14.662 Recitation 1

DFL, MM, FFL, and a quick Mundlak

Peter Hull

Spring 2015

Part 1: Review: DiNardo, Fortin, and Lemieux (1996)

Why All the Fancy New 'Metrics?

- Growing interest in the *distribution* of wages
- Would like to link distributional features of Y_i to other factors, X_i
  - As a descriptive task (e.g. "how much of the 90^{th}-10^{th} percentile gap in wages can we explain by differences in education?")
  - To answer causal questions (e.g. "what would happen to the 10^{th} percentile of earnings if we made community college free?")
- OLS/IV are all about *means;* to say something about other distributional features, we have to learn some new skills
- In some cases (e.g. "conditional" v. "unconditional" quantile regression), we have to face issues that OLS inherently sidesteps

DFL '96 Overview

- DFL extend the Oaxaca-Blinder mean-decomposition intuition to decompose wage distributions
- Basic idea: write

$$f(w;t_w,t_z) = \int_z f(w|z,t_w,t_z)dF(z|t_w,t_z)$$

where $w =$ wage, $z =$ individual attributes, $t_v =$ "time" (parameterizes distribution of v)

- Assume $f(w|z,t_w,t_z) = f(w|z,t_w)$, $dF(z|t_w,t_z) = dF(z|t_z)$:

$$\begin{aligned} f(w;t_w=t,t_z=t') &= \int_z f(w|z,t_w=t)dF(z|t_z=t') \\ &= \int_z f(w|z,t_w=t)\psi(z;t',t)dF(z|t_z=t) \end{aligned}$$

where $\psi(z;t',t) \equiv dF(z|t_z=t')/dF(z|t_z=t)$

DFL '96 Results

- $\psi(z;t',t)$ a "reweighting" that gives a "counterfactual" distribution of wages when $t' \neq t$ (like O-B)
  - Once you estimate $\psi(z;t',t)$, you can estimate (by KDE) "the density [of wages] that would have prevailed if individual attributes had remained at their 1979 level and workers had been paid *according to the wage schedule observed in 1988*"
- By Bayes' rule:

$$\psi(z;t',t) \equiv \frac{P(z|t')}{P(z|t)} = \frac{P(t'|z)\cdot P(z)/P(t')}{P(t|z)\cdot P(z)/P(t)} = \frac{P(t'|z)}{P(t|z)}\frac{P(t)}{P(t')}$$

  and it's easy to estimate these pieces (DFL use probit)
- DFL show this decomposition, while also accounting for changes in unionization rates and the min. wage (see notes for details). Find a lot of residual difference between 1979 and 1988 wage distribution
  - Reminder #1: decomposition order matters (as with O-B)
  - Reminder #2: partial equilibrium exercise (by assumption)

Part 2: Quantile Methods

Conditional QR: a Review

- The *quantile function* Q_Y is defined as the inverse of a CDF:

$$Q_Y(\tau|X_i) = y \iff F_Y(y|X_i) = \tau$$

It is thus invariant to monotone transformations $T(\cdot)$:

$$Q_Y(\tau|X_i) = y \implies P(Y_i \leq y|X_i) = \tau \implies$$
$$P(T(Y_i) \leq T(y)|X_i) = \tau \implies Q_{T(Y)}(\tau|X_i) = T(Q_Y(\tau|X_i)) = T(y)$$

- *Conditional QR* models $Q_Y(\tau|X_i)$ as a linear function of X_i:

$$Q_Y(\tau|X_i) = X_i'\beta_\tau$$

- This implies (can verify by writing out integrals and taking FOC):

$$\beta_\tau = \arg\min_b E\left[\rho_\tau(Y - X_i'b)\right]$$

$$\rho_\tau(\varepsilon) \equiv \begin{cases} \tau\varepsilon, & \varepsilon \geq 0 \\ (1-\tau)|\varepsilon|, & \varepsilon < 0 \end{cases}$$

Interpreting Conditional QR

- A linear $Q_Y(\tau|X_i)$ is consistent with a *location-scale* model:

$$Y_i = X_i'\alpha + X_i'\delta\varepsilon_i, \ \varepsilon_i \perp\!\!\!\perp X_i$$

  Since Y_i is monotone in ε_i conditional on X_i:

$$\begin{aligned} Q_Y(\tau|X_i) &= X_i'\alpha + X_i'\delta Q_\varepsilon(\tau|X_i) \\ &= X_i'\alpha + X_i'\delta Q_\varepsilon(\tau) = X_i'\beta_\tau \end{aligned}$$

- β_τ is the effect of X_i on the τ^{th} quantile of Y (<u>not</u> the effect on the τ^{th} quantile individual)
- If X_i is multidimensional, $\beta_{\tau,1}$ is the effect of $X_{i,1}$ on the τ^{th} quantile of Y, conditional on $X_{i,2} \ldots X_{i,k}$
  - Ex: $X_i = \begin{bmatrix} D_i & W_i' \end{bmatrix}'$ for D_i binary: $\beta_{\tau,1} =$ quantile treatment effect

Why is QR "Conditional" when OLS is not?

- Suppose $Y_i = \beta D_i + W_i'\gamma + (1+D_i)\varepsilon_i$ with $\varepsilon_i \perp\!\!\!\perp D_i, W_i$
  $\implies$ Both $E[Y|D_i, W_i]$ and $Q_Y(\tau|D_i, W_i)$ are linear

- Both QR and OLS give the *conditional* effect of D_i on Y_i:

$$
\begin{aligned}
E[Y_{1i}|W_i] - E[Y_{0i}|W_i] &= \beta + W_i'\gamma + E[2\varepsilon_i] - \left(W_i'y + E[\varepsilon_i]\right) \\
&= \beta \\
Q_{Y_1}(\tau|W_i) - Q_{Y_0}(\tau|W_i) &= \beta + W_i'\gamma + 2Q_\varepsilon(\tau) - \left(W_i'\gamma + Q_\varepsilon(\tau)\right) \\
&= \beta + Q_\varepsilon(\tau)
\end{aligned}
$$

- But not necessarily the *unconditional* effect:

$$
\begin{aligned}
E[Y_{1i}] - E[Y_{0i}] =& \beta + E[W_i'\gamma] + E[2\varepsilon_i] - \left(E[W_i'\gamma] + E[\varepsilon_i]\right) \\
=& \beta \\
Q_{Y_1}(\tau) - Q_{Y_0}(\tau) =& \beta + Q_{W'\gamma+2\varepsilon}(\tau) - Q_{W'\gamma+\varepsilon}(\tau) \\
\neq& \beta + Q_{W'\gamma}(\tau) + 2Q_\varepsilon(\tau) - \left(Q_{W'\gamma}(\tau) + Q_\varepsilon(\tau)\right)
\end{aligned}
$$

"Unconditioning" QR: Machado and Mata (2005)

Skorohod representation: $Y_i = Q_Y(\theta_i|X_i)$ for $\theta_i|X_i \sim U(0,1)$, because

$$\theta_i = F_Y(Y_i|X_i) \implies \theta_i|X_i \sim U(0,1)$$
$$Q_Y(\theta_i|X_i) = Q_Y(F_Y(Y_i|X_i)|X_i) = Y_i$$

M&M Marginalizing Method:

1. $\forall w \in supp(W_i)$, draw θ_i, simulate $\left(\widehat{Y_{1wi}}, \widehat{Y_{0wi}}\right)$ with $\widehat{Q_Y}(\theta_i|D_i, W_i)$
2. Average up $\left(\widehat{Y_{1wi}}, \widehat{Y_{0wi}}\right)$ by $\widehat{f_W}(w)$
3. Compute $\widehat{Q_{Y_1}}(\tau) - \widehat{Q_{Y_0}}(\tau)$

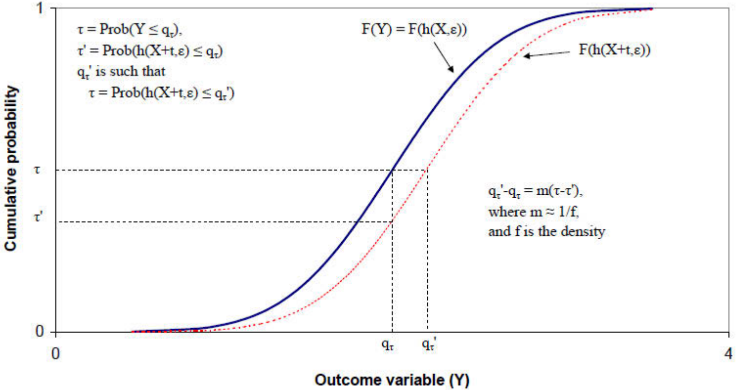
Simple, right?

...not really.

- Computationally demanding (especially if you bootstrap SEs!)
- Can be quite sensitive to linear approximation of $Q_Y(\theta_i|D_i, W_i)$
- Curse of dimensionality: $\widehat{f}_W(w)$ can be poorly estimated

"RIF-ing" QR: Firpo, Fortin, and Lemieux (2009)

Graphical intuition:



Unconditional effect on the τ^{th} quantile:

$$Q_{Y_1}(\tau) - Q_{Y_0}(\tau) \approx \frac{F_{Y_0}(Q_{Y_0}(\tau)) - F_{Y_1}(Q_{Y_0}(\tau))}{f_{Y_0}(Q_{Y_0}(\tau))}$$

Influence Functions: A Quick Overview

Q: "What happens to statistic $T_X(F)$ if I peturb F by adding mass at x"?
A:

$$IF(x;T_X,F) = \lim_{\varepsilon\to 0} \frac{T_X((1-\varepsilon)F+\varepsilon\delta_x) - T_X(F)}{\varepsilon}$$

- Ex. 1: $T_X(F) = E_{X\sim F}[X_i]$:

$$\begin{aligned} IF(x;T_X,F) &= \lim_{\varepsilon\to 0} \frac{E_{X\sim(1-\varepsilon)F+\varepsilon\delta_x}[X_i] - E_{X\sim F}[X_i]}{\varepsilon} \\ &= \lim_{\varepsilon\to 0} \frac{(1-\varepsilon)E_{X\sim F}[X_i] + \varepsilon E_{X\sim\delta_x}[X_i] - E_{X\sim F}[X_i]}{\varepsilon} \\ &= \lim_{\varepsilon\to 0} \frac{-\varepsilon E_{X\sim F}[X_i] + \varepsilon E_{X\sim\delta_x}[X_i]}{\varepsilon} = x - \mu_X \end{aligned}$$

- Ex. 2: $T_Y(F) = Q_{Y;F}(\tau)$:

$$IF(y;T_Y,F) = \frac{\tau - \mathbf{1}\{y \le Q_{Y;F}(\tau)\}}{f_Y(Q_{Y;F}(\tau))}$$

Recentered Influence Functions

- FFL define:

$$RIF(y; Q_{Y;F}(\tau), F_Y) = Q_{Y;F}(\tau) + \frac{\tau - \mathbf{1}\{y \leq Q_{Y;F}(\tau)\}}{f_Y(Q_{Y;F}(\tau))}$$

- Note the expectation of $RIF(x; T_X, F)$ is just $T_X(F)$:

$$\begin{aligned} E[RIF(Y_i; Q_{Y;F}(\tau), F_Y)] &= Q_{Y;F}(\tau) + \frac{\tau - E[\mathbf{1}\{Y_i \leq Q_{Y;F}(\tau)\}]}{f_Y(Q_{Y;F}(\tau))} \\ &= Q_{Y;F}(\tau) + \frac{\tau - \tau}{f_Y(Q_{Y;F}(\tau))} = Q_{Y;F}(\tau) \end{aligned}$$

- So if $E[RIF(Y_i; Q_{Y;F}(\tau), F_Y)|X_i] = X_i'\beta$,

$$\begin{aligned} Q_{Y;F}(\tau) &= E[RIF(Y_i; Q_{Y;F}(\tau), F_Y)] \\ &= E[E[RIF(Y_i; Q_{Y;F}(\tau), F_Y)|X_i]] \\ &= E[X_i']\beta \end{aligned}$$

- Coefficients of a conditional RIF also describe *unconditional* quantiles

Identifying RIFs

$$E[RIF(Y_i; Q_{Y;F}(\tau), F_Y)|X_i] = Q_{Y;F}(\tau) + \frac{\tau - E[\mathbf{1}\{Y_i \leq Q_{Y;F}(\tau)\}|X_i]}{f_Y(Q_{Y;F}(\tau))}$$
$$= Q_{Y;F}(\tau) + \frac{\tau - (1 - P(Y_i > Q_{Y;F}(\tau)|X_i))}{f_Y(Q_{Y;F}(\tau))}$$
$$= c_\tau + \frac{P(Y_i > Q_{Y;F}(\tau)|X_i)}{f_Y(Q_{Y;F}(\tau))}$$

If $E[RIF(Y_i; Q_{Y;F}(\tau), F_Y)|X_i] = X_i'\beta$,

$$c_\tau + \frac{P(Y_i > Q_{Y;F}(\tau)|X_i)}{f_Y(Q_{Y;F}(\tau))} = X_i'\beta$$
$$\implies E[T_i|X_i] = -a_\tau + f_Y(Q_{Y;F}(\tau))X_i'\beta$$

where $T_i = \mathbf{1}\{Y_i > Q_{Y;F}(\tau)\}$

Estimating RIFs

$$E[T_i|X_i] = -c_\tau + f_Y(Q_{Y;F}(\tau))X_i'\beta$$

So

$$T_i = -c_\tau + f_Y(Q_{Y;F}(\tau))X_i'\beta + \varepsilon_i$$
$$\text{where } E[\varepsilon_i|X_i] = 0$$

A regression!

Estimate (best linear approximation to the) RIF by:

1. Regressing $T_i = \mathbf{1}\{Y_i > Q_{Y;F}(\tau)\}$ on X_i
2. Dividing $\hat{\beta}$ by $\widehat{f_Y}(Q_{Y;F}(\tau))$
3. That's it!

RIF Limitations

- RIF approximation depends crucially on the estimated $\widehat{f_Y}(Q_{Y;F}(\tau))$
- RIF inherently *marginal*: influence f'n describes small changes in X_i
  - MM '05: "What is the avg. difference in quantiles of Y_{1i} and Y_{0i}?" (see also Chernozhukov et al. 2009)
  - FFL '09: "What is the avg. effect on the quantile of Y_i if we were to randomly switch one individual from $D_i = 0$ to $D_i = 1$?"
- As with all decomposition methods, RIFs reflect a "partial equilibrium": changes in D_i holding W_i fixed
- ...but at least it can describe the unconditional distribution!

Bonus: Mundlak as OVB

The Mundlak Decomposition

As David showed in class, the fixed-effects regression

$$Y_{ij} = \alpha + r^l S_{ij} + \mu_j + \varepsilon_{ij}$$

implies a decomposition of the coefficient from regressing Y_{ij} on S_{ij}:

$$r^s = r^l + \lambda b$$

where

$$\lambda = \frac{Cov(\mu_j, \bar{S}_j)}{Var(\bar{S}_j)}$$
$$b = \frac{Cov(\bar{S}_j, S_{ij})}{Var(S_i)}$$

We can think of λ as the return to mean establishment schooling and b as the association between worker and establishment schooling

Mundlak as OVB

We can derive this decomposition from the classical omitted variables bias formula:

$$\underbrace{r^s}_{\text{"short"}} = \underbrace{r^l}_{\text{"long"}} + \underbrace{1}_{\text{"effect of omitted"}} \underbrace{\frac{Cov(\mu_j, S_{ij})}{Var(S_{ij})}}_{\text{"regresion of omitted on included"}}$$

Define

$$\tilde{S}_{ij} = S_{ij} - \bar{S}_j$$

which is the "within establishment" variation in S_{ij} (i.e. the residual from regressing S_{ij} on establishment FEs. By construction

$$\begin{aligned} Cov(\bar{S}_j, S_{ij}) &= Cov(\bar{S}_j, \bar{S}_j + \tilde{S}_{ij}) \\ &= Var(\bar{S}_j) \end{aligned}$$

Mundlak as OVB (cont.)

Therefore,

$$r^s = r^l + \frac{Cov(\mu_j, \bar{S}_j + \tilde{S}_{ij})}{Var(\bar{S}_j + \tilde{S}_{ij})} = r^l + \frac{Cov(\mu_j, \bar{S}_j + \tilde{S}_{ij})}{Var(\bar{S}_j)} \frac{Var(\bar{S}_j)}{Var(\bar{S}_j + \tilde{S}_{ij})}$$
$$= r^l + \frac{Cov(\mu_j, \bar{S}_j)}{Var(\bar{S}_j)} \frac{Cov(\bar{S}_j, S_{ij})}{Var(\bar{S}_j)}$$

since $Cov(\mu_j, \tilde{S}_{ij}) = 0$, also by construction. This is Mundlak.
We can also use OVB intuition to estimate this decomposition; note that

$$r^s = r^l + \lambda \frac{Cov(\bar{S}_j, S_{ij})}{Var(S_i)}$$

is the OVB formula for the "long" regression of

$$Y_{ij} = \alpha^l + r^l S_{ij} + \lambda \bar{S}_j + \varepsilon^l_{ij}$$

which we can run to estimate λ(and then solve for b)!

14.662 Labor Economics II
Spring 2015