14.387 Recitation 1

Expectations, Regressions, and Controls

Peter Hull

Fall 2014

Part 1: Expectations and their properties

One variable

Scalar random variable x:

$$\text{Discrete } x: \quad E[x] \equiv \sum_z zPr(x = z)$$

$$\text{Continuous } x: \quad E[x] \equiv \int zf_x(z)dz$$

$$\text{Variance: } Var(x) \equiv E[(x - E[x])^2]$$

Random or fixed?

Two variables

Scalar random variables x and y:

$$\text{Discrete } y: \quad E[y|x] \equiv \sum_z z Pr(y = z|x)$$

$$\text{Continuous } y: \quad E[y|x] \equiv \int z f_{y|x}(z) dz$$

$$\text{Covariance: } Cov(x,y) \equiv E[(x - E[x])(y - E[y])]$$

Random or fixed?

- x and y are **uncorrelated** when $Cov(x,y) = 0$
- y is **mean-independent** of x when $E[y|x] = E[y]$

Which is stronger?

Two useful properties

- **Linearity**: for fixed a, b, c, and d

$$E[a+bx] = a + bE[x]$$
$$\implies Cov(a+bx, c+dy) = bdCov(x,y)$$

- The **Law of Iterated Expectations**:

$$E[E[y|x]] = E[y]$$

(Sloppy) proof of LIE in continuous case:

$$\begin{aligned} E[E[y|x]] &\equiv \int \left(\int z f_{y|x}(z|w) dz \right) f_x(w) dw \\ &= \int z \int f_{x,y}(w,z) dw dz \\ &= \int z f_y(z) dz \\ &\equiv E[y] \end{aligned}$$

Linearity and LIEing

- Mean independence implies uncorrelatedness:

$$\begin{aligned} E[(x-E[x])(y-E[y])] &= E[E[(x-E[x])(y-E[y])|x]] \\ &= E[(x-E[x])(E[y|x]-E[y])] \\ &= E[(x-E[x])\cdot 0] \\ &= 0 \end{aligned}$$

□

- Covariance with mean-zero r.v.s is the expectation of their product:

$$\begin{aligned} E[(x-E[x])(y-E[y])] &= E[xy-E[x]y-xE[y]+E[x]E[y]] \\ &= E[xy]-E[x]E[y]-E[x]E[y]+E[x]E[y] \\ &= E[xy]-E[x]E[y] \\ &= E[xy], \text{ if either } E[x]=0 \text{or } E[y]=0 \end{aligned}$$

□

Part 2: Regressions, large and small

Bivariate regression

Scalar random variables x_i and y_i:

$$(\alpha, \beta) = \arg\min_{a,b} E[(y_i - a - bx_i)^2]$$
$$\text{FOC:} -2E[(y_i - \alpha - \beta x_i)] = 0$$
$$-2E[(y_i - \alpha - \beta x_i)x_i] = 0$$

or

$$\alpha = E[y_i] - \beta E[x_i]$$
$$\beta E[x_i^2] = E[y_i x_i] - \alpha E[x_i]$$

Substituting:

$$\beta E[x_i^2] = E[y_i x_i] - E[y_i]E[x_i] + \beta E[x_i]^2$$
$$\beta = \frac{E[y_i x_i] - E[y_i]E[x_i]}{E[x_i^2] - E[x_i]^2} = \frac{Cov(y_i, x_i)}{Var(x_i)}$$

Multivariate regression

Scalar random variable y_i and $k \times 1$ random vector x_i:

$$\beta = \arg\min_b E[(y_i - x_i'b)^2]$$
$$\text{FOC:} -2E[x_i(y_i - x_i'\beta)] = 0$$

(A useful matrix-'metrics resource: The Matrix Cookbook)

$$\beta = E[x_i x_i']^{-1} E[x_i y_i]$$

How do we reconcile this with the last slide? (Where did α go? What about $Cov()$ and $Var()$?)

Partialling out

Scalar, mean-zero random variables y_i, x_{1i}, and x_{2i}:

$$(\beta,\gamma) = \arg\min_{b,c} E[(y_i - bx_{1i} - cx_{2i})^2]$$

$$\text{FOC}_\gamma: -2E[x_{2i}(y_i - bx_{1i} - \gamma x_{2i})] = 0$$

$$\text{IFT}: \gamma(b) = \frac{E[x_{2i}(y_i - bx_{\ i})]}{E[x_{2i}^2]}$$

Plug $\gamma(b)$ back in (sometimes called "concentrating out" γ):

$$\beta = \arg\min_b E\left[\left(y_i - bx_{1i} - \frac{E[x_{2i}(y_i - bx_{1i})]}{E[x_{2i}^2]}x_{2i}\right)^2\right]$$

$$= \arg\min_b E\left[\left(\left(y_i - \frac{E[x_{2i}y_i]}{E[x_{2i}^2]}x_{2i}\right) - b\left(x_{1i} - \frac{E[x_{2i}x_{1i}]}{E[x_{2i}^2]}x_{2i}\right)\right)^2\right]$$

A bivariate regression! But of what on what?

Partialling out (cont.)

- Special case of the **Frisch-Waugh (sometimes -Lovell) theorem:** If $x_i = [x_{1i}', x_{2i}']'$, $\tilde{x}_{1i}$ is the residual (vector) from regressing (each component of) x_{1i} on x_{2i}, and $\tilde{y}_i$ is the residual from regressing y_i on x_{2i}, then all three are equivalent:
  1. The component β_1 of $\beta = [\beta_1', \beta_2']'$ from regressing y_i on x_i
  2. $\tilde{\beta}_1$ from regressing y_i on $\tilde{x}_i$
  3. $\bar{\beta}_1$ from regressing $\tilde{y}_i$ on $\tilde{x}_i$
- Partialling out x_{2i} from y_i is unnecessary! Why? Back to our example:

$$y_i = \beta x_{1i} + \gamma x_{2i} + e_i$$

$$\tilde{y}_i = \beta \tilde{x}_{1i} + \tilde{e}_i$$

$$y_i = \beta \tilde{x}_{1i} + \tilde{e}_i + y_i - \tilde{y}_i$$

$$y_i = \beta \tilde{x}_{1i} + \left(\tilde{e}_i + \frac{E[x_{2i} y_i]}{E[x_{2i}^2]} x_{2i} \right)$$

Why must the last line be a *regression* (and not just an *equation*)?

From population to sample

- Regression is a **feature of data**: just like expectation, correlation, etc.
- It's a function of population second moments: so easy to estimate!

$$\hat{\beta} = E_n[x_i x_i']^{-1} E_n[x_i y_i]$$

- A more matrix-y way to write $\hat{\beta}$:

$$\begin{aligned} E_n[x_i x_i']^{-1} E_n[x_i y_i] &= \left(\frac{1}{n} \sum_i x_i x_i' \right)^{-1} \left(\frac{1}{n} \sum_i x_i y_i \right) \\ &= (X'X)^{-1} X'Y \end{aligned}$$

where

$$X = \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

Regression subtlety

- β is a **feature of data**. We know what it is and we know that it (probably) exists, given **any** y_i and x_i.
- We also how to estimate it; we know that (probably) $\hat{\beta} \xrightarrow{p} \beta$ (why?) (where "probably" $\equiv$ "given some innocuous technical conditions")

- ...ok...but then.... what's all the fuss about?
- Some common examples of fuss: "endogeneity," "simultaneity," "omitted variable bias," "selection bias," "measurement error," "division bias," etc. etc. etc.

The fuss.

Part 3: Controls: good and bad

You can't always get what you want

- *reg y x* is **always** going to give you a $\hat{\beta}$ estimating the β satisfying $E[x_i(y_i - x_i'\beta)] = 0$
- But what if this isn't what you want? (When might you want it?)
- Ex: suppose we want β from $y_i = \alpha + \beta x_i + \gamma a_i + \varepsilon_i$, where we know $E[\varepsilon_i | x_i, a_i] = 0$
    - We *reg y x* (maybe throw on a *, r*).
    - What do we get? What does $\hat{\beta}$ plim to? Could it be β?
- Obvious solution: just control for a_i. But why stop there?

Bad controls

- Goal: add *right* controls so that the regression β you get is the β you want (i.e. approximates the CEF you want)
- Ex: We randomly assign schooling $s_i \in \{0,1\}$. Want the *causal* effect of schooling on income y_i (a *causal* CEF)
  - Also measure race $b_i \in \{0,1\}$ and post-schooling occupation $x_i \in \{0,1\}$.
  - What regression should we run?
- Natural choice: β satisfying $E[s_i(y_i - \alpha - \beta s_i)] = 0$
  - Another choice: β satisfying $E[s_i(y_i - \alpha - \beta s_i - \gamma b_i)] = 0$. Better?
  - How about β satisfying $E[s_i(y_i - \alpha - \beta s_i - \delta x_i)] = 0$?

Controlling composition

- Potential outcomes: $\{y_{0i}, y_{\ i}\}$. Observe $y_i = y_{0i} + (y_{\ i} - y_{0i})s_i$
- Bivariate regression:

$$
\begin{aligned}
&E[y_i|s_i = 1] - E[y_i|s_i = 0] \\
&= E[y_{0i} + (y_{\ i} - y_{0i})s_i|s_i = 1] - E[y_{0i} + (y_{\ i} - y_{0i})s_i|s_i = 0] \\
&= E[y_{0i} + (y_{\ i} - y_{0i})|s_i = 1] - E[y_{0i}|s_i = 0] \\
&= E[y_{1i} - y_{0i}|s_i = 1] + (E[y_{0i}|s_i = 1] - E[y_{0i}|s_i = 0]) \\
&= \underbrace{E[y_{1i} - y_{0i}]}_{\text{Average treatment effect}} \qquad \text{(why?)}
\end{aligned}
$$

- Recover the CEF, and the CEF is *causal*.

Controlling composition (cont.)

- Potential occupations: $\{x_{0i}, x_{\ i}\}$. Observe $x_i = x_{0i} + (x_{\ i} - x_{0i})s_i$.
- Suppose three types T_i:
  1. *Always-zeros* ($T_i = AZ$): $x_{0i} = 0$, $x_{1i} = 0$
  2. *Always-ones* ($T_i = AO$): $x_{0i} = 1$, $x_{1i} = 1$
  3. *Switchers* ($T_i = SW$): $x_{0i} = 0$, $x_{1i} = 1$
- β satisfying $E[s_i(y_i - \alpha - \beta s_i - \delta x_i)] = 0$ will be a weighted average of
  1. β_0 satisfying $E[s_i(y_i - \alpha_0 - \beta_0 s_i)|x_i = 0] = 0$
  2. β_1 satisfying $E[s_i(y_i - \alpha_1 - \beta_1 s_i)|x_i = 1] = 0$
- Why? Think fixed-effects, or work through Frisch-Waugh algebra

Controlling composition (cont.)

- β (similar for β_0):

$$
\begin{aligned}
& E[y_i|s_i = 1, x_i = 1] - E[y_i|s_i = 0, x_i = 1] \\
& = E[y_{0i} + (y_{\ i} - y_{0i})|s_i = 1, x_i = 1] - E[y_{0i}|T_i = AO] \\
& = \underbrace{E[y_{\ i} - y_{0i}|T_i = AO \vee (T_i = SW \wedge s_i = 1)]}_{\text{Weighted avg. of type-specific treatment effects}} \\
& + \underbrace{E[y_{0i}|T_i = AO \vee (T_i = SW \wedge s_i = 1)] - E[y_{0i}|T_i = AO]}_{\text{Bias (no causal interpretation)}}
\end{aligned}
$$

- Recover the CEF (why?), but it's not a CEF we want (not causal)
- When would this CEF be causal?

14.387 Applied Econometrics: Mostly Harmless Big Data
Fall 2014