Econometrics Final Concepts

Ryan Safner

Fall 2017

Ordinary Least Squares (OLS) Regression

- Bivariate data and associations between variables (e.g. X and Y)
  - Apparent relationships are best viewed by looking at a scatterplot
    - Check for associations to be positive/negative, weak/strong, linear/nonlinear, etc
    - Y: dependent variable
    - X: independent variable
  - Correlation coefficient (r) can quantify the strength of an association

$$r = \frac{1}{n-1} \sum^{n} \left(\frac{X_i - \bar{X}}{s_X} \right) \left(\frac{Y_i - \bar{Y}}{s_Y} \right) = \frac{\sum^{n} Z_X Z_Y}{n-1}$$

    - $-1 \leq r \leq 1$ and r only measures *linear* associations
    - $|r|$ closer to 1 imply stronger correlation (near a perfect straight line)
    - Correlation does not imply causation! Might be confounding or lurking variables (e.g. Z) affecting X and/or Y

- Population regression model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

  - β_1: $\frac{\Delta Y}{\Delta X}$: the slope between X and Y, number of units of Y from a 1 unit change in X
  - β_0 is the Y-intercept: literally, value of Y when $X = 0$
  - ϵ_i is the error or residual, difference between actual value of $Y|X$ vs. predicted value of $\hat{Y}$

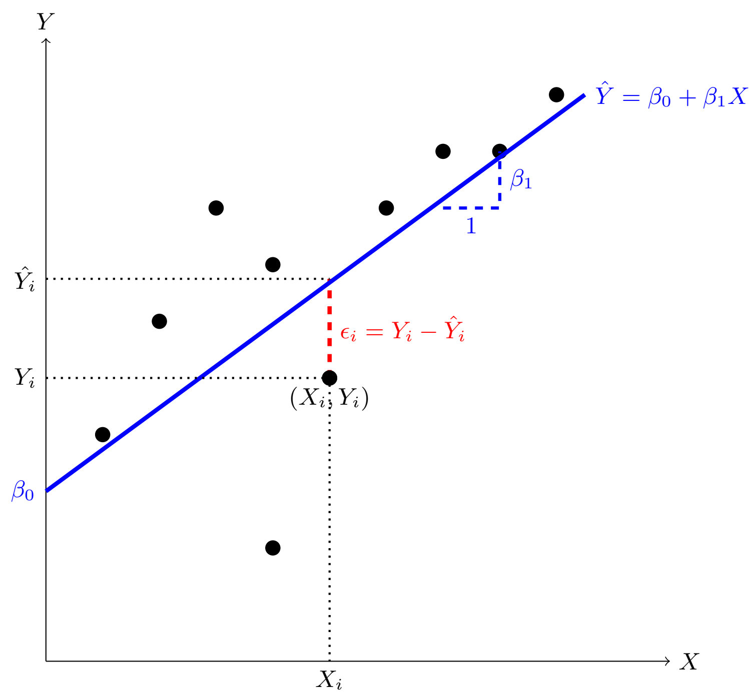
- Ordinary Least Squares (OLS) regression model

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

  - Least square estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ estimate population regression line from sample data
  - Minimize sum of squared errors (SSE) $min \sum^{n} \epsilon_i^2$ where $\epsilon_i = Y_i - \hat{Y}_i$
  - OLS regression line

$$\hat{\beta}_1 = \frac{cov(X,Y)}{var(X)} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = r_{X,Y} \frac{s_Y}{s_X}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$



Y
$\hat{Y} = \beta_0 + \beta_1 X$
β_1
1
$\hat{Y}_i$
$\epsilon_i = Y_i - \hat{Y}_i$
Y_i
(X_i, Y_i)
β_0
X
X_i

- Measures of Fit
  - R^2: fraction of total variation on Y explained by variation in X according to model

$$R^2 = \frac{ESS}{TSS}$$

$$R^2 = 1 - \frac{SSE}{TSS}$$

$$R^2 = r^2_{X,Y}$$

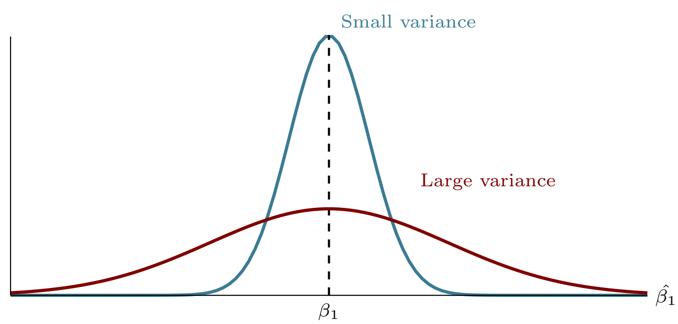
    * $ESS = \sum(\hat{Y}_i - \bar{Y})^2$
    * $TSS = \sum(Y_i - \bar{Y})^2$
    * $SSE = \sum \hat{\epsilon_i}^2$
  - Standard error of the regression (SER): average size of ϵ_i, average distance from regression line to data points

$$SER = \frac{1}{n-2}\sum \hat{\epsilon_i}^2 = \frac{SSE}{n-2}$$

- Hypothesis testing of β_1
  - $H_0 : \beta_1 = \beta_{1,0}$, often $H_0 : \beta_1 = 0$
  - Two sided alternative $H_1 : \beta_1 \neq 0$
  - One sided alternatives $H_1 : \beta_1 > 0$, $H_2 : \beta_1 < 0$
  - t-statistic

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE[\hat{\beta}_1]}$$

  - Compare t against critical value t^*, or compute p-value as usual
  - Confidence intervals (95%): $\hat{\beta}_1 \pm 1.96(SE[\hat{\beta}_1])$



$\hat{\beta}_1$ is a random variable, so it has its own sampling distribution with mean $E[\hat{\beta}_1]$ and standard error $se[\hat{\beta}_1]$

- Mean of OLS estimator $\hat{\beta}_1$ & Bias: Endogeneity & Exogeneity
  - X is **exogenous** if it is not correlated with the error term

$$corr(X, \epsilon) = 0$$

    * Equivalently, knowing X should not give you any information about ϵ:

$$E[\epsilon|X] = 0$$

      * If X is exogenous, OLS estimate on X is unbiased:

$$E[\hat{\beta}_1] = \beta_1$$

  - X is **endogenous** if it is correlated with the error term

$$corr(X, \epsilon) \neq 0$$

      * Equivalently, knowing X gives you information about ϵ:

$$E[\epsilon|X] \neq 0$$

      * If X is endogenous, OLS estimate on X is biased:

$$E[\hat{\beta}_1] = \beta_1 + corr(X, \epsilon)\frac{\sigma_\epsilon}{\sigma_X}$$

        · Can measure strength and direction (+ or −) of bias
        · Note: if unbiased, $corr(X, \epsilon) = 0$, so $E[\hat{\beta}_1] = \beta_1$

- Variance of OLS estimator $\hat{\beta}_1$, measuring precision of estimate

$$var[\hat{\beta}_1] = \frac{\hat{\sigma}^2}{n \times var(X)}$$

  and standard error

$$se[\hat{\beta}_1] = \sqrt{\frac{\hat{\sigma}^2}{n \times var(X)}}$$

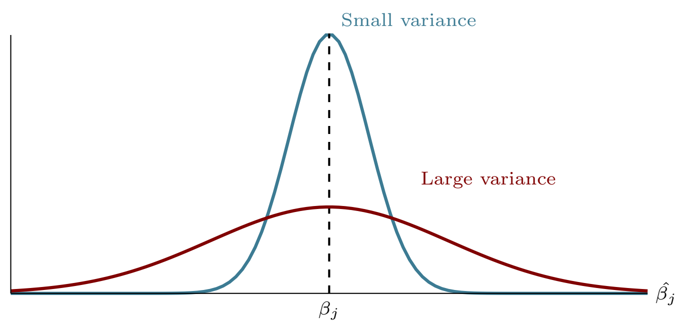
  - Affected by 3 major factors:
    1. Model fit, where SER=$\hat{\sigma}$
    2. Sample size n
    3. Variation in X_j

- Heteroskedasticity and homoskedasticity

  - Homoskedastic errors (ϵ) have the same variance over all values of X
  - Heteroskedastic errors (ϵ) have different variance over values of X
      * Heteroskedasticity does *not* bias our estimates, but incorrectly lowers variance & standard errors (inflating t-statistics and significance!)
      * Can correct for heteroskedasticity by using robust standard errors

Multivariate Regression

- Omitted Variable Bias
  - A variable Z causes omitted variable bias if:
    1. $corr(X, Z) \neq 0$, X and Z are correlated
    2. $corr(Z, Y) \neq 0$, Z is in the error term that explains Y
  - Omitted variable bias can be avoided by including Z in the regression (as X_2)
- Multivariate Regression Model
$$\widehat{Y_i} = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{\epsilon}_i$$
  - $\hat{\beta}_0$: predicted value of $\hat{Y}_i$ when $X_{1i} = 0; X_{2i} = 0$
  - $\hat{\beta}_1 = \dfrac{\Delta Y_i}{\Delta X_{1i}}$, marginal effect of X_{1i} on Y_i, holding X_{2i} constant
  - $\hat{\beta}_2 = \dfrac{\Delta Y_i}{\Delta X_{2i}}$, marginal effect of X_{2i} on Y_i, holding X_{1i} constant
- Measuring Omitted Variable Bias
  - Suppose we omit X_{2i} and run an Omitted Regression
$$Y_i = \alpha_0 + \alpha_1 X_{1i} + \nu_i$$
  - If we run an Auxiliary Regression of X_{2i} on X_{1i}:
$$X_{2i} = \delta_0 + \delta_1 X_{1i} + \tau_i$$
    * Size and significance of δ_1 measures relationship between X_{1i} and X_{2i}
$$\alpha_1 = \beta_1 + \beta_2 \delta_1$$
  - Biased estimate α_1 in Omitted Regression picks up:
    * True effect of X_{1i} on Y_i (β_1)
    * Effect of X_{2i} on Y_i (β_2) as pulled through the relationship between X_{1i} and X_{2i} (δ_1)
  - Conditions for Z being an omitted variable
    * Z_i must be a determinant of Y_i ($\beta_2 \neq 0$)
    * Z_i is correlated with X_{1i} ($\delta_1 \neq 0$)



$\hat{\beta}_j$ is a random variable, so it has its own sampling distribution with mean $E[\hat{\beta}_j]$ and standard error $se[\hat{\beta}_j]$

- Variance of OLS estimators $\hat{\beta}_j$

$$var[\hat{\beta}_j] = \frac{1}{(1-R_j^2)} * \frac{\hat{\sigma}^2}{n \times var[X_j]}$$

  and Standard error

$$s.e.[\hat{\beta}_j] = \sqrt{var[\hat{\beta}_j]}$$

- Affected by 4 major factors:
  1. Model fit, where SER=$\hat{\sigma}$
  2. Sample size n
  3. Variation in X_j
  4. Variance Inflation Factor (VIF) $\frac{1}{1-R_j^2}$
     - Independent variables are **multicollinear** if they are correlated

$$corr(X_j, X_l) \neq 0 \text{ for } j \neq l$$

     - Does not bias estimators, but increases their variance & standard errors
     - R_j^2 is the R^2 from an auxiliary regression of X_j on all other regressors
     - VIF quantifies how by many times the variance of $\hat{\beta}_j$ increased because of multicollinearity
       * $VIF > 10$ (or $\frac{1}{VIF} > 0.10$) is bad
     - **Perfect multicollinearity** when a regressor is an exact linear function of (an)other regressor(s) – cannot run a regression, a logical impossibility

$$|corr(X_1, X_2)| = 1$$

Dummy Variables

- Dummy variable
$$D_i = \begin{cases} 1 & \text{if } i \text{ meets condition} \\ 0 & \text{if } i \text{ does not meet condition} \end{cases}$$
- Dummy variables measure group means
$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 D_i$$
  - When $D_i = 0$ (Control group):
    - $\hat{Y}_i = \hat{\beta}_0$
    - $E[Y|D_i = 0] = \hat{\beta}_0 \iff$ the mean of Y when $D_i = 0$
  - When $D_i = 1$ (Treatment group):
    - $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 D_i$
    - $E[Y|D_i = 1] = \hat{\beta}_0 + \hat{\beta}_1 \iff$ the mean of Y when $D_i = 1$
  - Difference in group means:
$$\begin{aligned} &= E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \\ &= (\hat{\beta}_0 + \hat{\beta}_1) - (\hat{\beta}_0) \\ &= \hat{\beta}_1 \end{aligned}$$

- Transforming categorical variables into dummies
  - A categorical variable (e.g. region, class standing, etc) can be added to a regression by making each category option a dummy variable and including them all (minus one)
$$Y_i = \beta_0 + \beta_2 D_1 + \beta_2 D_2 + \beta_3 D_3$$
    where observations can fall into category 1, 2, 3, or 4
  - Including all category option dummies into a regression yields the **dummy variable trap**, where all dummies are perfectly multicollinear
  - Must drop one category dummy, the "reference group"
  - Coefficients on dummy variables are the difference between that category and the reference category:
    - $\beta_0 = Y$ for category 4 (omitted)
    - $\beta_1 =$ difference between category 1 and category 4 (omitted)
    - $\beta_2 =$ difference between category 2 and category 4 (omitted)
    - $\beta_3 =$ difference between category 3 and category 4 (omitted)
- **Interaction terms** measure if there is an additional effect of one variable on the value of another, 3 combinations:

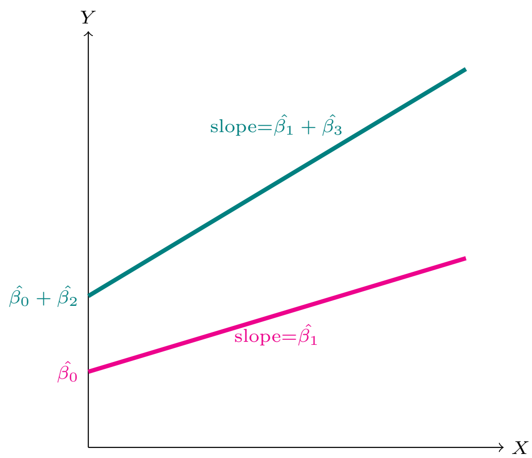
  1. Between a dummy and a continuous variable
$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 \mathbf{X_i} \times \mathbf{D_i}$$
     - Coefficients:
       - β_0: Y_i for $X_i = 0$ and $D_i = 0$
       - β_1: Effect of $X_i \to Y_i$ for $D_i = 0$
       - β_2: Effect on Y_i of difference between $D_i = 0$ and $D_i = 1$

    * β_3: Effect of *difference* of $X_i \to Y_i$ between $D_i = 0$ and $D_i = 1$
  - Easier to see as two different regression lines:
    * When $D_i = 0$ (Control group):

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

    * When $D_i = 1$ (Treatment group):

$$\hat{Y}_i = (\hat{\beta}_0 + \hat{\beta}_2) + (\hat{\beta}_1 + \hat{\beta}_3) X_i$$



    * Two regression lines may have (same/different) intercepts and (same/different) intercepts, test significance of:
      · β_2: difference in intercepts
      · β_3: difference in slopes

2. Between two dummy variables

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 \mathbf{D_{1i}} \times \mathbf{D_{2i}}$$

  - Coefficients:
    * β_0: value of Y for $D_{1i} = 0$ and $D_{2i} = 0$
    * β_1: effect on Y of $D_{1i} = 0 \to 1$ when $D_{2i} = 0$
    * β_2: effect on Y of $D_{2i} = 0 \to 1$ when $D_{1i} = 0$
    * β_3: *increment* to effect on Y of $D_{1i} = 0 \to 1$ when $D_{2i} = 1$ vs. when $D_{2i} = 0$
  - Compare difference in group means:
    * $D_{1i} = 0, D_{2i} = 0$: $\widehat{Y_i} = \hat{\beta}_0$
    * $D_{1i} = 0, D_{2i} = 1$: $\widehat{Y_i} = \hat{\beta}_0 + \hat{\beta}_2$
    * $D_{1i} = 1, D_{2i} = 0$: $\widehat{Y_i} = \hat{\beta}_0 + \hat{\beta}_1$
    * $D_{1i} = 1, D_{2i} = 1$: $\widehat{Y_i} = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3$

3. Between two continuous variables

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (\mathbf{X_{1i}} \times \mathbf{X_{2i}})$$

  - Marginal effects:
    * $\dfrac{\Delta Y_i}{\Delta X_{1i}} = \beta_1 + \beta_3 X_{2i}$ — marginal effect of $X_{1i} \to Y_i$ depends on X_{2i}
    * $\dfrac{\Delta Y_i}{\Delta X_{2i}} = \beta_2 + \beta_3 X_{1i}$ — marginal effect of $X_{2i} \to Y_i$ depends on X_{1i}

Transforming Variables

- Polynomial functions
$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2 + ... + \hat{\beta}_r X_i^r + \epsilon_i$$
where r is highest power X_i is raised to, a function with $r-1$ bends
  - Quadratic model
$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2 + \epsilon_i$$
    * Marginal effect of $X_i \to Y_i$:
$$\frac{d\,Y_i}{d\,X_i} = \hat{\beta}_1 + 2\hat{\beta}_2 X_i$$
    * Value of X_i where Y_i is minimized/maximized:
$$X_i^* = -\frac{1}{2}\frac{\beta_1}{\beta_2}$$
  - To determine if a higher-powered term is necessary, test significance of its associated coefficient (e.g. β_2 for quadratic model above)
  - To determine if a model is nonlinear, run F-test of all higher-powered terms
- Logarithmic functions (ln)
  - Natural Logs (ln) are used to talk about percentage changes, 3 types of models:
    1. Linear-log model:
$$Y = \beta_0 + \beta_1 \mathbf{ln(X)}$$
       * β_1: A 1% change in $X \to \frac{\beta_1}{100}$ unit change in Y
    2. Log-linear model:
$$\mathbf{ln(Y)} = \beta_0 + \beta_1 X$$
       * β_1: A 1 unit change in $X \to 100 \times \beta_1\%$ change in Y
    3. Log-log model:
$$\mathbf{ln(Y)} = \beta_0 + \beta_1 \mathbf{ln(X)}$$
       * β_1: A 1% change in $X \to \beta_1\%$ change in Y (**elasticity** between X and Y)
- Standardized coefficients
$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$$
  - To compare the magnitude of marginal effects (e.g. is $\beta_1 > \beta_2$) across variables of different units, standardize the variables by taking the Z-score of all observations
$$Variable_{std} = \frac{Variable - \overline{Variable}}{sd(Variable)}$$
  - Coefficients measure the # of standard deviations change of Y a 1 std. dev. change in X causes
- Joint Hypothesis Testing
  - Joint hypothesis tests against the null hypothesis of a value for multiple parameters, e.g.
$$H_0{:}\beta_1 = 0, \beta_2 = 0$$
$$H_1{:}H_0 \text{ is false}$$
  - Three common tests
    1. H_0: $\beta_1 = \beta_2 = 0$, testing if multiple variables do not affect Y

2. H_0: $\beta_1 = \beta_2$, testing if multiple variables have the same effect (must be same units)
3. H_0: all β's$= 0$, the model overall explains no variation in Y

– In general, with q restrictions:

$$H_0 : \beta_j = \beta_{j,0}, \beta_k = \beta_{k,0}, \text{...for } q \text{ restrictions}$$

– Use the F-statistic, (simplified homoskedastic formula below)

$$F_{q,n-k-1} = \frac{\left(\frac{(R_u^2 - R_r^2)}{q}\right)}{\left(\frac{(1-R_u^2)}{(n-k-1)}\right)}$$

– Compares the R^2's of two models:
  * Unrestricted model: regression with all coefficients
  * Restricted model: regression under the null hypothesis (e.g. where $\beta_1 = 0, \beta_2 = 0$)

– F tests if the increase in R^2 from including the suspect variables ($Restricted \rightarrow Unrestricted$) increases by a statistically significant amount

Time Series (Not on Exam)

- Time series data tracks (multiple variables describing) a single individual (person, country, city, etc) over time
- OLS Regression model:
$$Y_t = \hat{\beta_0} + \hat{\beta_1} X_t + \epsilon_t$$
where t represents an observation at time period t
- Lagged variable: 1^{st} lag of Y_t is the value of Y_{t-1}
  - Can take higher-order lags (e.g. 2^{nd} lag of Y is the value of Y_{t-2}, etc.
- Difference: 1^{st} difference of Y:
$$\Delta Y = Y_t - Y_{t-1}$$
using logs to express percentage change:
$$\Delta ln(Y_t) = ln(Y_t) - ln(Y_{t-1}) \times 100\%$$
  - Can take higher-order differences (e.g. 2^{nd} difference of $Y = Y_t - Y_{t-2}$
- **Autocorrelation** or **serial correlation** where error term ϵ_t depends on previous values of ϵ
$$corr(\epsilon_t, \epsilon_{t-1}) \neq 0$$
- First Order Autoregressive Model (AR1) models the error as a linear function of its 1^{st} lag:
$$\epsilon_t = \rho\epsilon_{t-1} + \nu_t$$
  - $-1 < \rho < 1$ ("rho") is the strength (and direction) of autocorrelation
    * If ρ is positive, positive autocorrelation: ϵ_t (residuals) tend to be the same sign ($+$ or $-$) as the ones before it (though signs can switch)
      · A residual plot will be fairly smooth and "sticky", with long periods of positive, and long periods of negative residuals
    * If ρ is negative, negative autocorrelation: ϵ_t (residuals) tend switch signs ($+$ or $-$) between consecutive periods
      · A residual plot will be very "spiky," constantly switching between positive and negative residuals over time
    * If $\rho \approx 0$, no autocorrelation: ϵ_t not significantly affected by ϵ_{t-1}
      · A residual plot is random with no obvious trend of residuals switching or sticking signs
  - ϵ_{t-1} is the (1^{st}) lagged error
  - ν_t is a random error term with mean $E[\nu] = 0$ and variance σ_v^2
  - Can run an auxiliary regression of residuals against lagged residuals and test significance of ρ to detect autocorrelation:
$$\hat{\epsilon_t} = \rho\widehat{\epsilon_{t-1}} + \nu_t$$
- Prais-Winston/Cochrane-Orcutt method of ρ-differencing data to remove autocorrelation:
$$Y_t - \rho Y_{t-1} = \beta_0(1-\rho) + \beta_1(X_t - \rho X_{t-1}) + \nu_t$$
$$\tilde{Y_t} = \tilde{\beta_0} + \beta_1\tilde{X_t} + \nu_t$$
  - Generates a Durbin-Watson (DW) statistic:
$$DW \approx 2(1-\rho)$$

    * DW near 2: no autocorrelation (autocorrelation fixed)
    * DW near 0: positive autocorrelation (needs more fixes)
    * DW near 4: negative autocorrelation (needs more fixes)

- Dynamic model includes lagged dependent variable in regression

$$Y_t = \gamma Y_{t-1} + \beta_0 + \beta_1 X_t + \epsilon_t$$

  - Marginal effect of $X_1 \to Y$:
    * Short-term: β_1 (standard OLS)
    * Long-term: a 1 unit change in X will change Y by $\frac{\beta_1}{1-\gamma}$
  - Often soaks up autocorrelation by including Y_{t-1}

- Trends and Stationarity
  - A variable is stationary if it has the same distribution over time (good)
  - Non-stationarity (bad)
    * Trend: persistent long-term movement or tendency in data, a nonrandom function of time
    * Random walk: random trend over time, where the best prediction of Y_t is Y_{t-1}, ϵ_t is a random error

$$Y_t = Y_{t-1} + \epsilon_t$$

  - Determine stationarity by examining γ in dynamic model (again:)

$$Y_t = \gamma Y_{t-1} + \epsilon_t$$

    * $\gamma < 1$: standard dynamic model (OLS is fine)
    * $\gamma > 1$: data "explodes"
    * $\gamma = 1$: "unit root," Y follows a random walk (our model is useless)
  - Dickey-Fuller test: tests against H_0: $\gamma = 1$ to check for nonstationarity in each variable

Panel Data

- Panel data tracks the same individuals (a cross-section) over time (time-series)

$$\widehat{Y_{it}} = \beta_0 + \beta_1 X_{it} + \epsilon_{it}$$

  with N number of i groups and T number of t time periods

- A **pooled model** simply runs this as normal OLS regression
  - Biased: ignores factors correlated with X in ϵ
  - Systematic differences across groups i that may be stable over time
  - Systematic differences across time t that may be stable across groups

- (One-Way) Fixed effects model

$$\widehat{Y_{it}} = \beta_0 + \beta_1 X_{it} + \alpha_{\mathbf{i}} + \nu_{it}$$

  - α_i: group-fixed effect (pulled from error term ϵ_{it})
    * Includes *all* differences across groups that do not change over time! (e.g. geography, culture, etc. of Maryland vs. Alaska)
    * Does *not* include variables that change over time!
    * Estimates a different intercept for each group
  - Least Squares Dummy Variable (LSDV) Approach: can estimate via creating & including a dummy variable for each group (minus 1 to avoid dummy variable trap)

$$\widehat{Y_{it}} = \beta_0 + \beta_1 X_{it} + \sum_{i=1}^{N-1} \alpha_i D_i$$

    where α_i is a coefficient and D_i is a dummy variable for group i, for example:

$$\widehat{Y_{it}} = \beta_0 + \beta_1 X_{it} + \beta_2 Alabama_i + \beta_3 Alaska_i + ...$$

- Two-Way Fixed effects model

$$\widehat{Y_{it}} = \beta_0 + \beta_1 X_{it} + \alpha_{\mathbf{i}} + \tau_{\mathbf{t}} + \nu_{it}$$

  - τ_i: time-fixed effect (pulled from error term ϵ_{it})
    * Includes *all* differences over time that do not change across groups! (e.g. all States experience recession in 2008, or federal law change)
    * Does *not* include variables that are different across groups!
    * Estimates a different intercept for each time period
  - Least Squares Dummy Variable (LSDV) Approach: can estimate via creating & including a dummy variable for each group and each time period (minus 1 for each to avoid dummy variable trap)

$$\widehat{Y_{it}} = \beta_0 + \beta_1 X_{it} + \sum_{i=1}^{N-1} \alpha_i D_i + \sum_{t=1}^{T-1} \tau_i D_t$$

    where α_i and τ_t are coefficients, D_i is a dummy variable for group i, and D_t is a dummy variable for time period t, for example:

$$\widehat{Y_{it}} = \beta_0 + \beta_1 X_{it} + \beta_2 Alabama_i + \beta_3 Alaska_i + ... + \beta_{51} 2000_t + \beta_{52} 2001_t + ...$$

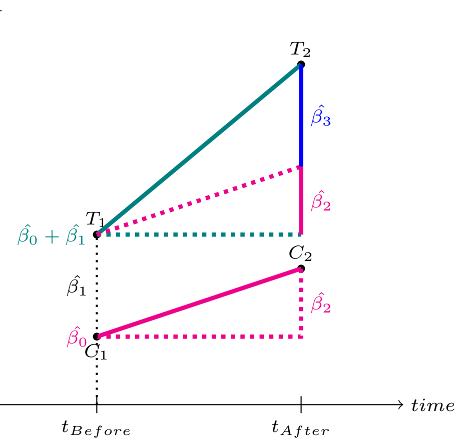
- Difference-in-Differences model

$$\widehat{Y_{it}} = \beta_0 + \beta_1 \text{Treated}_i + \beta_2 \text{After}_{it} + \beta_3 (\text{Treated}_i \times \text{After}_t) + \epsilon_{it}$$

  – Where:

    * $\text{Treated}_i = 1$ if unit i is in treatment group
    * $\text{After}_{it} = 1$ if observation it is after treatment period

	Control	Treatment	Group Diff. (ΔY_i)
Before	β_0	$\beta_0 + \beta_1$	β_1
After	$\beta_0 + \beta_2$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_1 + \beta_3$
Time Diff. (ΔY_t)	β_2	$\beta_2 + \beta_3$	β_3
			Diff-in-diff ($\Delta\Delta Y$)



$$\Delta\Delta Y = (Treated_{after} - Treated_{before}) - (Control_{after} - Control_{before})$$

  – OLS Coefficients:

    * $\hat{\beta}_0$: value of Y for control before treatment
    * $\hat{\beta}_1$: difference between treatment and control (before treatment)
    * $\hat{\beta}_2$: time difference between before and after treatment
    * $\hat{\beta}_3$: difference-in-difference: effect of treatment

  – Values of Y for different groups:

    * Y for Control Group Before: $\hat{\beta}_0$
    * Y for Control Group After: $\hat{\beta}_0 + \hat{\beta}_2$
    * Y for Treatment Group Before: $\hat{\beta}_0 + \hat{\beta}_1$
    * Y for Treatment Group After: $\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3$
    * Treatment Effect: $\hat{\beta}_3$

  – Key assumption about *counterfactual*: if not for treatment, the treated group would change the same over time as the control group (parallel time trends, magenta dotted line)

  – Can generalize the model with two way fixed effects:

$$\widehat{Y_{it}} = \alpha_i + \tau_t + \beta_3 (\text{Treated}_i \times \text{After}_t) + X_{it} + \epsilon_{it}$$

* α_i: group-fixed effects, where some groups receive treatment and others do not
* τ_t: time-fixed effects, where some periods are before treatment and others are after
* X_{it}: other control variables
* This allows for multiple treatments to happen at different times!

Instrumental Variables

- A variable IV can act as an instrumental variable if:

    1. Inclusion Condition: IV statistically significantly explains X

    $$corr(X, IV) \neq 0$$

    2. Exclusion Condition: IV doesn't directly affect Y (not in ϵ)

    $$corr(Y, IV) = 0$$

- IV only affects Y *through* its relationship with X
- IV removes endogenous variation of X (correlated with ϵ) and only uses *exogenous* variation of X (correlated with IV but not with ϵ) to estimate causal effect of $X \rightarrow Y$
- Implement in regression with Two Stage Least Squares (2SLS)

    1. First Stage (Auxiliary Regression)

    $$\widehat{X_{1i}} = \widehat{\gamma}_0 + \widehat{\gamma}_1 IV_i + \widehat{\gamma}_2 X_{2i} + \widehat{\nu}_i$$

        - Use instrument and other control variables (e.g. X_2) to predict value of X_{1i}
        - Can test the inclusion restriction by testing significance of γ_1 ($|t$-statistic$| > 3$)

    2. Second Stage

    $$\widehat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 \widehat{X_{1i}} + \hat{\beta}_2 X_{2i} + \widehat{\epsilon}_i$$

        - Use the predicted value of $\widehat{X_{1i}}$ from First Stage
        - There is no statistical test for the exclusion restriction, must argue why IV does not affect Y (except through X)