline{\mu} \equiv \int_{-\infty}^{F^{-1}(q)} \frac{y}{q} f(y)\, dy \leq \mu \leq \int_{F^{-1}(1-q)}^{\infty} \frac{y}{q} f(y)\, dy \equiv \overline{\mu}$$

- Next slide: example where $q = 0.4$ and f is a mixture of normals.

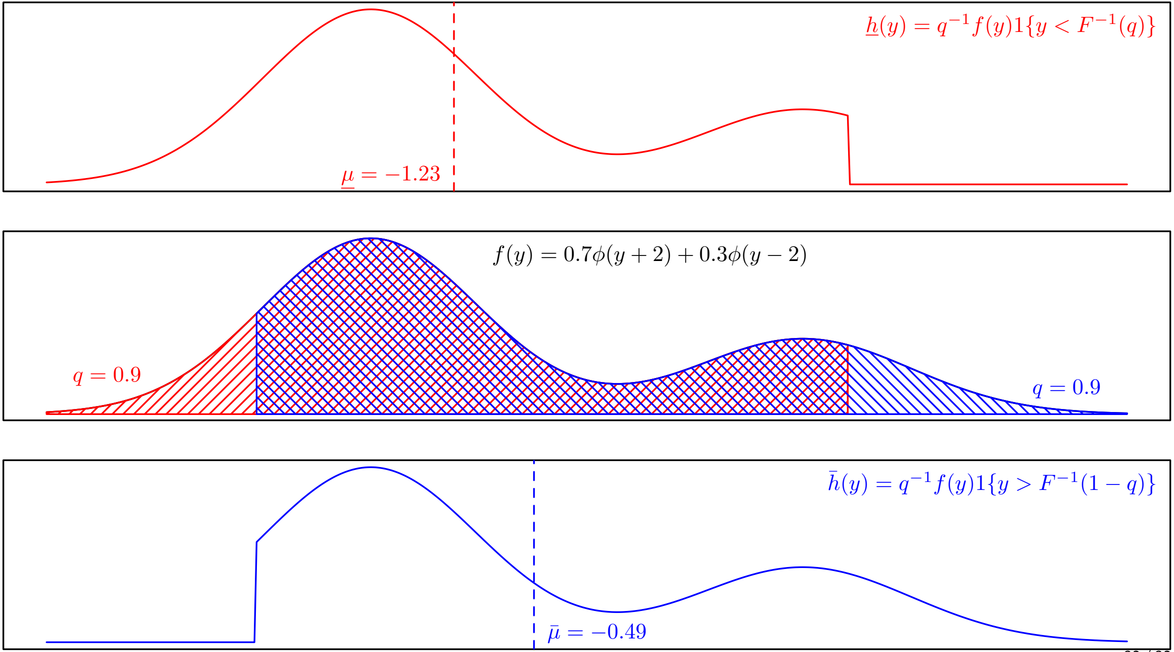

$\underline{h}(y) = q^{-1} f(y) 1\{y < F^{-1}(q)\}$
$\underline{\mu} = -2.69$
$f(y) = 0.7\phi(y+2) + 0.3\phi(y-2)$
$q = 0.4$
$q = 0.4$
$\bar{h}(y) = q^{-1} f(y) 1\{y > F^{-1}(1-q)\}$
$\bar{\mu} = 1.4$

What does this have to do with always-takers and compliers?

$$\underline{\mu} \equiv \int_{-\infty}^{F^{-1}(q)} \frac{y}{q} f(y)\, dy \leq \mu \leq \int_{F^{-1}(1-q)}^{\infty} \frac{y}{q} f(y)\, dy \equiv \overline{\mu}$$

$$\mathbb{E}(Y|D=1, Z=1, Y \leq y_q) \leq \mathbb{E}(Y_1|T=a) \leq \mathbb{E}(Y|D=1, Z=1, Y \geq y_{1-q})$$

- f is the density of $Y|D=1, Z=1$, a mixture of compliers and always-takers
- $q = 0.4 \implies$ then 40% always-takers, but we don't know who they are
- Consider the two most extreme possibilities: pack them all in the **bottom** or **top** 40% of the distribution of $Y|D=1, Z=1$.
- Resulting bounds for $\mathbb{E}(Y_1|T=a)$ are sharp; **tightest** when few compliers (q large)


$\underline{h}(y) = q^{-1} f(y) 1\{y < F^{-1}(q)\}$
$\underline{\mu} = -1.23$
$f(y) = 0.7\phi(y+2) + 0.3\phi(y-2)$
$q = 0.9$
$q = 0.9$
$\bar{h}(y) = q^{-1} f(y) 1\{y > F^{-1}(1-q)\}$
$\bar{\mu} = -0.49$