POLS/CS&SS 503:
Advanced Quantitative Political Methodology

LINEAR REGRESSION ESTIMATOR

April 14, 2015

Jeffrey B. Arnold

UNIVERSITY of WASHINGTON
DEPARTMENT OF POLITICAL SCIENCE

CENTER for STATISTICS and the SOCIAL SCIENCES

Overview

Regression Coefficient Anatomy

Linear Regression Population Model

Sampling Distribution

Estimators and mean squared error

Gauss-Markov Theorem

Coeficients of a simple regression

$$Y_i = A + BX_i + E_i$$

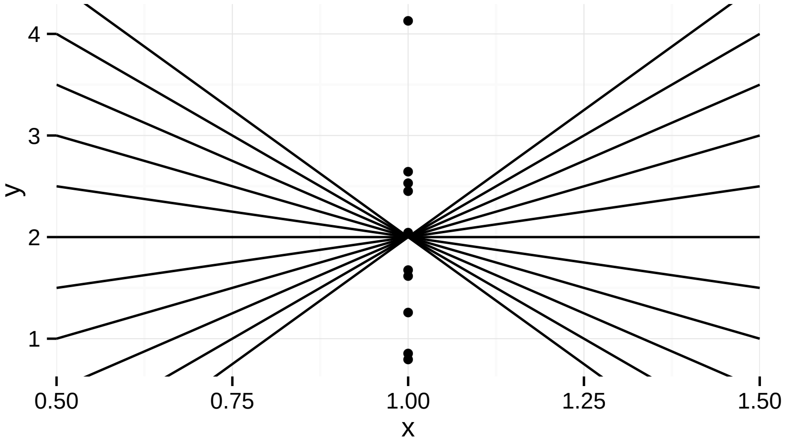
The least squares coefficients are

$$A = \bar{Y} + B\bar{X}$$

$$B = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_i (X_i - \bar{X})^2}$$

- State B in terms of covariance of X and Y and variance?
- State B in terms of correlation of X and Y and standard deviations?
- What values can B take if $\mathrm{SD}(X) = \mathrm{SD}(Y) = 1$?
- What is $\hat{Y}$ for $X = \bar{X}$?
- What happens to B as $\mathrm{V}(X)$ decreases? $\mathrm{V}(Y)$ decreases? If $\mathrm{V}(X) = 0$

Least squares when $\mathbb{V}X = 0$



Least squares coefficients are unidentified if $\mathrm{V}\, x = 0$

- If $\mathrm{V}\, x = 0$ then least squares solution is unidentified
- There is no unique value of A, B that $\arg\min_{A,B} \sum_i E_i^2$

```r
y <- c(1, 2, 3, 4, 5)
x <- 1
ybar <- mean(y)
ybar
```

```
## [1] 3
```

```r
# A = 2, B = 1
sum((y - 2 - 1 * x) ^ 2)
```

```
## [1] 10
```

```r
# A = -7, B = 10
sum((y + 7 - 10 * x) ^ 2)
```

```
## [1] 10
```

Coefficients of a multiple regression

$$\vec{Y} = Xb + e$$

- $b = (X'X)^{-1}X'y$. *Not that intuitive!*
- Coefficient b_j is

$$b_k = \frac{\mathrm{C}(y, \tilde{x}_j)}{\mathrm{V}(\tilde{x}_k)}$$

- Where $\tilde{x}_j$ are the residulals of x_j on all X_h where $h \neq j$

$$\tilde{X}_{j,i} = X_{j,i} - \tilde{A} - \sum_{h \neq j} \tilde{B}_h X_h$$

Regression example

See `multiple_regression_anatomy.R`

Least Squares coefficients are unidentified if $(X'X)^{-1}$ does not exist

- Common cases in which $(X'X)^{-1}$ does not exist:
  - Number of observations less than $k+1$
  - X_k is constant
  - X_k is a linear function of other variables: $X_k = \sum_{j \neq k} c_j X_j$.
    - dummy variables for all categories of a categorical variable
    - variable multiplied by the constant of another variable

Which of these would be cases of collinearity and why?

- There is a variable that takes values “white”, “black”, “hispanic”, “asian”, “other”. You include a dummy variable for each category.
- GDP, GDP per capita, and population
- Log GDP, log GDP per capita, and log population
- GDP in millions of dollars; GDP in trillions of dollars
- GDP measured in nominal value; GDP measured in real terms
- Regression with 3 variables and 4 observations

Regression Coefficient Anatomy

Linear Regression Population Model

Sampling Distribution

Estimators and mean squared error

Gauss-Markov Theorem

Population model in a simple regression

$$Y_i = \alpha + \beta X_i + \epsilon_i$$

Assumptions for statistical inference

1. *X is not invariant:* $\mathrm{V}(X) > 0$
2. *Linearity*. Average value of error given x is 0. $E(\epsilon_i) = E(\epsilon_i | x_i) = 0$

$$\mu_i = E(Y_i) = \mathrm{E}(Y|X_i) = \mathrm{E}(\alpha + \beta X_i + \epsilon_i) = \alpha + \beta x_i$$

3. *Constant variance* Variance of the errors is the same regardless of the value of X

$$\mathrm{V}(Y|x_i) = E(\epsilon_i^2) = \sigma_\epsilon^2$$

4. *Independence:* Observations are sampled independently. $\mathrm{Cor}(\epsilon_i, \epsilon_j) = 0$ for all $i \neq j$.
5. Fixed X or X measured without error and independent of the error.
6. Errors are normally distributed $\epsilon_i \sim N\left(0, \sigma_\epsilon^2\right)$

Population model in a multiple regression

$$Y_i = \alpha + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \ldots + \beta_k x_{i,k} + \epsilon_i$$

Assumptions for statistical inference

1. *X is not invariant* and no X is a perfect linear function of the others.
2. *Linearity*. $\mathrm{E}(\epsilon_i) = 0$
3. *Constant variance* $V(\epsilon_i) = \sigma_\epsilon^2$
4. *Independence Observations are sampled independently*. $\mathrm{Cor}(\epsilon_i, \epsilon_j) = 0$ for all $i \neq j$.
5. *Fixed X or X measured without error and independent of the error*
6. *Normality* Errors are normally distributed $\epsilon_i \sim N\left(0, \sigma_\epsilon^2\right)$

Definitions

population The observations of interest. May be theoretical

sample The data you have.

parameter A function of the population distribution

statistic A function of the sample

sampling distribution The distribution of a statistic calculated from the distribution of samples of a given size drawn from a population.

See `Sampling_Distributions.Rmd`

Sampling Distribution of Simple Regression Coefficients

The sampling distributions of A, B given $Y_i = \alpha + \beta X_i + \epsilon_i$

- expected values (linearity)

$$E(A) = \alpha$$
$$E(B) = \beta$$

- variances (linearity, constant variance, independence)

$$V(A) = \frac{\sigma_\epsilon^2}{n} \cdot \frac{\sum x_i^2}{\sum (x_i - \bar{x})^2}$$
$$V(B) = \frac{\sigma_\epsilon^2}{\sum (x_i - \bar{x})^2} = \frac{\sigma_\epsilon^2}{n\, V(x)}$$

- normal distribution (normal errors)

$$A \sim N(E(A), V(A))$$
$$B \sim N(E(B), V(B))$$

Coefficient sampling distributions in multiple regression

The sampling distributions of B_k given $Y_i = \alpha + \sum \beta_j X_{j,i} + \epsilon_i$

- Expected value: $\mathrm{E}(B_K) = \beta_k$
- Variance:

$$\mathrm{V}(B_j) = \frac{1}{(1 - R_j^2)} \frac{\sigma_\epsilon^2}{\sum (x_{i,j} - \bar{x}_j)^2}$$
$$= \frac{\sigma_\epsilon^2}{\sum_i \left(x_{i,j} - \hat{x}_{i,j}\right)^2}$$

  Where R_j^2 is R^2 from regression of X_j on other X, and $\hat{x}_{ij}$ are fitted values from that regression.

- Normally distributed if errors are normally distributed or as $n \to \infty$.
- b is multivariate normally distributed

$$b \sim N\left(\beta, \sigma_\epsilon^2 (X'X)^{-1}\right)$$

Regression Coefficient Anatomy

Linear Regression Population Model

Sampling Distribution

Estimators and mean squared error

Gauss-Markov Theorem

Let's define some things

statistic: Function of a sample, e.g. Sample mean $\bar{x} = \frac{1}{n} \sum x_i$

parameter: Function of the population distribution, e.g. Expected value μ of the normal distribution.

estimator: Method to use a sample statistic (estimate) to infer a population parameter (estimand)

How to determine if an estimator is good?

- Is $\hat{\beta} = \left(X'X\right)^{-1}X'y$ a good estimator for β?
- Would another estimator be better?
- First, need criteria to by which to judge estimators

What makes an estimator good?

- Bias
- Variance
- Efficiency (mean squared error)
- Consistency

Bias and Variance

Bias

On average how far off is the estimator?

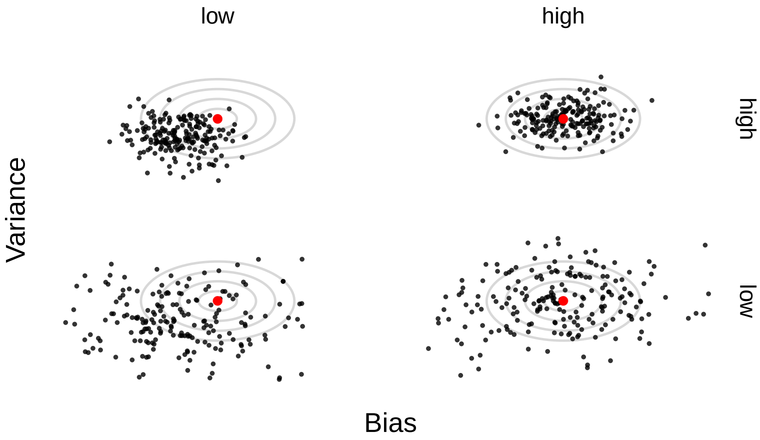
$$\text{bias}(\hat{\beta}) = \mathrm{E}(\hat{\beta}) - \beta$$

Variance

Does the estimator give similar results in different samples?

$$\mathrm{V}(\hat{\beta}) = \mathrm{E}\left((\beta - \mathrm{E}(\hat{\beta}))^2\right)$$

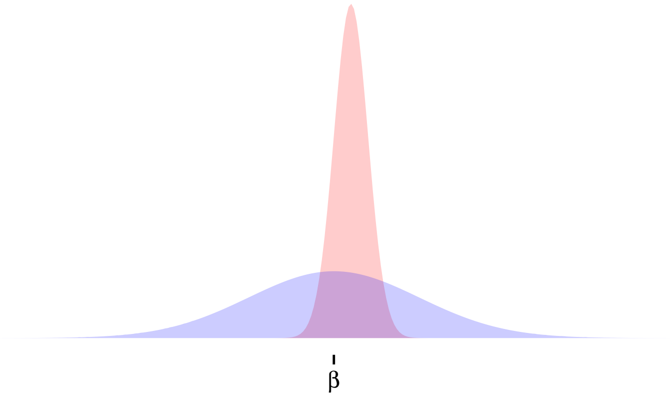
Bias and Variance Visualized



What makes an estimator good?

- Unbiased methods may still miss the truth by a large amount, just direction not systematic
- Unbiased estimates can be horrible: random draw from numbers 0–24 for time of day
- Biased estimates are not necessarily terrible: a clock that's 2 minutes fast

You may prefer a biased, low variance estimator to an unbiased, high variance estimator



Mean Squared Error (MSE)

- MSE is

$$MSE(\hat{\beta}) = \mathrm{E}\left((\hat{\beta} - \beta)^2\right)$$

- MSE trades off bias and variance

$$\begin{aligned} MSE(\hat{\beta}) &= \mathrm{E}((\hat{\beta} - \mathrm{E}(\hat{\beta}))^2) + \mathrm{E}(\mathrm{E}(\hat{\beta}) - \beta))^2 \\ &= \mathrm{V}(\hat{\beta}) + \left(\text{bias}(\hat{\beta}, \beta)\right)^2 \end{aligned}$$

- root mean squared error (RMSE) $\sqrt{\text{MSE}}$: on average how far is an estimate from the truth
- An **efficient** estimator has the smallest MSE
- What is the MSE of an unbiased estimator?

$$MSE(\hat{\beta}) = \mathrm{V}(\hat{\beta}) + \left(\text{bias}(\hat{\beta}, \beta)\right)^2 = \mathrm{V}(\hat{\beta}) + 0 = \mathrm{V}(\hat{\beta})$$

MSE Example

- Suppose population parameter $\beta = 1$
- Consider two estimators $\hat{\beta}_1$ and $\hat{\beta}_2$.
  - $\hat{\beta}_1 \sim N(1, 1^2)$
  - $\hat{\beta}_2 \sim N(0.5, 0.5^2)$
- What are the bias, variance, and MSE of each estimator?

Consistency

- A consistent estimator converges to the parameter value as the number of observations grows

$$E(\hat{\beta} - \beta) \to 0 \qquad \text{as } n \to \infty$$

- A concern of econometricians
- May not be as much a concern in finite, small sample sizes
- We will mainly be concerned with efficiency, secondarily with bias, rarely with consistency

LS assumptions and consequences of violations

	Assumption		Consequence of violation
1	No perfect collinearity	$\text{rank}(X) = k, k < n$	Coefficients unidentified
2	X is exogenous	$\text{E}(X\epsilon) = 0$	Biased, even as $n \to \infty$
3	Disturbances have mean 0	$\text{E}(\epsilon) = 0$	Biased, even as $n \to \infty$
4	No serial correlation	$\text{E}(\epsilon_i \epsilon_j) = 0, i \neq j$	Unbiased but ineff. Wrong se.
5	Homoskedastic errors	$\text{E}(\epsilon' \epsilon') = \sigma^2 I$	Unbiased but ineff. Wrong se.
6	Normal errors	$\epsilon \sim N(0, \sigma^2)$	se wrong unless $n \to \infty$

Assumptions stronger from top to bottom, 4 and 5 could be combined

Unbiasedness of LS

- Only need assumptions 1-3 (no collinearity, X exogenous, $\mathrm{E}(\epsilon) = 0$
- Start with

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'(X\beta + \epsilon) \\ &= \beta + (X'X)^{-1}X'\epsilon\end{aligned}$$

- Take the expectation

$$\begin{aligned}\mathrm{E}(\hat{\beta}) &= \mathrm{E}(\beta) + \mathrm{E}(X'X)^{-1}X'\epsilon) \\ &= \mathrm{E}(\beta) + X'X)^{-1}X'\,\mathrm{E}(\epsilon) \\ &= \mathrm{E}(\beta)\end{aligned}$$

- Since $\mathrm{E}(\hat{\beta}) = \mathrm{E}(\beta)$, LS is unbiased.

Gauss-Markov

- If make assumptions 1–5: LS is the best linear unbiased estimator (BLUE)
- LS estimator is **linear** because $\hat{\beta} = My$, where $M = (X'X)^{-1}X'$
- **best** is best mean squared error (MSE).
- If LS is unbiased, then its mean squared error is the same as its ...?
- Could exist other non-linear unbiased estimators with smaller MSE, e.g. Robust regression when population has fat tailed errors
- If errors are Gaussian, LS is Minimum Variance Unbiased (MVU).
- MVU = for *all* estimators that are unbiased. $\hat{\beta}$ has smallest variance (and MSE).

References

- Some slides derived from Christopher Adolph *Linear Regression in Matrix Form / Propoerties & Assumptions of Linear Regression*. Used with permission. <http://faculty.washington.edu/cadolph/503/topic3.pw.pdf>
- Material included from
    - Fox Ch 6, 9.3
    - Angrist and Pischke, Chapter 3

dasfjasda