Specification of Conditional Expectation Functions

Econometrics II

Douglas G. Steigerwald

UC Santa Barbara

Overview

Reference: B. Hansen Econometrics Chapter 2.9-2.17, 2.31-2.32

Why focus on $\mathbb{E}\left(y|x\right)$*?*

- conditional mean is the "best predictor"
- yields marginal effects
  - are the measured marginal effects causal?
- specification of conditional mean
  - exact for discrete covariates as $\mathbb{E}\left(y|x\right)$ is linear in x by definition
- conditional variance is also informative

Why Focus on the CEF?

let $g(x)$ be an arbitrary predictor of y
suppose our goal is to minimize

$$\mathbb{E}\left(y - g(x)\right)^2$$

- if $\mathbb{E}y^2 < \infty$

$$\mathbb{E}\left(y - g(x)\right)^2 \geq \mathbb{E}\left(y - \mathbb{E}(y|x)\right)^2$$

  - conditional mean is the "best predictor"

Proof

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

$$\mathbb{E}(y - g(x))^2 = \mathbb{E}(e + m(x) - g(x))^2$$

- because $\mathbb{E}(e|x) = 0$, e is uncorrelated with any function of x:

$$\mathbb{E}(e + m(x) - g(x))^2 = \mathbb{E}e^2 + \mathbb{E}(m(x) - g(x))^2$$

  - $g(x) = m(x)$ is the minimizing value

Marginal Effects

$\mathbb{E}(y|x)$ can be interpreted in terms of marginal effects

- effect of a change in x_1 holding other covariates constant
  - cannot hold "all else" constant
- causal effect - marginal effect of (continuous) x_1 on y

$$\frac{\partial y}{\partial x_1} = \frac{\partial \mathbb{E}(y|x)}{\partial x_1} + \frac{\partial e}{\partial x_1}$$

- for marginal effects to be causal, we must establish (assume)

$$\frac{\partial e}{\partial x_1} = 0$$

CEF Derivative

- we measure - marginal effect on CEF
- the formula for this derivative differs for continuous and discrete covariates

$$\nabla_1 m(x) = \begin{cases} \frac{\partial}{\partial x_1}\mathbb{E}\left(y|x_1, x_2, \ldots, x_K\right) & \text{if } x_1 \text{ is continuous} \\ \mathbb{E}\left(y|1, x_2, \ldots, x_K\right) - \mathbb{E}\left(y|0, x_2, \ldots, x_K\right) & \text{if } x_1 \text{ is discrete} \end{cases}$$

- effect of a change in x_1 on $\mathbb{E}\left(y|x\right)$ holding other covariates constant
- potentially varies as we change the set of covariates

Effect of Covariates

- changing covariates changes $\mathbb{E}(y|x)$ and e

$$\begin{aligned} y &= \mathbb{E}(y|x_1) + e_1 \\ y &= \mathbb{E}(y|x_1, x_2) + e_2 \end{aligned}$$

  - $\mathbb{E}(y|x_1, x_2)$ reveals greater detail about the behavior of y
- e is the unexplained portion of y
- Adding Covariates Theorem: If $\mathbb{E}y^2 < \infty$

$$Var(y) \geq Var(e_1) \geq Var(e_2)$$

- How restrictive is the finite moment assumption?

Finite Moment Assumption

CEF Specification: No Covariates

- $m(x) = \mu$

$$y = m(x) + e$$

becomes

$$y = \mu + e$$

- intercept-only model

Binary Covariate

- binary covariate

$$x_1 = \begin{cases} 1 \text{ if } gender = man \\ 0 \text{ if } gender = woman \end{cases}$$

  - more commonly termed an indicator variable
- conditional expectation

$$\mu_1 \text{ for men and } \mu_0 \text{ for women}$$

- conditional expectation function is linear in x_1

$$\mathbb{E}\left(y|x_1\right) = \beta_1 x_1 + \beta_2$$

  - 2 covariates (including the intercept)
  - $\beta_2 = \mu_0 \quad \beta_1 = \mu_1 - \mu_0$

Multiple Indicator Variables

- notation: $x_2 = 1\,(\textit{union})$
- there are 4 conditional means

μ_{11} for union men and μ_{10} for nonunion men (and similarly for women)

- conditional expectation is linear in (x_1, x_2, x_1x_2)

$$\mathbb{E}\,(y|x_1) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4$$

  - β_4 mean for nonunion women
  - β_1 male wage premium for nonunion workers
  - β_2 union wage premium for women
  - β_3 difference in union wage premium for men and women
  - x_1x_2 is the interaction term (4 covariates)

- p indicator variables, 2^p conditional means

Categorical Covariates

$$x_3 = \begin{cases} 1 \text{ if } race = white \\ 2 \text{ if } race = black \\ 3 \text{ if } race = other \end{cases}$$

no meaning in terms of magnitude, only indicates category

- $\mathbb{E}(y|x_3)$ is not linear in x_3
- represent x_3 with two indicator variables: $x_4 = 1(black)$ and $x_5 = 1(other)$
- conditional expectation is linear in (x_4, x_5)

$$\mathbb{E}(y|x_4, x_5) = \beta_1 x_4 + \beta_2 x_5 + \beta_3$$

  - β_3 mean for white workers
  - β_4 black-white mean difference
  - β_5 other-white mean difference
  - no individual who is both black and other - no interaction

Continuous Covariates: Linear CEF

- $\mathbb{E}(y|x)$ is linear in x: $\mathbb{E}(y|x) = x^{\mathrm{T}}\beta$

$$y = x^{\mathrm{T}}\beta + e$$

  - $x = (x_1, \ldots, x_{k-1}, 1)^{\mathrm{T}} \qquad \beta = (\beta_1, \ldots, \beta_k)^{\mathrm{T}}$
- derivative of $\mathbb{E}(y|x)$

$$\nabla_x \mathbb{E}(y|x) = \beta$$

  - coefficients are marginal effects
  - marginal effect - impact of change in one covariate holding all other covariates fixed
  - marginal effect is a constant
- general existence of CEF Existence of the Conditional Mean

Measures of Conditional Distributions

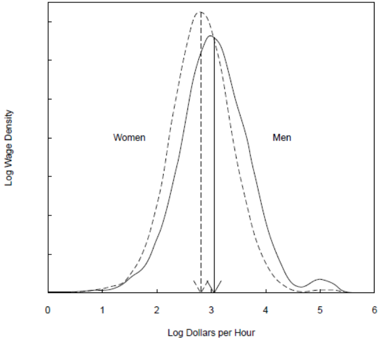
- common measure of location - conditional mean
- common measure of dispersion - conditional variance
- conditional variance

$$\sigma^2(x) := Var(y|x) = \mathbb{E}\left((y - \mathbb{E}(y|x))^2 |x\right)$$

- $\sigma^2(x) = \mathbb{E}\left(e^2|x\right)$
  - often report $\sigma(x)$, same unit of measure as y
- alternative representation

$$y = \mathbb{E}(y|x) + \sigma(x)\epsilon \qquad \mathbb{E}(\epsilon|x) = 0 \quad \mathbb{E}\left(\epsilon^2|x\right) = 1$$

Conditional Standard Deviation



- $\sigma_{\text{men}} = 3.05 \quad \sigma_{\text{women}} = 2.81$
- men have higher average wage and more dispersion

Heteroskedasticity

homoskedasticity

$$\mathbb{E}\left(e^2|x\right) = \sigma^2 \quad \text{(does not depend on } x\text{)}$$

heteroskedasticity

$$\mathbb{E}\left(e^2|x\right) = \sigma^2\left(x\right) \quad \text{(does depend on } x\text{)}$$

- unconditional variance $\mathbb{E}\left(\mathbb{E}\left(e^2|x\right)\right)$ constant by construction
    - formally, *conditional* heteroskedasticity
- heteroskedasticity is the leading case for empirical analysis

Proofs

1. Proof of Adding Covariates Theorem

Review

- Why focus on $\mathbb{E}\left(y|x\right)$?
- "best" predictor

Suppose $\mathbb{E}\left(y|x\right) = x^{\mathrm{T}}\beta$. How to do you interpret β?

- $\nabla_x \mathbb{E}\left(y|x\right)$

What is required for causality?

- $\nabla_x e = 0$

When is $\mathbb{E}\left(y|x\right)$ known?

- x discrete

What is the leading empirical approach for dispersion?

- (conditional) heteroskedasticity $\mathbb{E}\left(e^2|x\right) = \sigma^2\left(x\right)$

Finite Moment Assumption

- Consider the family of Pareto densities

$$f(y) = ay^{-a-1} \qquad y > 1$$

  - a indexes the decay rate of the tail
    - larger a implies tail declines to zero more quickly

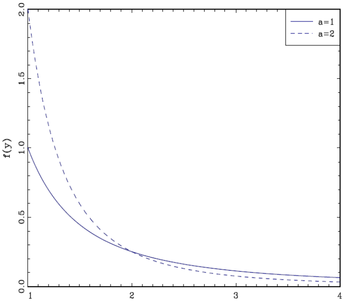


Figure 2.11: Pareto Densities, $a = 1$ and $a = 2$

Tail Behavior and Finite Moments

- for the Pareto densities

$$\mathbb{E}\left|y\right|^r = \left\{ \begin{array}{cc} a\int_1^\infty y^{r-a-1}dy = \frac{a}{a-r} & r < a \\ \infty & r \geq a \end{array} \right\}$$

  - r^{th} moment is finite iff $r < a$

- extend beyond Pareto distribution using tail bounds
  - $f\left(y\right) \leq A\left|y\right|^{-a-1}$ for some $A < \infty$ and $a > 0$
    - ★ $f\left(y\right)$ is bounded below a scale of a Pareto density
    - ★ for $r < a$

$$\begin{aligned} \mathbb{E}\left|y\right|^r &= \int_{-\infty}^{\infty}\left|y\right|^r f\left(y\right)dy \leq \int_{-1}^{1} f\left(y\right)dy + 2A\int_1^\infty y^{r-a-1}dy \\ &\leq 1 + \frac{2A}{a-r} < \infty \end{aligned}$$

Tail Behavior

- if the tail of a density declines at rate $|y|^{-a-1}$ or faster
  - then y has finite moments up to (but not including) a
- intuitively, restriction that y has finite r^{th} moment means the tail of the density declines to zero faster than y^{-r-1}
  - finite mean (but not variance), density declines faster than $\frac{1}{y^2}$
  - finite variance (but not third moment), density declines faster than $\frac{1}{y^3}$
  - finite fourth moment, density declines faster than $\frac{1}{y^5}$

Return to Covariates

Conditional Mean Existence

- If $\mathbb{E}\,|y| < \infty$ then there exists a function $m(x)$ such that for all measurable sets $\mathcal{X}$

$$\mathbb{E}\left(1\left(x \in \mathcal{X}\right) y\right) = \mathbb{E}\left(1\left(x \in \mathcal{X}\right) m\left(x\right)\right) \qquad (1)$$

  - from probability theory e.g. Ash (1972) Theorem 6.3.3
  - $m(x)$ is almost everywhere unique
    - if $h(x)$ satisfies (1) then there is a set $\mathcal{S}$ such that $\mathbb{P}(\mathcal{S}) = 1$ and $m(x) = h(x)$ for $x \in \mathcal{S}$
    - $m(x)$ is called the conditional mean and is written $\mathbb{E}(y|x)$
    - (1) establishes $\mathbb{E}(y) = \mathbb{E}(\mathbb{E}(y|x))$

General Nature of Conditional Mean

- $\mathbb{E}(y|x)$ exists for all finite mean distributions
  - y can be discrete or continuous
  - x can be scalar or vector valued
  - components of x can be discrete or continuous
- if (y, x) have a joint continuous distribution with density $f(y, x)$ then
  - the conditional density $f_{y|x}(y|x)$ is well defined
  - $\mathbb{E}(y|x) = \int_{\mathbb{R}} y f_{y|x}(y|x)\, dy$

Return to Conditional Mean

Proof of Adding Covariates Theorem

Theorem: If $\mathbb{E}y^2 < \infty$,

$$Var\,(y) \geq Var\,(y - \mathbb{E}\,(y|x_1)) \geq Var\,(y - \mathbb{E}\,(y|x_1, x_2))$$

- $\mathbb{E}y^2 < \infty$ implies existence of all conditional moments in the proof
- First establish $Var\,(y) \geq Var\,(y - \mathbb{E}\,(y|x_1))$
  - let $z_1 = \mathbb{E}\,(y|x_1)$
  - $y - \mu = (y - z_1) + (z_1 - \mu)$
    - ★ note $\mathbb{E}\,((z_1 - \mu)\,(y - z_1)\,|x_1) = 0$
  - $\mathbb{E}\,(y - \mu)^2 = \mathbb{E}\,(y - z_1)^2 + \mathbb{E}\,(z_1 - \mu)^2$
    - ★ y and z_1 both have mean μ
  - $Var\,(y) = Var\,(y - z_1) + Var\,(z_1)$
- $Var\,(y) \geq Var\,(y - \mathbb{E}\,(y|x_1))$

Completion of Proof

- Second establish $Var\left(y - \mathbb{E}\left(y|x_1\right)\right) \geq Var\left(y - \mathbb{E}\left(y|x_1, x_2\right)\right)$
  - let $z_2 = \mathbb{E}\left(y|x_1, x_2\right)$ also with mean μ
    - ★ note $\mathbb{E}\left(\left(z_2 - \mu\right)\left(y - z_2\right)|x_1, x_2\right) = 0$
  - $Var\left(y\right) = Var\left(y - z_2\right) + Var\left(z_2\right)$
- need to show $Var\left(z_2\right) \geq Var\left(z_1\right)$
  - $\left(\mathbb{E}\left(z_2|x_1\right)\right)^2 \leq \mathbb{E}\left(z_2^2|x_1\right)$ (conditional Jensen's inequality)
  - $\mathbb{E}\left(\mathbb{E}\left(z_2|x_1\right)\right)^2 \leq \mathbb{E}\left(\mathbb{E}\left(z_2^2|x_1\right)\right)$ (unconditional expectations)
    - ★ $\mathbb{E}\left(z_2|x_1\right) = z_1 \quad \mathbb{E}\left(\mathbb{E}\left(z_2^2|x_1\right)\right) = \mathbb{E}\left(z_2^2\right)$ (LIE)
  - $Var\left(z_1\right) \leq Var\left(z_2\right)$
- $Var\left(y - \mathbb{E}\left(y|x_1\right)\right) \geq Var\left(y - \mathbb{E}\left(y|x_1, x_2\right)\right). \quad \blacksquare$

Return to Proofs