

# Econometrics Cheat Sheet

by Tyler Ransom, University of Oklahoma  
@tyleransom

## Data & Causality

Basics about data types and causality.

### Types of data

Experimental	Data from randomized experiment
Observational	Data collected passively
Cross-sectional	Multiple units, one point in time
Time series	Single unit, multiple points in time
Longitudinal (or Panel)	Multiple units followed over multiple time periods

### Experimental data

- Correlation  $\implies$  Causality
- Very rare in Social Sciences

## Statistics basics

We examine a **random sample** of data to learn about the population

Random sample	Representative of population
Parameter ( $\theta$ )	Some number describing population
Estimator of $\theta$	Rule assigning value of $\theta$ to sample e.g. Sample average, $\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$
Estimate of $\theta$	What the estimator spits out for a particular sample ( $\hat{\theta}$ )
Sampling distribution	Distribution of estimates across all possible samples
Bias of estimator $W$	$E(W) - \theta$
Efficiency	$W$ efficient if $Var(W) < Var(\tilde{W})$
Consistency	$W$ consistent if $\hat{\theta} \rightarrow \theta$ as $N \rightarrow \infty$

## Hypothesis testing

The way we answer yes/no questions about our population using a sample of data. e.g. “Does increasing public school spending increase student achievement?”

null hypothesis ( $H_0$ )	Typically, $H_0 : \theta = 0$
alt. hypothesis ( $H_a$ )	Typically, $H_0 : \theta \neq 0$
significance level ( $\alpha$ )	Tolerance for making Type I error; (e.g. 10%, 5%, or 1%)
test statistic ( $T$ )	Some function of the sample of data
critical value ( $c$ )	Value of $T$ such that reject $H_0$ if $ T  > c$ ; $c$ depends on $\alpha$ ; $c$ depends on if 1- or 2-sided test
$p$ -value	Largest $\alpha$ at which fail to reject $H_0$ ; reject $H_0$ if $p < \alpha$

## Simple Regression Model

Regression is useful because we can estimate a *ceteris paribus* relationship between some variable  $x$  and our outcome  $y$

$$y = \beta_0 + \beta_1 x + u$$

We want to estimate  $\hat{\beta}_1$ , which gives us the effect of  $x$  on  $y$ .

## OLS formulas

To estimate  $\hat{\beta}_0$  and  $\hat{\beta}_1$ , we make two assumptions:

1.  $E(u) = 0$
2.  $E(u|x) = E(u)$  for all  $x$

When these hold, we get the following formulas:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$
$$\hat{\beta}_1 = \frac{\widehat{Cov}(y, x)}{\widehat{Var}(x)}$$

fitted values ( $\hat{y}_i$ )	$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$
residuals ( $\hat{u}_i$ )	$\hat{u}_i = y_i - \hat{y}_i$
Total Sum of Squares	$SST = \sum_{i=1}^N (y_i - \bar{y})^2$
Expl. Sum of Squares	$SSE = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2$
Resid. Sum of Squares	$SSR = \sum_{i=1}^N \hat{u}_i^2$
R-squared ( $R^2$ )	$R^2 = \frac{SSE}{SST}$ ; “frac. of var. in $y$ explained by $x$ ”

## Algebraic properties of OLS estimates

- $\sum_{i=1}^N \hat{u}_i = 0$  (mean & sum of residuals is zero)
- $\sum_{i=1}^N x_i \hat{u}_i = 0$  (zero covariance bet.  $x$  and resids.)
- The OLS line (SRF) always passes through  $(\bar{x}, \bar{y})$
- $SSE + SSR = SST$
- $0 \leq R^2 \leq 1$

## Interpretation and functional form

- Our model is restricted to be **linear in parameters**
- But not linear in  $x$
- Other functional forms can give more realistic model

Model	DV	RHS	Interpretation of $\beta_1$
Level-level	$y$	$x$	$\Delta y = \beta_1 \Delta x$
Level-log	$y$	$\log(x)$	$\Delta y \approx (\beta_1/100) [1\% \Delta x]$
Log-level	$\log(y)$	$x$	$\% \Delta y \approx (100\beta_1) \Delta x$
Log-log	$\log(y)$	$\log(x)$	$\% \Delta y \approx \beta_1 \% \Delta x$
Quadratic	$y$	$x + x^2$	$\Delta y = (\beta_1 + 2\beta_2 x) \Delta x$

Note: DV = dependent variable; RHS = right hand side

## Multiple Regression Model

Multiple regression is more useful than simple regression because we can more plausibly estimate *ceteris paribus* relationships (i.e.  $E(u|x) = E(u)$  is more plausible)

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u$$

$\hat{\beta}_1, \dots, \hat{\beta}_k$ : **partial effect** of each of the  $x$ 's on  $y$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \dots - \hat{\beta}_k \bar{x}_k$$
$$\hat{\beta}_j = \frac{\widehat{Cov}(y, \text{residualized } x_j)}{\widehat{Var}(\text{residualized } x_j)}$$

where “residualized  $x_j$ ” means the residuals from OLS regression of  $x_j$  on all other  $x$ 's (i.e.  $x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_k$ )

## Gauss-Markov Assumptions

1.  $y$  is a linear function of the  $\beta$ 's
2.  $y$  and  $x$ 's are randomly sampled from population
3. No perfect multicollinearity
4.  $E(u|x_1, \dots, x_k) = E(u) = 0$  (Unconfoundedness)
5.  $Var(u|x_1, \dots, x_k) = Var(u) = \sigma^2$  (Homoskedasticity)

When (1)-(4) hold: OLS is unbiased; i.e.  $E(\hat{\beta}_j) = \beta_j$   
When (1)-(5) hold: OLS is Best Linear Unbiased Estimator

## Variance of $u$ (a.k.a. “error variance”)

$$\hat{\sigma}^2 = \frac{SSR}{N - K - 1}$$
$$= \frac{1}{N - K - 1} \sum_{i=1}^N \hat{u}_i^2$$

## Variance and Standard Error of $\hat{\beta}_j$

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}, j = 1, 2, \dots, k$$

where

$$SST_j = (N - 1)Var(x_j) = \sum_{i=1}^N (x_{ij} - \bar{x}_j)^2$$

$R_j^2 = R^2$  from a regression of  $x_j$  on all other  $x$ 's

Standard deviation:  $\sqrt{Var}$

Standard error:  $\sqrt{\widehat{Var}}$

$$se(\hat{\beta}_j) = \sqrt{\frac{\hat{\sigma}^2}{SST_j(1 - R_j^2)}}, j = 1, \dots, k$$

## Classical Linear Model (CLM)

Add a 6th assumption to Gauss-Markov:

6.  $u$  is distributed  $N(0, \sigma^2)$

Need this to know what the exact *distribution* of  $\hat{\beta}_j$  is

- If A(6) fails, need **asymptotics** to test  $\beta$ 's
- Then, interpret distr. of  $\hat{\beta}_j$  as asymptotic (not exact)

## Testing Hypotheses about the $\beta$ 's

- Under A (1)–(6), can test hypotheses about the  $\beta$ 's
- Or, (much more plausible) A (1)–(5) + asymptotics

### *t*-test for simple hypotheses

To test a simple hypothesis like

$$H_0 : \beta_j = 0$$
$$H_a : \beta_j \neq 0$$

use a *t*-test:

$$t = \frac{\hat{\beta}_j - 0}{se(\hat{\beta}_j)}$$

where 0 is the null hypothesized value.

Reject  $H_0$  if  $p < \alpha$  or if  $|t| > c$  (See: Hypothesis testing)

### *F*-test for joint hypotheses

Can't use a *t*-test for joint hypotheses, e.g.:

$$H_0 : \beta_3 = 0, \beta_4 = 0, \beta_5 = 0$$
$$H_a : \beta_3 \neq 0 \text{ OR } \beta_4 \neq 0 \text{ OR } \beta_5 \neq 0$$

Instead, use *F* statistic:

$$F = \frac{(SSR_r - SSR_{ur}) / (df_r - df_{ur})}{SSR_{ur} / df_{ur}} = \frac{(SSR_r - SSR_{ur}) / q}{SSR_{ur} / (N - k - 1)}$$

where

$$SSR_r = SSR \text{ of restricted model (if } H_0 \text{ true)}$$
$$SSR_{ur} = SSR \text{ of unrestricted model (if } H_0 \text{ false)}$$
$$q = \# \text{ of equalities in } H_0$$
$$N - k - 1 = \text{Deg. Freedom of unrestricted model}$$

Reject  $H_0$  if  $p < \alpha$  or if  $F > c$  (See: Hypothesis testing)

Note:  $F > 0$ , always

### Qualitative data

- Can use qualitative data in our model
- Must create a **dummy variable**
- e.g. “Yes” represented by 1 and “No” by 0

**dummy variable trap:** Perfect collinearity that happens when too many dummy variables are included in the model

$$y = \beta_0 + \beta_1 \textit{happy} + \beta_2 \textit{not.happy} + u$$

The above equation suffers from the dummy variable trap because units can only be “happy” or “not happy,” so including both would result in perfect collinearity with the intercept

### Interpretation of dummy variables

Interpretation of dummy variable coefficients is always relative to the excluded category (e.g. *not.happy*):

$$y = \beta_0 + \beta_1 \textit{happy} + \beta_2 \textit{age} + u$$

$\beta_1$ : avg. *y* for those who are happy *compared to* those who are unhappy, holding fixed age

### Interaction terms

**interaction term:** When two *x*'s are multiplied together

$$y = \beta_0 + \beta_1 \textit{happy} + \beta_2 \textit{age} + \beta_3 \textit{happy} \times \textit{age} + u$$

$\beta_3$ : difference in *age slope* for those who are happy *compared to* those who are unhappy

### Linear Probability Model (LPM)

When *y* is a dummy variable, e.g.

$$\textit{happy} = \beta_0 + \beta_1 \textit{age} + \beta_2 \textit{income} + u$$

$\beta$ 's are interpreted as *change in probability*:

$$\Delta \Pr(y = 1) = \beta_1 \Delta x$$

By definition, homoskedasticity is violated in the LPM

### Time Series (TS) data

- Observe one unit over many time periods
- e.g. US quarterly GDP, 3-month T-bill rate, etc.
- New G-M assumption: no serial correlation in  $u_t$
- Remove random sampling assumption (makes no sense)

### Two focuses of TS data

1. Causality (e.g.  $\uparrow$  taxes  $\xrightarrow{?}$   $\downarrow$  GDP growth)
2. Forecasting (e.g. AAPL stock price next quarter?)

### Requirements for TS data

To properly use TS data for causal inf / forecasting, need data free of the following elements:

Trends:  $y$  always  $\uparrow$  or  $\downarrow$  every period  
Seasonality:  $y$  always  $\uparrow$  or  $\downarrow$  at regular intervals  
Non-stationarity:  $y$  has a unit root; i.e. not stable  
Otherwise,  $R^2$  and  $\hat{\beta}_j$ 's are misleading

### AR(1) and Unit Root Processes

AR(1) model (Auto Regressive of order 1):

$$y_t = \rho y_{t-1} + u_t$$

Stable if  $|\rho| < 1$ ; Unit Root if  $|\rho| \geq 1$   
“Non-stationary,” “Unit Root,” “Integrated” are all synonymous

### Correcting for Non-stationarity

Easiest way is to take a first difference:

First difference: Use  $\Delta y = y_t - y_{t-1}$  instead of  $y_t$   
Test for unit root: Augmented Dickey-Fuller (ADF) test  
 $H_0$  of ADF test:  $y$  has a unit root

### TS Forecasting

A good forecast minimizes **forecasting error**  $\hat{f}_t$ :

$$\min_{f_t} E(e_{t+1}^2 | I_t) = E[(y_{t+1} - f_t)^2 | I_t]$$

where  $I_t$  is the **information set**

RMSE measures forecast performance (on future data):

$$\text{Root Mean Squared Error} = \sqrt{\frac{1}{m} \sum_{h=0}^{m-1} \hat{e}_{T+h+1}^2}$$

Model with lowest RMSE is best forecast

- Can choose  $f_t$  in many ways
- Basic way:  $\hat{y}_{T+1}$  from linear model
- ARIMA, ARMA-GARCH are cutting-edge models

### Granger causality

$z$  **Granger causes**  $y$  if, after controlling for past values of  $y$ , past values of  $z$  help forecast  $y_t$

### CLM violations

#### Heteroskedasticity

- Test: Breusch-Pagan or White tests ( $H_0$  : homosk.)
- If  $H_0$  rejected, SEs, t-, and F-stats are invalid
- Instead use heterosk.-robust SEs and t- and F-stats

#### Serial correlation

- Test: Breusch-Godfrey test ( $H_0$  : no serial corr.)
- If  $H_0$  rejected, SEs, t-, and F-stats are invalid
- Instead use HAC SEs and t- and F-stats
- HAC: “Heterosk. and Autocorrelation Consistent”

#### Measurement error

- Measurement error in  $x$  can be a violation of A4
- **Attenuation bias:**  $\beta_j$  biased towards 0

#### Omitted Variable Bias

When an important  $x$  is excluded: **omitted variable bias**.

Bias depends on two forces:

1. Partial effect of  $x_2$  on  $y$  (i.e.  $\beta_2$ )
2. Correlation between  $x_2$  and  $x_1$

Which direction does the bias go?

	$Corr(x_1, x_2) > 0$	$Corr(x_1, x_2) < 0$
$\beta_2 > 0$	Positive Bias	Negative Bias
$\beta_2 < 0$	Negative Bias	Positive Bias

Note: “Positive bias” means  $\beta_1$  is too big;  
“Negative bias” means  $\beta_1$  is too small

## How to resolve $E(u|\mathbf{x}) \neq 0$

How can we get unbiased  $\hat{\beta}_j$ 's when  $E(u|\mathbf{x}) \neq 0$ ?

- Include lagged  $y$  as a regressor
- Include proxy variables for omitted ones
- Use instrumental variables
- Use a natural experiment (e.g. diff-in-diff)
- Use panel data

## Instrumental Variables (IV)

A variable  $z$ , called the instrument, satisfies:

1.  $cov(z, u) = 0$  (**not** testable)
2.  $cov(z, x) \neq 0$  (testable)

$z$  typically comes from a **natural experiment**

$$\hat{\beta}_{IV} = \frac{cov(z, y)}{cov(z, x)}$$

- SE's much larger when using IV compared to OLS
- Be aware of **weak instruments**

### When there are multiple instruments:

- use Two-stage least squares (2SLS)
- exclude at least as many  $z$ 's as endogenous  $x$ 's

1st stage: regress endogenous  $x$  on  $z$ 's and exogenous  $x$ 's

2nd stage: regress  $y$  on  $\hat{x}$  and exogenous  $x$ 's

**Test for weak instruments:** Instrument is weak if

- 1st stage  $F$  stat  $< 10$
- or 1st stage  $|t| < \sqrt{10} \approx 3.2$

## Difference in Differences (DiD)

Can get causal effects from **pooled cross sectional data**

A nat. experiment divides units into treatment, control groups

$$y_{it} = \beta_0 + \delta_0 d_{2t} + \beta_1 dT_{it} + \delta_1 d_{2t} \times dT_{it} + u_{it}$$

where

- $d_{2t}$  = dummy for being in time period 2
- $dT_{it}$  = dummy for being in the treatment group
- $\hat{\delta}_1$  = **difference in differences**

$$\hat{\delta}_1 = (\bar{y}_{treat,2} - \bar{y}_{control,2}) - (\bar{y}_{treat,1} - \bar{y}_{control,1})$$

### Extensions:

- Can also include  $x$ 's in the model

- Can also use with more than 2 time periods
- $\hat{\delta}_1$  has same interpretation, different math formula

### Validity:

- Need  $y$  changing across time and treatment for reasons only due to the policy
- a.k.a. **parallel trends assumption**

### Panel data

Follow same sample of units over multiple time periods

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \dots + \beta_k x_{itk} + \underbrace{a_i + u_{it}}_{\nu_{it}}$$

- $\nu_{it}$  = **composite error**
- $a_i$  = unit-specific unobservables
- $u_{it}$  = idiosyncratic error
- Allow  $E(a|\mathbf{x}) \neq 0$
- Maintain  $E(u|\mathbf{x}) = 0$

Four different methods of estimating  $\beta_j$ 's:

1. Pooled OLS (i.e. ignore composite error)
2. First differences (FD):

$$\Delta y_i = \beta_1 \Delta x_{i1} + \dots + \beta_k \Delta x_{ik} + \Delta u_i$$

estimated via Pooled OLS on transformed data

3. Fixed effects (FE):

$$y_{it} - \bar{y}_i = \beta_1 (x_{it1} - \bar{x}_{i1}) + \dots + \beta_k (x_{itk} - \bar{x}_{ik}) + (u_{it} - \bar{u}_i)$$

estimated via Pooled OLS on transformed data

4. Random effects (RE):

$$y_{it} - \theta \bar{y}_i = \beta_0 (1 - \theta) + \beta_1 (x_{it1} - \theta \bar{x}_{i1}) + \dots + \beta_k (x_{itk} - \theta \bar{x}_{ik}) + (\nu_{it} - \theta \bar{\nu}_i)$$

estimated via FGLS, where

$$\theta = 1 - \sqrt{\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_a^2}}$$

$$\hat{\beta}_{RE} \rightarrow \hat{\beta}_{FE} \text{ as } \theta \rightarrow 1$$

$$\hat{\beta}_{RE} \rightarrow \hat{\beta}_{POLS} \text{ as } \theta \rightarrow 0$$

**RE assumes**  $E(a|\mathbf{x}) = 0$

### Cluster-robust SEs

- Serial correlation of  $\nu_{it}$  is a problem
- Use **cluster-robust** SEs
- These correct for serial corr. and heterosk.
- Cluster at the unit level

## Binary dependent variables

Three options for estimation when  $y$  is binary (0/1):

- Linear Probability Model
- Logit
- Probit

Latter two are *nonlinear* models:

$$P(y = 1 | \mathbf{x}) = G(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$$

where  $G(\cdot)$  is some nonlinear function satisfying  $0 < G(\cdot) < 1$

## Trade-offs with logit/probit

### Disadvantages

- Now it's much harder to estimate and interpret  $\beta$ 's!
- Can no longer use OLS; instead use maximum likelihood
- Nonlinear  $G(\cdot) \implies$ 
  - Must use chain rule to compute slope
  - Slope of tangent line depends on  $\mathbf{x}$ !

### Main advantage

- Now  $0 < \hat{y} < 1$ , which is more realistic
- (Recall: in LPM, possible to have negative probabilities)

## Common choices for $G(\cdot)$

### Logit model:

$$G(\mathbf{x}\beta) = \frac{\exp(\mathbf{x}\beta)}{1 + \exp(\mathbf{x}\beta)} = \Lambda(\mathbf{x}\beta)$$

### Probit model:

$$G(\mathbf{x}\beta) = \int_{-\infty}^{\mathbf{x}\beta} \phi(z) dz = \Phi(\mathbf{x}\beta)$$

where  $\phi(\cdot)$  is the standard normal pdf

## Interpreting logit/probit parameter estimates

- $\beta$ 's that come from logit/probit  $\neq \beta$ 's from LPM
- But, *sign* is same
- In logit/probit, we have

$$\frac{\partial p(\mathbf{x})}{\partial x_j} = \beta_j g(\mathbf{x}\beta)$$

where  $g(\mathbf{x}\beta)$  is the first derivative of  $G(\mathbf{x}\beta)$

- In LPM, we have  $\frac{\partial p(\mathbf{x})}{\partial x_j} = \beta_j$