Conditional Expectation Functions

Econometrics II

Douglas G. Steigerwald

UC Santa Barbara

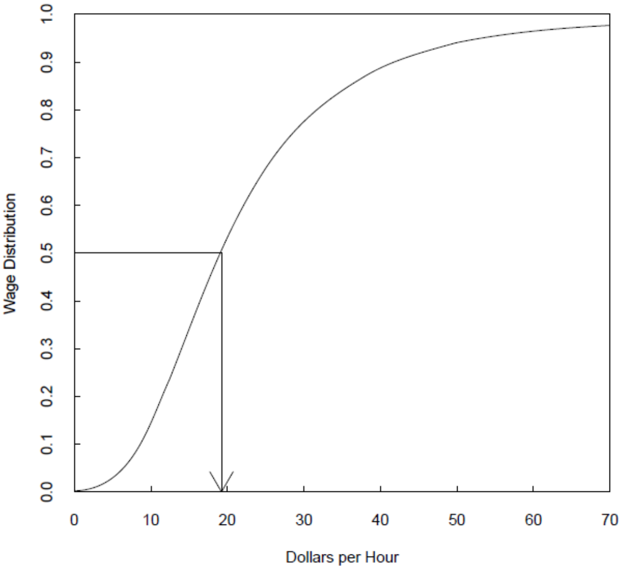
Overview

Reference: B. Hansen Econometrics Chapters 1 and 2.0 - 2.8

- most commonly applied econometrics tool
  - least-squares estimation (regression)
- tool to estimate
  - approximate conditional mean of dependent variable
  - as a function of covariates (regressors)
  - $(y, x_1, \ldots, x_K) := (y, x^{\mathrm{T}})$
- data is observational *not* experimental
  - causality is difficult to infer
- example - wages
  - random variable before measurement
  - observed wages are outcomes of the random variable
  - underpins the application of statistics to economics

Distribution of Wages

- probability distribution
  - $F(u) = \mathbb{P}(wage \leq u)$
- median - measure of location (central tendency)
  - If F is continuous, m uniquely solves $F(m) = \frac{1}{2}$
  - Otherwise, $m = \inf\left\{u : F(u) \geq \frac{1}{2}\right\}$
  - not a linear operator, some calculations are tricky
  - robust to tail perturbations
- nonparametric distribution estimate (following slide)
  - 50,742 full-time non-military wage earners March 2009 CPS
  - $\widehat{m} = \$19.23$


Wage Distribution
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
10
20
30
40
50
60
70
Dollars per Hour

Quantiles

a useful way to summarize a probability distribution

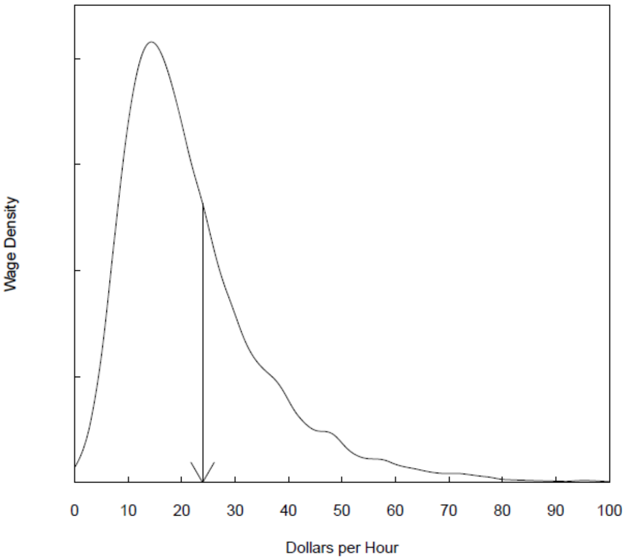
- for any $\alpha \in (0, 1)$, the α^{th} quantile is
  - If F is continuous, q_α uniquely solves $F(q_\alpha) = \alpha$
  - Otherwise, $q_\alpha = \inf \{u : F(u) \geq \alpha\}$
  - $q_{0.5} = m$
- quantile function, q_α, viewed as a function of α is the inverse of F
- if α is represented in percentage terms (10% instead of .1), quantiles are referred to as percentiles
  - $q_{0.5} = m$ is called the 50th percentile
  - $q_{0.9}$ is called the 90th percentile

Density of Wages

If F is differentiable, density function exists

$$f(u) = \frac{d}{du} F(u)$$

- $F(u)$ and $f(u)$ contain the same information
- density is easier to interpret visually
- mean - measure of location
  - if F is continuous, $\mu := \mathbb{E}(u) = \int_{-\infty}^{\infty} uf(u)\, du$
  - formal definition, 240A Lecture on Random Variables and Distributions
  - linear operator, not robust
- nonparametric density estimate for wages (following slide)
- $\widehat{\mu} = \$23.90$
- data are skew, 64% of workers earn less than $\widehat{\mu}$

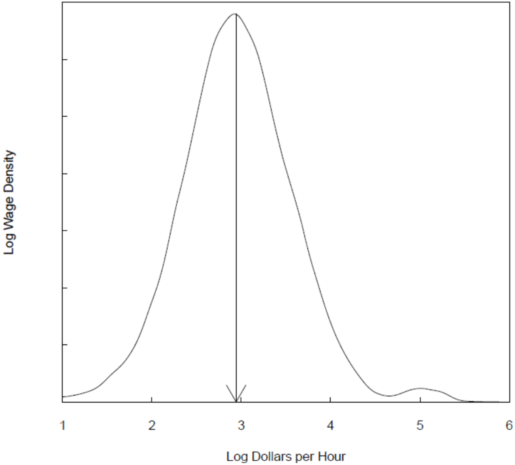


Wage Density
0
10
20
30
40
50
60
70
80
90
100
Dollars per Hour

Density of Log Wage

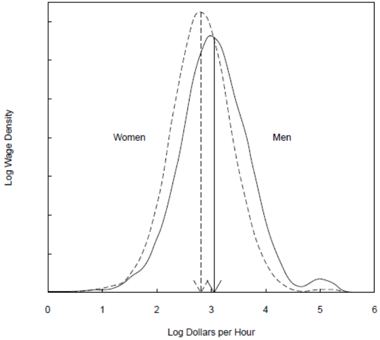
gains can be made by taking the natural logarithm of wages

- skewness and thick tails can be reduced
- $\widehat{\mu}$ for log wage is a much better measure of central tendency



Conditional Expectations

Is the wage distribution the same for all workers?



- Men versus women (43% of workers are women)
  - $\mathbb{E}\left(\log\left(wage\right)\left|gender = man\right.\right) = 3.05$
  - $\mathbb{E}\left(\log\left(wage\right)\left|gender = woman\right.\right) = 2.81$
  - 24% difference in average wages between men and women

Toolkit: Log Differences

If y^* is c% greater than y

$$\begin{aligned} y^* &= \left(1 + \frac{c}{100}\right) y \\ \log y^* - \log y &= \log\left(1 + \frac{c}{100}\right) \approx \frac{c}{100} \end{aligned}$$

key logic $\log(1 + x) \approx x$

- example: $100*(\log w - \log z) = c$
  - then w is approximately c% larger than z
  - approximation is quite good for $|c| \leq 10$

Approximation Accuracy

Gender and Race

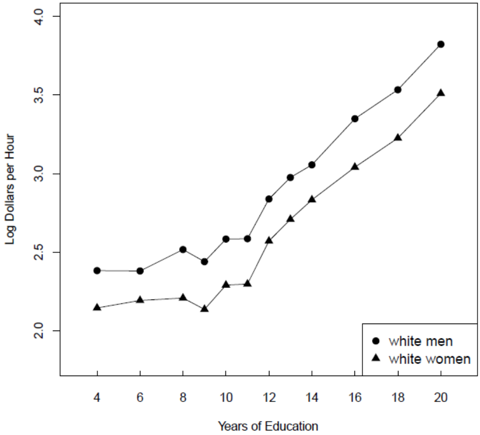
- conditional means reduce distributions to a single summary measure
  - primary focus of regression analysis
  - major focus of econometrics

Conditional Means

	White	Black	Other
Men	3.07	2.86	3.03
Women	2.82	2.73	2.86

- male-female wage gap
  - 25% for whites 13% for blacks
- black-white wage gap
  - 21% for men 9% for women

Education



- after 9 years, conditional mean increases at a different rate
- male-female gap is constant across education levels
  - constant percentage difference in wages

Conditional Expectation Function

Discrete Conditioning Variables

CEF

$$\mathbb{E}\left(\log\left(wage\right)|gender, race, education\right)$$

simplify notation

$$\mathbb{E}\left(y|x_1, x_2, \ldots, x_k\right) = m\left(x_1, x_2, \ldots, x_k\right)$$

for $x = \left(x_1, x_2, \ldots, x_k\right)^{\mathrm{T}}$

$$\mathbb{E}\left(y|x\right) = m\left(x\right)$$

- CEF $\mathbb{E}\left(y|x\right)$ is a random variable because it is a function of x
- given x, it is not random
  $\mathbb{E}\left(\log\left(wage\right)|gender = man, race = white, education = 12\right) = 2.84$

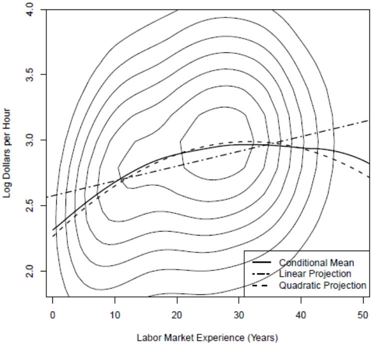
Conditional Expectation Function

Continuous Variables with Joint Density Function

$f(y, x)$ is the joint density function for (y, x)

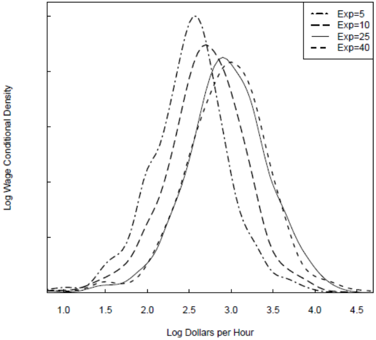
- $y = \log(wage) \quad x = experience$

for white men with 12 years of education contours of $f(y, x)$ are



Conditional Density

a "slice" of the joint density contours yields the conditional density

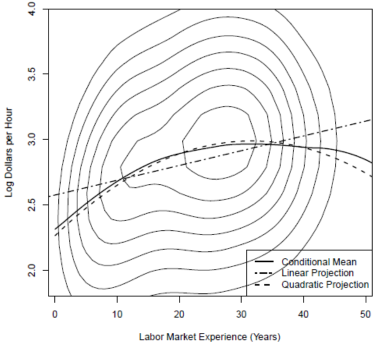


- shifts right and becomes more diffuse as experience increases
  - little change as experience increases past 25 years
- the conditional density is denoted $f_{y|x}(y|x)$ Conditional Density

Conditional Expectation Function

$$m(x) := \mathbb{E}(y|x) = \int_{\mathbb{R}} y f_{y|x}(y|x)\, dy$$

- mean of idealized subpopulation with value x
  - x continuous implies this subpopulation is infinitely small



- conditional mean (CEF) is nonlinear

Error

define: CEF error $\quad e = y - m(x)$

$$y = m(x) + e$$
$$\mathbb{E}(e|x) = 0$$

note, $\mathbb{E}(e|x) = 0$ is not a restriction, these equations hold by definition

Error Properties Theorem: (derived from $f(y, x)$)

1. $\mathbb{E}(e|x) = 0 \Rightarrow \mathbb{E}(e) = 0$
2. $\mathbb{E}(h(x)\, e) = 0 \qquad \text{if } \mathbb{E}\,|h(x)\, e| < \infty$
3. $\mathbb{E}\,|y|^r < \infty \Rightarrow \mathbb{E}\,|e|^r < \infty \qquad (r \geq 1)$

Error Properties

1. $\mathbb{E}\left(e|x\right) = 0$

   - *not* a restriction, but a definition
   - called mean independence
     - mean independence $\not\Rightarrow$ independence
     - $e = x\epsilon$ with $\epsilon \sim \mathcal{N}\left(0, 1\right)$ independent of $x \Rightarrow e|x \sim \mathcal{N}\left(0, x^2\right)$
     - empirics : e and x are rarely assumed independent

2. $\mathbb{E}\left(h\left(x\right)e\right) = 0$

   - e is uncorrelated with any function of the covariates

3. $\mathbb{E}\left|y\right|^r < \infty \Rightarrow \mathbb{E}\left|e\right|^r < \infty$

   - $\mathbb{E}y^2 < \infty \Rightarrow Var\left(e\right) < \infty$

Toolkit: Law of Iterated Expectations

To show property 1

Simple Law

$$\mathbb{E}\left(\mathbb{E}\left(y|x\right)\right) = \mathbb{E}\left(y\right) \qquad \text{if } \mathbb{E}\left(y\right) < \infty$$

note $\mathbb{E}\left(\mathbb{E}\left(y|x\right)\right) = \int_{\mathbb{R}} \mathbb{E}\left(y|x\right) f_x\left(x\right) dx$

General Law (allows for 2 sets of conditioning variables)

$$\mathbb{E}\left(\mathbb{E}\left(y|x_1, x_2\right)|x_1\right) = \mathbb{E}\left(y|x_1\right) \qquad \text{if } \mathbb{E}\left(y\right) < \infty$$

- "smaller information set wins"

Toolkit: Conditioning Theorem

To show property 2

condition on $x \longrightarrow$ effectively treat x as constant
Simple Property

$$\mathbb{E}\left(g\left(x\right)|x\right) = g\left(x\right) \qquad \text{for any function } g\left(\cdot\right)$$

example $\mathbb{E}\left(x|x\right) = x$

General Property (allows for an additional random variable)

$$\mathbb{E}\left(g\left(x\right)y|x\right) = g\left(x\right)\mathbb{E}\left(y|x\right)$$

$$\mathbb{E}\left(g\left(x\right)y\right) = \mathbb{E}\left(g\left(x\right)\mathbb{E}\left(y|x\right)\right)$$

Proofs

1. Proof of Simple LIE
2. Proof of General LIE
3. Proof of Conditioning Theorems
4. Proof of Error Properties Theorem

Review

- Implication of observational data?
- causality is difficult to infer

Should we model $\mathbb{E}(y|x)$ as linear in x?

- no

What are the key properties of $e = y - \mathbb{E}(y|x)$?

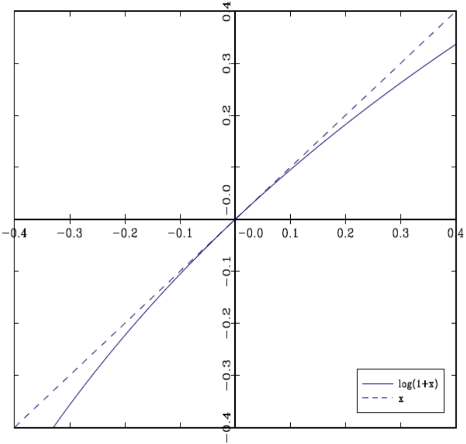
- $\mathbb{E}(e|x) = 0 \qquad$ (by construction)
- uncorrelated with any function of x

Approximation Accuracy

Taylor Series expansion

$$log(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \cdots = x + O\left(x^2\right) \approx x$$

highly accurate if $|x| \leq .1$



Landau Notation (Big O)

we say $f(x) = O(g(x))$ as $x \to 0$ if

$$|f(x)| \leq M|g(x)| \text{ for all } x \leq x_0$$

Consider $f(x) = -\frac{1}{2}x^2 + \frac{1}{3}x^3$

$$\left|-\frac{1}{2}x^2 + \frac{1}{3}x^3\right| \leq \frac{1}{2}\left|x^2\right| + \frac{1}{3}\left|x^3\right|$$

for suitably chosen x_0, if $x \leq x_0$

$$\frac{1}{2}\left|x^2\right| + \frac{1}{3}\left|x^3\right| \leq \frac{1}{2}\left|x^2\right| + \frac{1}{3}\left|x^2\right| = \frac{5}{6}x^2$$

so $f(x) = O\left(x^2\right)$ as $x \to 0$

Return to Log Wage

Definition of Conditional Density

- if (y, x) have joint density $f(y, x)$ then
  - x has marginal density

$$f_x(x) = \int_{\mathbb{R}} f(y, x)\, dy$$

- for any x such that $f_x(x) > 0$, the conditional density of y given x is defined as

$$f_{y|x}(y|x) = \frac{f(y, x)}{f_x(x)}$$

- consider $f(\log(wage)|\text{experience} = 5)$

$$\frac{f(y, x = 5)}{\mathbb{P}(x = 5)} \frac{\leftarrow \text{ the "slice"}}{\leftarrow \text{ the scale factor}}$$

  - if their are fewer individuals with 5 years of experience than with 10 years of experience, the higher conditional density could correspond to workers with 5 years of experience, even if the joint density is higher for workers with 10 years of experience

Return to Conditional Density

Proof of Simple Law of Iterated Expectations

Simple LIE: $\mathbb{E}\left(\mathbb{E}\left(y|x\right)\right) = \mathbb{E}\left(y\right)$

- assume (y, x) have joint density $f(y, x)$ (for convenience)
  - $\mathbb{E}(y|x)$ is a function of the random variable x only
  - to calculate its expectation, integrate with respect to the density $f_x(x)$ of x

$$\mathbb{E}\left(\mathbb{E}\left(y|x\right)\right) = \int_{\mathbb{R}^k} \mathbb{E}\left(y|x\right) f_x\left(x\right) dx$$

  - which equals (by substitution and by noting that $f_{y|x}(y|x) f_x(x) = f(y, x)$)

$$\begin{aligned}\int_{\mathbb{R}^k} \left(\int_{\mathbb{R}} y f_{y|x}\left(y|x\right) dy \right) f_x\left(x\right) dx &= \int_{\mathbb{R}^k} \int_{\mathbb{R}} y f_{y,x}\left(y, x\right) dy dx \\ &= \mathbb{E}\left(y\right),\end{aligned}$$

because $\int_{\mathbb{R}^k} f_{y,x}(y, x)\, dx = f_y(y)$. $\blacksquare$

Return to Proofs

Proof of General Law of Iterated Expectations

General LIE: $\mathbb{E}\left(\mathbb{E}\left(y|x_1,x_2\right)|x_1\right) = \mathbb{E}\left(y|x_1\right)$

- assume (y, x_1, x_2) have joint density $f(y, x_1, x_2)$ (for convenience)
  - $\mathbb{E}\left(y|x_1,x_2\right)$ is a function of the random variables x_1 and x_2
  - integrate with respect to the density of x_2 given x_1

$$\begin{aligned}
\mathbb{E}\left(\mathbb{E}\left(y|x_1,x_2\right)|x_1\right) &= \int_{\mathbb{R}^{k_2}} \mathbb{E}\left(y|x_1,x_2\right) f\left(x_2|x_1\right) dx_2 \\
&= \int_{\mathbb{R}^{k_2}} \left(\int_{\mathbb{R}} y f\left(y|x_1,x_2\right) dy \right) f\left(x_2|x_1\right) dx_2
\end{aligned}$$

  - note that $f\left(y|x_1,x_2\right) f\left(x_2|x_1\right) = \frac{f(y,x_1,x_2)}{f(x_1,x_2)} \frac{f(x_1,x_2)}{f(x_1)} = f\left(y,x_2|x_1\right)$, so

$$\begin{aligned}
&= \int_{\mathbb{R}^{k_2}} \int_{\mathbb{R}} y f\left(y,x_2|x_1\right) dy dx_2 \\
&= \mathbb{E}\left(y|x_1\right),
\end{aligned}$$

the mean of y given the value of x_1. $\blacksquare$

Return to Proofs

Proof of Conditioning Theorems

General CT 1: $\mathbb{E}\left(g\left(x\right)y|x\right) = g\left(x\right)\mathbb{E}\left(y|x\right)$

- assume $\left(y, x_1, x_2\right)$ have joint density $f\left(y, x_1, x_2\right)$ (for convenience)

$$\begin{aligned}
\mathbb{E}\left(g\left(x\right)y|x\right) &= \int_{\mathbb{R}} g\left(x\right) y f_{y|x}\left(y|x\right) dy \\
&= g\left(x\right) \int_{\mathbb{R}} y f_{y|x}\left(y|x\right) dy \\
&= g\left(x\right)\mathbb{E}\left(y|x\right),
\end{aligned}$$

where $\mathbb{E}\left|g\left(x\right)y\right| < \infty$ is needed to ensure the first equality is well defined. ■

General CT 2: $\mathbb{E}\left(g\left(x\right)y\right) = \mathbb{E}\left(g\left(x\right)\mathbb{E}\left(y|x\right)\right)$

- Proof: application of simple LIE. ■

Return to Proofs

Proof of Error Properties Theorem

Parts 1 and 2 follow trivially

Part 3: $\mathbb{E}\,|y|^r < \infty \Rightarrow \mathbb{E}\,|e|^r < \infty \qquad (r \geq 1)$

- $e = y - m(x)$
- $\left(\mathbb{E}\,|e|^r\right)^{1/r} = \left(\mathbb{E}\,|y - m(x)|^r\right)^{1/r}$
  - $\left(\mathbb{E}\,|y - m(x)|^r\right)^{1/r} \leq \left(\mathbb{E}\,|y|^r\right)^{1/r} + \left(\mathbb{E}\,|m(x)|^r\right)^{1/r}$ (Minkowski's Inequality - generalizes Triangle Inequality)
    - $\mathbb{E}\,|\mathbb{E}(y|x)|^r \leq \mathbb{E}\,|y|^r$ for any $r \geq 1$ (Conditional Expectation Inequality)
  - the two parts on the right are finite by $\mathbb{E}\,|y|^r < \infty$
- $\left(\mathbb{E}\,|e|^r\right)^{1/r} < \infty$ implies $\mathbb{E}\,|e|^r < \infty$. $\blacksquare$

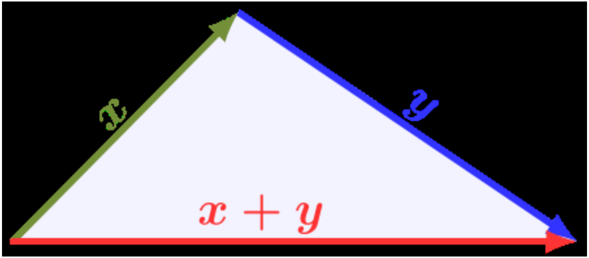
Background : Triangle Inequality

Let x and y be real numbers. The triangle inequality is

$$|x+y| \leq |x| + |y| \, .$$

- Proof
  - $x + y \leq |x| + |y|$
  - $-(x+y) = (-x) + (-y) \leq |x| + |y|$
  - $|x+y| \leq \max\{-(x+y), x+y\}$ $\blacksquare$

To understand why it is called the triangle inequality, let x and y be vectors



Background : Triangle Inequality for a Random Variable

- Let x be a random variable
  - for convenience, x is a discrete random variable
  - set of possible values $\{x_1, x_2, \ldots, x_n\}$
  - with probabilities $\{p_1, p_2, \ldots, p_n\}$

The triangle inequality is

$$|\mathbb{E}x| \leq \mathbb{E}\,|x|\,.$$

- Proof
  - $\left|\sum_{i=1}^{n} x_i p_i\right| \leq \sum_{i=1}^{n} |x_i p_i| = \sum_{i=1}^{n} |x_i|\, p_i \qquad \blacksquare$

Return to Proofs