Importance of the Classical Linear Model

Small data still matters.

- Experimental data can be expensive
- Data may be aggregated
    - Policy regions
    - Markets
    - Prior studies
- Some units of observation are limited
    - Space shuttle launches
    - Viral pandemics
    - Elections

Why Can't we use the Large-Sample Model?

- Coefficients may have high bias
- Standard error estimators have high variance
- Standard error estimates may have high bias

$\Longrightarrow$ Our coefficients could be far from the truth

$\Longrightarrow$ We can't trust our estimates of uncertainty

Twenty (Noisy) Questions

- Is it in a school?
- Is it bigger than a breadbox?
- Is it red?
- It is alive?

More assumptions

↓

Fewer unknowns

↓

More mileage from data

Five (Noisy) Questions about Pets

The Classical Linear Model (CLM)

Key assumption: $f_{Y|X}$ belongs to a parametric family.

Goal: Identify which member is the true one.

Unit Plan

Plan for the Week

Three sections:

1.
2.
3.

Plan for the Week (cont'd)

At the end of this week, you will be able to:

- Understand statistical inference based on the classical linear model
- Use regression diagnostics to assess all CLM assumptions
- Understand how to leverage transformations to help meet CLM assumptions

Part 1: Importance of the Classical Linear Model

Reading: The CLM Assumptions

Reading: The CLM Assumptions

Read *Foundations of Agnostic Statistics* Chapter 5 through section 5.1.1.

Pay attention to the discussion of the disturbance, and notice how the last paragraph seems to imply a particular causal model.

The CLM Assumptions, Part 1

CLM Assumption 1

I.I.D. Data. $(Y_1, \boldsymbol{X}_1), (Y_2, \boldsymbol{X}_2), \ldots, (Y_n, \boldsymbol{X}_n)$ are independent and identically distributed.

Common violations:

- Clusters
- Dependencies among family members, competitors, geographic neighbors
- Dependencies from one time period to the next

We denote a representative datapoint as $(Y, \boldsymbol{X})$.

CLM Assumption 2

No Perfect Collinearity. $E[X^T X]$ exists and is invertible.

$\implies$ No X_i can be written as a linear combination of the other X's.

Perfect Collinearity Example 1

$$\widehat{Price} = .5\ Donuts + 0.0\ Dozens$$

or

$$\widehat{Price} = 0.0\ Donuts + 6.0\ Dozens$$

Perfect Collinearity Example 2

$$\widehat{Voters} = 200\, Positive\_Ads + 100\, Negative\_Ads + 0\, Total\_Ads$$

or

$$\widehat{Voters} = 100\, Positive\_Ads + 0\, Negative\_Ads + 100\, Total\_Ads$$

The CLM Assumptions, Part 2

CLM Assumption 3

Linear Conditional Expectation. The conditional expectation of Y given $\boldsymbol{X}$ exists and has the linear form,

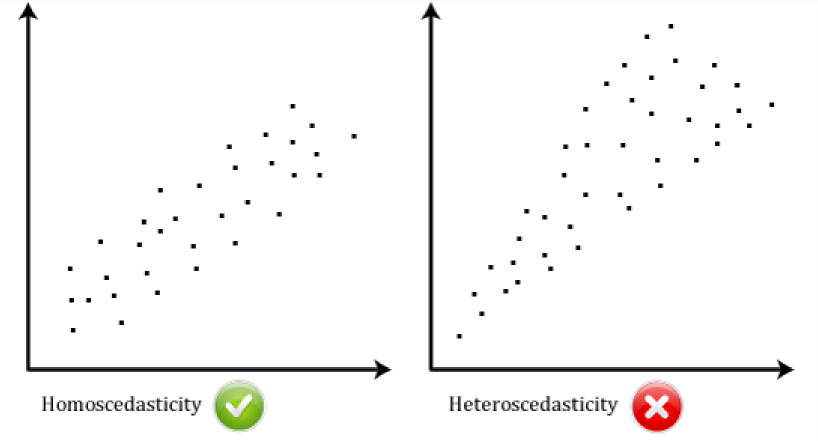
$$\mathrm{E}[Y|\boldsymbol{X} = \boldsymbol{x}] = \boldsymbol{x}\beta$$

Where β is a vector of parameters, for all $\boldsymbol{x} \in \mathrm{Supp}[\boldsymbol{X}]$

CLM Assumption 4

Homoskedasticity. Letting $\epsilon = Y - \boldsymbol{X}\beta$, the conditional variance $\mathbb{V}[\epsilon|\boldsymbol{X}]$ is a constant, which we label σ^2.

Understanding Homoskedasticity



CLM Assumption 5

Normally Distributed Errors.. Letting $\epsilon = Y - \boldsymbol{X}\beta$, the conditional distribution of ϵ given $\boldsymbol{X} = \boldsymbol{x}$ is normal.

- Given previous assumptions, $\epsilon \sim N(0, \sigma^2)$.

Common CLM Errors

NOT CLM Assumptions:

- Normality of X_i or Y
- No outliers
- No high Collinearity

CLM Model Summary

1. I.I.D. Data
2. No Perfect Collinearity
3. Linear Conditional Expectation
4. Homoskedastic Errors
5. Normally Distributed Errors

Part 2: Properties of the CLM

OLS is Unbiased under the CLM

OLS is Unbiased under the CLM

Assume CLM 1-3.

OLS is Unbiased under the CLM

Assume CLM 1-3. $\mathrm{E}[Y|\boldsymbol{X}] = \boldsymbol{X}\beta$

$$\mathrm{E}[\boldsymbol{Y}|\mathbb{X}] = \begin{bmatrix} \mathrm{E}[Y_1|\mathbb{X}] \\ \mathrm{E}[Y_2|\mathbb{X}] \\ \vdots \\ \mathrm{E}[Y_n|\mathbb{X}] \end{bmatrix} = \begin{bmatrix} X_1\boldsymbol{\beta} \\ X_2\boldsymbol{\beta} \\ \vdots \\ X_n\boldsymbol{\beta} \end{bmatrix} = \mathbb{X}\boldsymbol{\beta}$$

$$\begin{aligned}\mathrm{E}[\hat{\boldsymbol{\beta}}] &= \mathrm{E}[(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\boldsymbol{Y}] = \mathrm{E}\Big[\mathrm{E}[(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\boldsymbol{Y}|\mathbb{X}]\Big] \\ &= \mathrm{E}\Big[(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathrm{E}[\boldsymbol{Y}|\mathbb{X}]\Big] = \mathrm{E}\Big[(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbb{X}\boldsymbol{\beta}\Big] \\ &= \mathrm{E}[\boldsymbol{\beta}] = \boldsymbol{\beta}\end{aligned}$$

Classical Standard Errors

Standard Errors under the CLM

Two choices:

- Robust Standard Errors
- Classical Standard Errors

Classical Standard Errors: Intuition

Fewer unknowns $\implies$ more accurate estimates

- All data points share a common variance $\mathbb{V}[Y_i] = \sigma^2$.
- Leverage all data points to estimate a single number.

A Simple Equation for Sampling Variance

Theorem: OLS variance under homoskedasticity

Under CLM 1-4, the variance of the OLS coefficients is given by,

$$\mathrm{V}[\hat{\beta}] = \sigma^2 \mathrm{E}[\boldsymbol{X}^T \boldsymbol{X}]^{-1}$$

Classical Standard Errors

The classical variance estimator is

$$\widehat{\mathbb{V}}_C[\hat{\beta}] = \hat{\sigma}^2(\mathbb{X}^T\mathbb{X})^{-1}$$

Where $\hat{\sigma}^2$ is the *residual variance*, given by

$$\hat{\sigma}^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}\hat{\epsilon}_i^2$$

- Consistent under CLM 1-4
- Unbiased if $\mathbb{X}$ is nonrandom.

Should you use Classical Standard Errors?

YES

- Classical standard errors are efficient.
- If you transform variables to fit CLM, want to take advantage of extra precision.
- t-Tests, other Wald tests are based on classical standard errors.

NO

- Without homoskedasticity, classical standard errors are inconsistent
- Hard to assess homoskedasticity in small samples

Sampling Distributions under the CLM

CLM Sampling Distributions

Normality of OLS Coefficients

Under CLM 1-5, $\hat{\beta}$ is distributed multivariate normal.

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\boldsymbol{\epsilon}$$

- Each β_i is normal
- Any linear combination (e.g. $\beta_i - \beta_j$) is normal

Testing Coefficients under the CLM

Theorem: t-statistics for OLS Coefficients

Assume CLM 1-5 and the null hypothesis:

$H_0 : \beta_i = \mu_0$

Let $\hat{\sigma}_i$ be the classical standard error for $\hat{\beta}_i$.

$$t = \frac{\hat{\beta}_i - \mu_0}{\hat{\sigma}_i}$$

is distributed T with $n - k - 1$ degrees of freedom.

Take-Aways

Under CLM 1-5

- t-tests based on classical standard errors are valid
- Confidence intervals based on classical standard errors are valid

Efficiency Theorems for OLS

OLS and Alternative Estimators

OLS is one of many possible estimators for β:

- Should we weight some datapoints more than others?
- What is the right form for the cost function?
- Should we reduce influence of outliers?

Understanding Efficiency

Desirable estimator properties

- Unbiased: $\mathsf{E}[\hat{\beta}] = \mathsf{E}[\beta]$
- Efficient: $\mathsf{V}[\hat{\beta}]$ is small.

Efficiency: More precision with less data

- Is the OLS estimator efficient?

The Gauss-Markov Theorem

Under CLM 1-4, out of all estimators that are

1. Unbiased
2. Linear (of the form $\mathbb{M}\boldsymbol{Y}$ for some random matrix $\mathbb{M}$)

OLS has the minimum variance.

Remember the phrase: OLS is BLUE

- Best
- Linear
- Unbiased
- Estimator

The Rao-Blackwell Theorem

Under CLM 1-5, out of all estimators that are

1. Unbiased

OLS has the minimum variance.

Reading Assignment

Reading Assignment

Make sure you remember the material in section 5.2 through page 189.

Then read the last paragraph of page 189 through 191.

Maximum Likelihood Estimation of the CLM

Likelihood

Likelihood: A function that takes values for a model's parameter's as inputs, and yields the probability of the (fixed) data as output.

Maximum Likelihood Estimators

If the model is true:

- Consistent and asymptotically efficient.

If the model is not true:

- Consistently estimates the parameters that minimize KL divergence.

Maximum Likelihood Estimation of the CLM

Data $\boldsymbol{Y}, \mathbb{X}$. Find ML estimator for CLM

$$L(b, s|\boldsymbol{Y}, \mathbb{X}) = \prod_{i=1}^{n} \phi\big(Y_i, (\boldsymbol{X_i b}, s^2)\big)$$

$$\ln(L) = \ln \prod_{i=1}^{n} \phi(\ldots) = \sum_{i=1}^{n} \ln \phi(\ldots)$$

$$= \sum_{i=1}^{n} ln \frac{1}{s\sqrt{2\pi}} e^{\frac{-(Y_i - \boldsymbol{X_i b})^2}{2s^2}}$$

$$= \sum_{i=1}^{n} \left[ln \frac{1}{s\sqrt{2\pi}} - \frac{(Y_i - \boldsymbol{X_i b})^2}{2s^2} \right]$$

$$\operatorname{argmin} \ln(L) = \operatorname{argmin} \sum_{i=1}^{n} (Y_i - \boldsymbol{X_i b})^2. \qquad \beta_{\boldsymbol{ML}} = \beta_{\boldsymbol{OLS}}$$

Can We Believe the CLM?

All models are wrong, but some are useful.

George Box

The CLM Assumptions

1. I.I.D. Data
2. No Perfect Collinearity
3. Linear Conditional Expectation
4. Homoskedastic Errors
5. Normally Distributed Errors

OLS when the CLM is false

The OLS estimator is consistent for the parameter values that minimize KL divergence between the true distribution and the model distribution.

Can We Believe the CLM?

> All models are wrong; they are, at best, approximations of reality. But, even without assuming that they are exactly true, when employed and interpreted correctly, they can nonetheless be useful for obtaining estimates of features of probability distributions.
>
> \- Aronow and Miller

Assessing the CLM assumptions - Software Demo

This one will be a screenshare using R Studio