Regression

Econometrics II

Douglas G. Steigerwald

UC Santa Barbara

Overview

Reference: B. Hansen Econometrics Chapter 2.20-2.23, 2.25-2.29

- Best Linear Projection is often called Regression
  - coined by Francis Galton 1886
- Regression coefficient: $\beta = (Var(x))^{-1} Cov(x, y)$
  - regression measures correlation not causality
- (y, x) jointly normally distributed yields "classic regression model"
- short regression coefficient is often a bound ("omitted variable bias")
- random coefficient model implies $\mathbb{E}(y|x)$ is linear in x

Origin

Francis Galton

- Galton: introduces regression, correlation, standard deviation
- 1886 paper studies joint distribution of heights (y, x)
  - y childrens height x parents height
  - in essence, he estimates $\mathbb{E}(y|x)$ - slope is $2/3$
    - on average, children don't attain the height of their parents
    - regression to the mean
- he deduces that this conclusion is a fallacy if the marginal distribution of y and x are the same (heights are stable)
- if heights are stable, $\mu_y = \mu_x = \mu$
  - childrens heights are a linear combination of average height and parents height
    - $y = x\beta + \alpha + u$
    - $\alpha = (1 - \beta)\mu$
    - $\mathcal{P}(y|x) = (1 - \beta)\mu + x\beta$

Regression Fallacy

- if heights are stable, $Var(y) = Var(x)$
  - slope coefficient is always less than 1
    - ★ $\beta = Cov(x, y) / Var(x) = Corr(x, y)$
    - ★ $-1 \leq Corr(x, y) \leq 1$
    - ★ $\beta < 1$ in general
- regression fallacy: $\beta < 1$ implies $Var(y) < Var(x)$
  - clearly false, we derived $\beta < 1$ with $Var(y)$ equal to $Var(x)$!
  - subtle, Secrist (1933) *The Triumph of Mediocrity in Business*
    - ★ department stores: regress 1920 profit on 1930 profit
    - ★ interprets $\beta < 1$ as triumph of mediocrity
- what condition implies $Var(y) < Var(x)$?
  - $Var(y) = \beta^2 Var(x) + Var(u)$
    - ★ $Var(y) / Var(x) < 1$ if $\beta^2 < 1 - Var(u) / Var(x)$

Reverse Regression

- Galton also noted, could regress x on y
  - reverse regression (parent height on child height)
- $x = y\beta^* + \alpha^* + u^*$
  - $\beta^* = Corr(x, y) = \beta$
  - $\alpha^* = (1 - \beta)\,\mu = \alpha$
- coefficients are identical
  - a natural feature of joint distributions (not causal)
- inverting a projection (or CEF) does not yield a projection (or CEF)
  - $x = y\frac{1}{\beta} - \frac{\alpha}{\beta} - \frac{1}{\beta}u$
    - ★ a valid equation, but not a linear projection (nor a CEF)
    - ★ hence projection of x on y does not have slope $1/\beta$

Regression Coefficients

- in regression, often the intercept is separated, if so

$$y = x^T \beta + \alpha + u$$

  - x does not contain the value 1
  - $\mathbb{E}y = \mathbb{E}x^T \beta + \alpha$ or $\mu_y = \mu_x^{\mathrm{T}} \beta + \alpha$
    - $\alpha = \mu_y - \mu_x^{\mathrm{T}} \beta$
- the regression model is sometimes written as
  - $y - \mu_y = (x - \mu_x)^{\mathrm{T}} \beta + u$
    - centered (deviations-from-mean) form
    - $\beta_{lpc} = \left(\mathbb{E}\left((x - \mu_x)(x - \mu_x)^{\mathrm{T}}\right)\right)^{-1} \mathbb{E}\left((x - \mu_x)\left(y - \mu_y\right)\right)$

$$\beta = (Var\,(x))^{-1}\, Cov\,(x, y)$$

Coefficient Decomposition

- $x^{\mathrm{T}} = \left(x_1, x_2^{\mathrm{T}}\right)$ with $\dim\left(x_1\right) = 1$

$$y = x_1\beta_1 + x_2^{\mathrm{T}}\beta_2 + u \quad \textit{Matrix Expression}$$

- $\beta_1 = \left(\mathbb{E}\left(\widetilde{x}\widetilde{x}^{\mathrm{T}}\right)\right)^{-1}\mathbb{E}\left(\widetilde{x}\widetilde{y}\right)$
  - the correlation between x_1 and y after correlation with x_2 is removed
- "Two-Step" Procedure
  - regression 1: x_1 on x_2
    - $x_1 = x_2^{\mathrm{T}}\gamma_2 + u_1$
  - regression 2: y on u_1
    - $\beta_1 = \mathbb{E}\left(u_1 y\right)/\mathbb{E}u_1^2$

Normal Regression

Assume (y, x) are jointly normally distributed, which implies (u, x) are jointly normally distributed

- Best linear projection

$$y = x^{\mathrm{T}}\beta + u \quad \text{where } \beta = \left(\mathbb{E}\left(xx^{\mathrm{T}}\right)\right)^{-1}\mathbb{E}\left(xy\right)$$

- by construction $\mathbb{E}\left(xu\right) = 0$
  - together with jointly normal distribution, implies x and u are independent
    - ★ $\mathbb{E}\left(u|x\right) = \mathbb{E}\left(u\right) = 0$
    - ★ $\mathbb{E}\left(u^2|x\right) = \mathbb{E}\left(u^2\right) = \sigma^2$
- jointly normal distribution yields "classic regression model"

Review: Best Linear Predictor Error

- best linear predictor (linear approximation)

$$\mathcal{P}\left(y|x\right) = x^{\mathrm{T}}\beta_{lpc}$$

- decomposition

$$y = x^{\mathrm{T}}\beta_{lpc} + u \qquad u = e + \left(\mathbb{E}\left(y|x\right) - x^{\mathrm{T}}\beta_{lpc}\right)$$

- error consists of two components
  - e - deviation from conditional mean $\mathbb{E}\left(e|x\right) = 0$
  - $\mathbb{E}\left(y|x\right) - x^{\mathrm{T}}\beta_{lpc}$ error from approximation of conditional mean

Approximation Error

Example 1

- $y = x + x^2 \qquad x \sim \mathcal{N}(0,1)$
  - $\mathbb{E}(y|x) = x + x^2$
  - $e \equiv 0$
- linear approximation

$$y = \beta x + \alpha + u$$

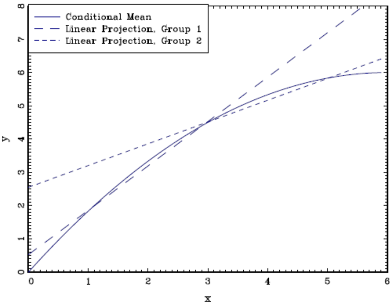
  - $\alpha = \mu_y - \beta\mu_x = 1 \qquad \beta = \mathbb{E}(xy)/\mathbb{E}x^2 = 1$
  - approximation $\mathcal{P}(y|x) = x + 1$ differs sharply from $\mathbb{E}(y|x)$
- linear predictor error

$$u = 0 + \left(x + x^2\right) - (x + 1) = x^2 - 1$$

  - u is a function of x, but uncorrelated with x
    - ⋆ $\mathbb{E}(xu) = \mathbb{E}\left(x^3 - x\right) = 0$

Approximation Error: Grouped Data

- β differs over groups could indicate nonlinear $\mathbb{E}(y|x)$
  - group 1 - $x \sim \mathcal{N}(2,1)$ group 2: $x \sim \mathcal{N}(4,1)$



- here $\nabla_x \mathbb{E}(y|x)$ does not differ over groups
  - $y|x \sim \mathcal{N}(m(x),1)$ $m(x) = 2x - \frac{x^2}{6}$ $\nabla_x \mathbb{E}(y|x) = 2 - \frac{x}{3}$
- linear approximation does differ over groups
  - implication: conditional mean is nonlinear

Approximation Error: Omitted Covariates

- long regression (Goldberger coined these terms)

$$y = x_1^{\mathrm{T}}\beta_1 + x_2^{\mathrm{T}}\beta_2 + u \qquad \mathbb{E}\left(xu\right) = 0 \qquad \beta = \left(\mathbb{E}\left(xx^{\mathrm{T}}\right)\right)^{-1}\mathbb{E}\left(xy\right)$$

- short regression

$$y = x_1^{\mathrm{T}}\gamma_1 + u_1 \qquad \mathbb{E}\left(x_1 u_1\right) = 0 \qquad \gamma_1 = \left(\mathbb{E}\left(x_1 x_1^{\mathrm{T}}\right)\right)^{-1}\mathbb{E}\left(x_1 y\right)$$

  - recall, both coefficient and error change

- the linear projection coefficient from the short regression is
  - $\gamma_1 = \beta_1 + \left(\mathbb{E}\left(x_1 x_1^{\mathrm{T}}\right)\right)^{-1}\mathbb{E}\left(x_1 x_2\right)\beta_2 + \left(\mathbb{E}\left(x_1 x_1^{\mathrm{T}}\right)\right)^{-1}\mathbb{E}\left(x_1 u\right)$
    - ★ $\mathbb{E}\left(xu\right) = 0 \Rightarrow \mathbb{E}\left(x_1 u\right) = 0$
  - $\gamma_1 \neq \beta_1$ unless $\mathbb{E}\left(x_1 x_2\right) = 0$ (or $\beta_2 = 0$)
    - ★ "omitted variable" bias

Short Regression Coefficient

in many cases, short regression coefficient is a bound

- Long regression: linear projection of log(wage) on $x = (ed, ab)$
  - ab intellectual ability, unobserved
- Short regression: linear projection of log(wage) on $x = ed$
  - ed and ab likely positively correlated
  - conditional on ed, ab likely increases wages
  - therefore $\gamma_1 = \beta_1 + c$, $c > 0$
    - ★ an upper bound (not very useful here)

Approximation: Random Coefficient Model

- (linear) random coefficient model yields a linear CEF
- random coefficient model

$$y = x^{\mathrm{T}} \eta$$

  - η individual specific component
    - random
    - independent of x
    - equals $\nabla_x y$, true causal effect
    - i.e. change in response variable due to change in covariate
- Example: $y = \log(wages) \quad x = ed$
  - η individual-specific return to schooling

Random Coefficient Model Yields Linear CEF

- $\eta = \beta + v$
  - $\mathbb{E}(\eta) = \beta \quad Var(\eta) = \Omega$
  - distribution of v is independent of x, mean 0 covariance Ω
- Conditional Expectation Function
  - $\mathbb{E}(y|x) = x^{\mathrm{T}}\mathbb{E}(\eta|x) = x^{\mathrm{T}}\mathbb{E}(\eta) = x^{\mathrm{T}}\beta$

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

  - $e = x^{\mathrm{T}}v$
    - ★ $\mathbb{E}(e|x) = 0$
    - ★ $Var(e|x) = x^{\mathrm{T}}\Omega x$
- (conditional) heteroskedasticity in a regression can be evidence that y is generated by a random coefficient model

Review

- How do we express β in terms of covariances? (consider a single covariate)
- $\beta = Cov\left(x, y\right) / Var\left(x\right)$

How does this change if $Var\left(x\right) = Var\left(y\right)$?

- $\beta = Cov\left(x, y\right) / \left(sd\left(x\right) sd\left(y\right)\right) = Corr\left(x, y\right)$

How does this change, if we regress x on y?

- $\beta = Cov\left(x, y\right) / Var\left(y\right) = Corr\left(x, y\right)$ Identical!

How does the short (regression) coefficient differ from the long coefficient?

- $\gamma_{1(short)} = \beta_{1(long)} + \left(Var\left(x_1\right)\right)^{-1} Cov\left(x_1, x_2\right) \beta_{2(long)}$

When does the short (regression) coefficient provide a useful bound on the long coefficient?

- when the bias attenuated the measured response (sign of $\beta_{1(long)}$ differs from sign of $Cov\left(x_1, x_2\right) \beta_{2(long)}$)

Matrix Expression

- partition regressor matrix, with $Q_{12} = \mathbb{E}\left(x_1 x_2^{\mathrm{T}}\right)$

$$\left(\begin{array}{cc} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{array} \right)^{-1} := \left(\begin{array}{cc} Q^{11} & Q^{12} \\ Q^{21} & Q^{22} \end{array} \right)$$

- $Q^{11} = Q_{11} - Q_{12} Q_{22}^{-1} Q_{21}$
  - variation in the component of x_1 that is uncorrelated with x_2
- $y - x_2 Q_{22}^{-1} \mathbb{E}\left(x_2 y\right)$
  - component of y that is uncorrelated with x_2
- $Q^{12} = -Q^{11} Q_{12} Q_{22}^{-1}$
- $\beta = Q^{-1} \mathbb{E}\left(xy\right)$
  - $\beta_1 = Q^{11} \mathbb{E}\left(x_1 y\right) - Q^{12} \mathbb{E}\left(x_2 y\right)$
    - ★ $Q^{12} = -Q^{11} Q_{12} Q_{22}^{-1}$
  - $\beta_1 = Q^{11} \mathbb{E}\left(x_1 \left(y - x_2 Q_{22}^{-1} \mathbb{E}\left(x_2 y\right)\right)\right)$ *Return to Decomposition*