ETC2410

Fin

Semester 1 2018

Week 1

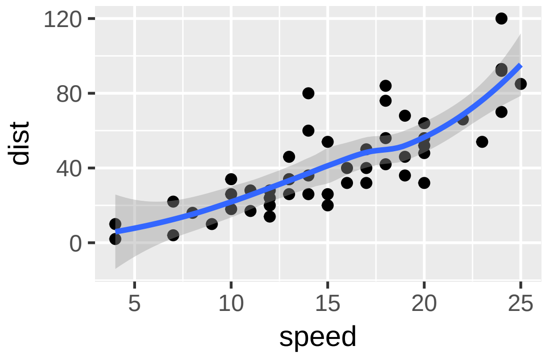
- Stages of empirical analysis

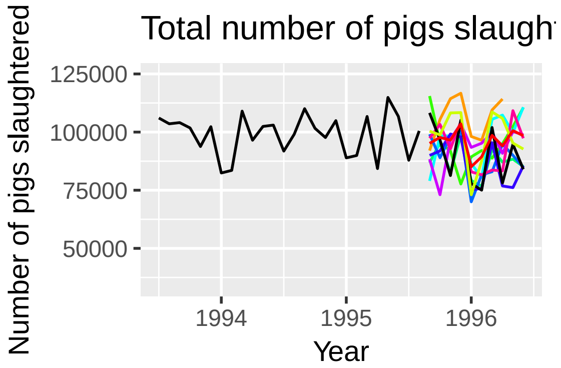
1. Understanding the problem
2. Formulating an appropriate conceptual model to tackle the problem
3. Collecting appropriate data
4. Lookong at the data (Descriptive Analytics)
5. Estimating the model, making inference, predictions and policy presciptions as appropriate
6. Evaluation, learning and improving each of the previous steps, and iterating until the problem is solved

- A general statistical model can be written as:

$$y = f(x) + e$$

Where y denotes the random variablewe want to predict or analyze; x is the variable that are related or can explain some behaviour of y; random error e accounts for the many factors that affect y that we have omitted from this simple model



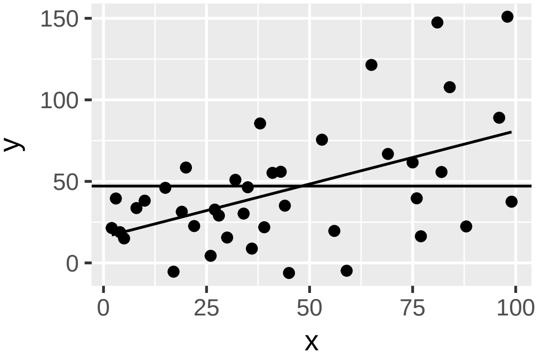


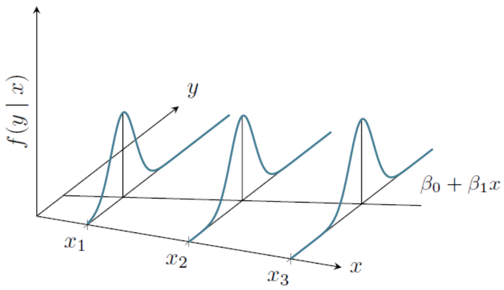
- Converting an economic model to a statistical model
- conomic theory describes the systematic part - $f(i)$
- e is the nonsystematic, random error component that we know is present, but cannot be observed

- Adding random errors converts our economic model into a statistical one that gives us a basis for statistical inference
- Simple Linear Regression
  A complete model for a random variable is a model of its probability distribution. We model the conditional distribution of y given x, a model for $E(y|x)$.

y	x
Dependent Variable	*Explained Variable*
Response Variable	*Predicted Variable*
Independent Variable	*Explanatory Variable*
Control Variable	*Predictor Variable*
Regressand	*Regressor*

- Population and Sample format
  - Population
    - $E(y|x) = \beta_0 + \beta_1 x$
  - Sample Estimation
    - $\hat{y} = \widehat{E(y|x)} = \hat{\beta}_0 + \hat{\beta}_1 x$
- **Law of Probability**
  If A and B are mutually exclusice events then $P(A\ or\ B) = P(A) + P(B)$
  If A and B are two events, then $P(A|B) = \frac{P(A\ and\ B)}{P(B)}$
- ***Model***





$E(y|x) = \beta_0 + \beta_1 x$
$y = \beta_0 + \beta_1 x + u$ where u is a random variable with $E(u) = 0$ and $E(u|x) = 0$

- Experimental and Non-experimental data

Non-experimental data are not accumulated through controlled experiments on individuals, firms, or segments of the economy.

(observational data, retrospective data)

Experimental data are often collected in laboratory environments in the natural sciences, but they are much more difficult to obtain in the social sciences.

- Elasticity

The percentage change in one variable given a 1% ceteris paribus increase in another variable.

$$\frac{\Delta Q/Q}{\Delta P/P}$$

- Ceteris paribus (other relevent factors being equal)
- Predictive analytics (no causality) and Prescriptive analytics (causal)

Predictive analytics: using some variables to predict a target variable without any requirement of causality

Prescriptive analytics to measure the causal relationship between variables, to prescribe how to achieve a desired change in the target variable by manipulating the cause.

We can always use multiple regression for prediction.
We can sometimes use multiple regression to tease out causal relationships

- Cross-sectional, Time-series, Panel or Longitudinal data, Pooled

A **Cross-sectional** data set consists of a sample of individuals, households, firms, cities, states, contries, or a variety of other units, *taken at a given point in time.*

A **Time Series** data set consists of observations on a variable or several variables *over time.*

A **Pooled** data set combine data (on multiple variable) from *different individuals* over time.

A **Panel or Longitudinal** data set consists of a time series for *each* cross-sectional member in the data set.

Time series data can be used to accomplish two important tasks for which cross-sectional data are inadequate. These are to:

1. Forecast future values of a variable:
   eg. stock prices, consumer price index, gross domestic product, annual homicide rates one or several days /months /quarters / years ahead.
2. Estimate the dynamic causal effect of one variable x on another variable y:
   eg. estimate the effect on alcohol consumption of an increase in the tax on alcohol, both initially and subsequently as consumers adjust to the new tax.

- **Cross-sectional data, time series data, panel data**
  Panel data: can be used to address questions that coonot be adequately addressed using either cross-section or time series data.
- **Univariate time series**: A time series data set consisting of observations on a single variable.
- **Multivariate time series**: A time series data set consisting of observations on several variables.
- Notation
  y_t : the value of the time series in time period t.
  T : sample size
- **What time series data can do, but cross-section cannot:**
- Forecast future values of a variable

- Estimate the dynamic causal effect of one variable x (estimate the causal effect on y, over several time periods, of a change in x today. e.g. tax on alcohol)
- **Properties**

1. Observations on time series data are ordered.
2. Time series data is generally charaacterized by some form of temporal dependence.

Because of temporal dependence, it is implaisible to assume that the random variable y_t and y_{t-1} are i.i.d. The strength of temporal dependence can differ between time series. When the dependence is strong the time series is called *persistent*.

- Notation

cross section analysis:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + u_i \ , \quad i = 1, 2, \ldots, n$$

time series analysis:

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + u_t \ , \quad t = 1, 2, \ldots, n$$

Week 2

- Random variable

A **random variable** is a rule that assigns a numerical outcome to an event in each possible state of the world.

- A **discrete random variable** has a finite number of distinct outcomes. For example, rolling a die is a random variable with 6 distinct outcomes.
- A **continuous random variable** can take a continuum of values within some interval. For example, rainfall in Melbourne in May can be any number in the range from 0.00 to 200.00 mm.
- discrete random variable

The probability density function (pdf) for a discrete random variable X is a function f with f(xi) = pi, i = 1, 2, . . . , m and f(x) = 0 for all other x.

$$P(X = x) = \begin{cases} P(X = 1) & x = 1 \\ P(X = 0) & x = 0 \end{cases}$$

$$P(X = 1) = p_1, \ P(X = x_2) = p_2, \ \ldots, P(X = x_m) = p_m$$

- Properties

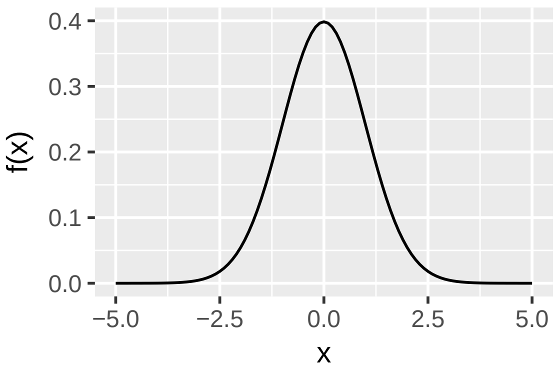
$$\sum_{i=1}^{m} p_i = p_1 + p_2 + \cdots + p_m = 1$$

$$0 \le p \le 1$$

- continuous random variable

The probability density function (pdf) for a continuous random variable X is a function f such that $P(a \leq X \leq b)$ is the area under the pdf between a and b

The total area under the pdf is equal to 1.



- Expected value

$$E(X) = p_1 x_1 + p_2 x_2 + \cdots + p_M x_m = \sum_{i=1}^{m} p_i x_i$$

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

- median, mode
- Variance

$$\sigma_X^2 = Var(X) = E(X - \mu_x)^2$$

Variance is a measure of spread of the distribution of X around its mean.

If X is an action with different possible outcomes, then Var(X) gives an indication of *riskiness* of that action.

- Standard deviation

$$\sigma_X = sd(X) = \sqrt{E(X - \mu_x)^2}$$

In finance, standard deviation is called the *volatility* in X.
The advantage of standard deviation over variance is that it has the same units as X.

- Properties of the Expected Value

1. For any constant c, $E(c) = c$.
2. For any constants a and b,

$$E(aX + b) = aE(X) + b$$

3. Expected value is a linear operator, meaning that expected value of sum of several variables is the sum of their expected values:

$$E(X + Y + Z) = E(X) + E(Y) + E(Z)$$

$$E(a + bX + cY + dZ) = a + bE(X) + cE(Y) + dE(Z)$$

$$E(X^2) \neq (E(X))^2$$

$$E(\log X) \neq \log\left(E(X)\right)$$

$$Var(X) = E(X^2) - \mu^2$$

Population parameter	its estimater
$\mu_y = E(y)$	$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$
$\sigma_y^2 = E(y-\mu_y)^2$	$s_y^2 = \hat{\sigma}_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \hat{y})^2$
$\sigma_y = \sqrt{\sigma_y^2}$	$\hat{\sigma_y} = \sqrt{\hat{\sigma}_y^2}$
$\sigma_{xy} = E(x-\mu_x)(y-\mu_y)$	$\hat{\sigma}_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$
$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$	$\hat{\rho}_{xy} = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y}$
$E(y\|x) = \beta_0 + \beta_1 x$	$\hat{\beta}_1 = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y} \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

Independent properties

$$P(A|B) = P(A)$$

$$P(AB) = P(A) \times P(B)$$

$$E(XY) = E(x)E(y)$$

$$Cov(x,y) = E(xy) - E(x)E(y) = 0$$

$$Corr(x,y) = 0$$

- Diversification in econometrics - Averaging

$$E(\frac{1}{2}(X_1 + X_2)) = \frac{1}{2}(\mu + \mu) = \mu$$

$$\begin{aligned} Var(\frac{1}{2}(X_1 + X_2)) &= \frac{1}{4}Var(X_1) + \frac{1}{4}Var(X_2) \\ &= \frac{1}{4}(\sigma^2 + \sigma^2) = \sigma^2/2 \end{aligned}$$

Week 3

A complete model for a random variable is a model of its probability distribution.

e.g. $E(y|x) = \beta_0 + \beta_1 x$

- **Law of Probability**

1. Probability of any event is a number between 0 and 1. The probabilities of all possible outcomes of a random variable add up to 1
2. If A and B are mutually exclusice events then $P(A\ or\ B) = P(A) + P(B)$
3. If A and B are two events, then $P(A|B) = \frac{P(A\ and\ B)}{P(B)}$

yx	1	2	3	$marginal\ f_y$
1	0.40	0.24	0.04	
2	0	0.16	0.16	
$marginal\ f_x$				

P(y = 1|x = 1) = ?
p(y = 2|x = 2) = ?

yx	1	2	3	$marginal\ f_y$
1	0.40	0.24	0.04	
2	0	0.16	0.16	
$marginal\ f_x$				

$$P(y=1|x=1) = \frac{P(AB)}{P(B)} = \frac{P(y=1\ \&\ x=1)}{p(x=1)} = \frac{0.4}{0.4} = 1$$

$$P(y=2|x=1) = \frac{P(AB)}{P(B)} = \frac{P(y=2\ \&\ x=1)}{p(x=1)} = \frac{0}{0.4} = 0$$

$$P(y=1|x=2) =$$

$$P(y=2|x=2) =$$

$$P(y=1|x=3) =$$

$$P(y=2|x=3) =$$

Expected value?

yx	1	2	3	$marginal\ f_y$
1	0.40	0.24	0.04	
2	0	0.16	0.16	
$marginal\ f_x$				

$$\begin{array}{rr}
P(y=1|x=1) = 1 & P(y=2|x=1) = 0 \\
P(y=1|x=2) = 0.6 & P(y=2|x=2) = 0.4 \\
P(y=1|x=3) = 0.2 & P(y=2|x=3) = 0.8
\end{array}$$

$$E(y|x=1) = 1 \times P(y=1|x=1) + 2 \times P(y=2|x=1) = 1$$
$$E(y|x=2) =$$
$$E(y|x=3) =$$

$$E(y|x=2) = 1.4$$
$$E(y|x=3) = 1.8$$

$$E(y|x) = \alpha + \beta x$$

When y and x have many possible outcomes or when they are continuous random variables we cannot enumerate the joint density and perform the same exercise.

$$y = \beta_0 + \beta_1 x + u \quad \text{where} \quad E(u|x) = 0$$

Implies that

$$E(u) = 0$$

- **L.I.E.** (Law of Iterated Expections)

$$E(E(X|Y)) = E(X)$$

Recall: A complete model for a random variable is a model of its probability distribution.

$$y = \beta_0 + \beta_1 x + u \quad \text{where} \quad E(u|x) = 0$$

Add assumptions:

$$Var(u|x) = \sigma^2$$
$$u|x \sim N$$

We make the assumption that in the big scheme of things, data are generated by this model, and we want to use observed data to learn the unknowns β_0, β_1 and σ^2 in order to predict y using x.

The method to get there parameters:

$$\begin{array}{rcl}
\text{Population parameter} & | & \text{its estimater} \\
\mu_y = E(y) & | & \bar{y} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} y_i \\
\sigma_y^2 = E(y-\mu_y)^2 & | & s_y^2 = \hat{\sigma}_y^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(y_i-\hat{y})^2 \\
\sigma_y = \sqrt{\sigma_y^2} & | & \hat{\sigma_y} = \sqrt{\hat{\sigma}_y^2} \\
\sigma_{xy} = E(x-\mu_x)(y-\mu_y) & | & \hat{\sigma}_{xy} = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) \\
\rho_{xy} = \dfrac{\sigma_{xy}}{\sigma_x\sigma_y} & | & \hat{\rho}_{xy} = \dfrac{\hat{\sigma}_{xy}}{\hat{\sigma}_x\hat{\sigma}_y} \\
E(y|x) = \beta_0 + \beta_1 x & | & \hat{\beta}_1 = \dfrac{\hat{\sigma}_{xy}}{\hat{\sigma}_x\hat{\sigma}_y} \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}
\end{array}$$

- Consistency of $\hat{\sigma}^2$

$$s_y^2 = \hat{\sigma}_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \hat{y})^2$$

$$\begin{aligned}
E[\hat{\sigma}_y^2] &= E\left[\frac{1}{n-1}\sum_{i=1}^{n}\big(y_i - \bar{y}\big)^2\right] \\
&= E\bigg[\frac{1}{n-1}\sum_{i=1}^{n}\bigg((y_i - \mu) - (\bar{y} - \mu)\bigg)^2\bigg] \\
&= E\bigg[\frac{1}{n-1}\sum_{i=1}^{n}\bigg((y_i - \mu)^2 - 2(\bar{y} - \mu)(y_i - \mu) + (\bar{y} - \mu)^2\bigg)\bigg] \\
&= E\bigg[\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \mu)^2 - \frac{2}{n-1}(\bar{y} - \mu)\sum_{i=1}^{n}(y_i - \mu) + \frac{n}{n-1}(\bar{y} - \mu)^2\bigg] \\
&= E\bigg[\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \mu)^2 - \frac{2}{n-1}(\bar{y} - \mu)\cdot n\cdot(\bar{y} - \mu) + \frac{n}{n-1}(\bar{y} - \mu)^2\bigg] \\
&= E\bigg[\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \mu)^2 - \frac{n}{n-1}(\bar{y} - \mu)^2\bigg] \\
&= \frac{1}{n-1}\sum_{i=1}^{n}E((y_i - \mu)^2) - \frac{n}{n-1}E((\bar{y} - \mu)^2) \\
&= \frac{n}{n-1}\sigma^2 - \frac{n}{n-1}\frac{1}{n}\sigma^2 \\
&= \sigma^2
\end{aligned}$$

$$\begin{aligned}
E[\hat{\sigma}_y^2] &= E\left[\frac{1}{n}\sum_{i=1}^{n}\big(y_i - \bar{y}\big)^2\right] \\
&= \frac{n}{n}\sigma^2 - \frac{n}{n}\frac{1}{n}\sigma^2 \\
&= \frac{n-1}{n}\sigma^2
\end{aligned}$$

- Geometry

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}\beta_0 + \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}\beta_1 + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \text{and} \quad \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

$$\underset{n\times 1}{\mathbf{y}} = \underset{n\times(k+1)}{\mathbf{X}}\;\underset{(k+1)\times 1}{\beta} + \underset{n\times 1}{\mathbf{u}}$$

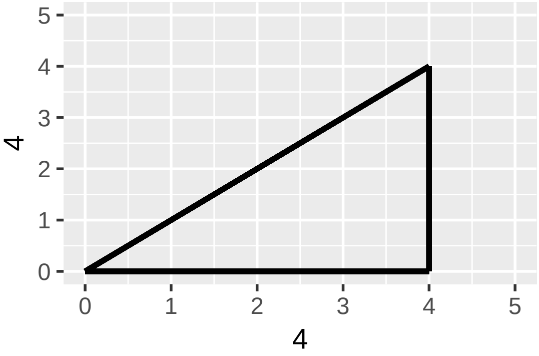
$$\underset{n \times (k+1)}{\mathbf{X}} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \qquad \underset{(k+1) \times 1}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

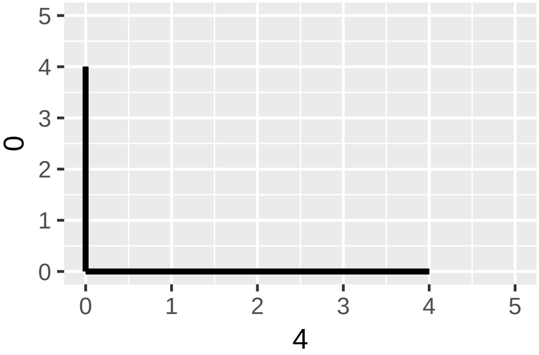
- Vector
  - $length(\mathbf{u}) = (\mathbf{u}'\mathbf{u})^{1/2}$

$$length \begin{bmatrix} 4 \\ 4 \end{bmatrix} = \left(\begin{bmatrix} 4 & 4 \end{bmatrix} \begin{bmatrix} 4 \\ 4 \end{bmatrix} \right)^{1/2} = \sqrt{32}$$

  - For $\mathbf{u}$ and $\mathbf{v}$ of the same dimension
    - $\mathbf{u}'\mathbf{v} = 0 \Leftrightarrow \mathbf{u}$ and $\mathbf{v}$ are perpendicular (orthogonal) to wach other

$$\begin{bmatrix} 4 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 4 \end{bmatrix} = 0$$





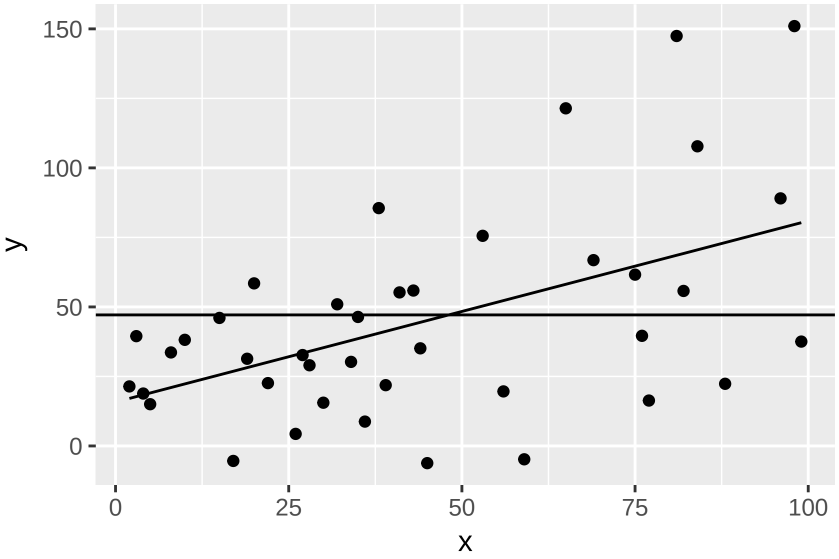
$\mathbf{Y}$ be explained as a combination of columns of $\mathbf{X}$ (the column space of $\mathbf{X}$) with zero $\hat{\mathbf{u}}$ (orthogonal to each other)

- ***Derivation of $\hat{\beta}$ in matrix***

$$\begin{aligned} \mathbf{y} = \mathbf{X}\hat{\beta} + \hat{\mathbf{u}} &\Rightarrow \hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\beta} \\ \mathbf{X}'\hat{\mathbf{u}} &= 0 \\ \mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\beta}) = 0 &\Rightarrow \mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\hat{\beta} \\ &\Rightarrow \hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \end{aligned}$$

For $\mathbf{X'X}$ to be invertible ($(\mathbf{X'X})^{-1}$ exists while $(\mathbf{X'X})^{-1}(\mathbf{X'X}) = \mathbf{I}$), columns of $\mathbf{X}$ must be linearly independent.
The vector of OLS predicted value $\hat{\mathbf{y}}$ is the prthogmal projection of $\mathbf{y}$ in the column space of $\mathbf{X}$.

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} = \mathbf{X(X'X)^{-1}X'y}$$



$\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{u}}$

$\mathbf{y}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{u}}$ form a right-angled triangle. + $length(\hat{\mathbf{y}}) = (\hat{\mathbf{y}}'\hat{\mathbf{y}})^{1/2} \Rightarrow$

$$\mathbf{y'y} = \hat{\mathbf{y}}'\hat{\mathbf{y}} + \hat{\mathbf{u}}'\hat{\mathbf{u}}$$

i.e.

$$\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} \hat{y}_i^2 + \sum_{i=1}^{n} \hat{u}_i^2$$

Subtracting $n\bar{y}^2$ from both sides

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} \hat{u}_i^2$$

Using $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i \Rightarrow \bar{y} = \bar{\hat{y}}$ since $(1 \;\; 1 \;\; \cdots \;\; 1)\hat{\mathbf{u}} = 0$ i.e.

$$SST = SSE + SSR$$

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

Week 4

- An estimator (sample → population) is *an unbiased estimater* of a parameter of interest if its expected value is the parameter of interest.
- Under the following assumptions $E(\hat{\beta}) = \beta$. i.e. unbiased.

Multiple Regression Model Assumptions

MLR.1 Linear in Parameters $y = \beta_0 = \beta_1 x_1 + \cdots + \beta_k x_k + u$	**E.1 Linear in parameters** $\underset{\mathbf{n\times 1}}{\mathbf{y}} = \underset{\mathbf{n\times(k+1)}}{\mathbf{X}} \underset{\mathbf{(k+1)\times 1}}{\beta} + \underset{\mathbf{n\times 1}}{\mathbf{u}}$
MLR.2 Random Sampling *We have a sample of n observations*	**E.3 Zero Conditional Mean** $E(\mathbf{u}\|\mathbf{X}) = \underset{\mathbf{(n\times 1)}}{\mathbf{0}}$
MLR.4 Zero Conditional Mean $E(u\|x_1, x_2, \cdots, x_k) = 0$	
MLR.3 No Perfect Collinearity *None of* $x_1,\ x_1,\ \cdots,\ x_k$ *is a constant and there are no exact linear relationships among them*	**E.2 No Perfect Collinearity** $\mathbf{X}$ *has rank* $k+1$
MLR.5 Homoskedasticity $Var(u\|x_1,\ x_2,\ \cdots,\ x_k) = \sigma^2$	**E.4 Homo + Randomness** $Var(\mathbf{u}\|\mathbf{X}) = \sigma^2 \mathbf{I}_n$
MLR.6 or E.5 Normality Conditional on $\mathbf{X}$ the population errors are normally distributed	

- Results from the assumptions
  E.1-E.3 Ubiasedness
  E.1-E.4 BLUE E.1-E.5 T-test, F-test

- ***Unbiasedness***

To show $E(\hat{\beta}) = \beta$, we need **E.2** (E.2 $\Rightarrow$ $\mathbf{X'X}$ can have a inverse).
Using **E.1**

$$\hat{\beta} = (\mathbf{X'X})^{-1}\mathbf{X'y} = (\mathbf{X'X})^{-1}\mathbf{X'}(\mathbf{X}\beta + \mathbf{u}) = \beta + (\mathbf{X'X})^{-1}\mathbf{X'u}$$

$$E(\hat{\beta}) = E[\beta + (\mathbf{X'X})^{-1}\mathbf{X'u}] = \beta + E((\mathbf{X'X})^{-1}\mathbf{X'u}) \overset{\mathbf{E.3}}{=} \beta$$

since $E(\mathbf{u}|\mathbf{X}) = 0 \Rightarrow E((\mathbf{X'X})^{-1}\mathbf{X'u}) = 0$

Zero conditional mean is not a problem for *predictive analysis* because the x set is all we have, we don't want to and won't get causal effect.

However, for *presctiptive analysis*, if we don't controll variables that may be affect by the x we are interested in, we won't get causal effect.

True model:

$$wage = \beta_0 + \beta_1 educ + \beta_2 ability + u$$

Estimate:

$$wage = \hat{\alpha}_0 + \hat{\alpha}_1 educ + \hat{v}$$

$$E(\hat{\alpha}_1) = \beta_1 + \beta_2 \frac{\partial\ ability}{\partial\ educ} \neq \beta_1$$

"omitted variable bais"

Zero conditional mean requires **Strictly exogenous** in time series $E(u_t|X_{s1}, X_{s2}, \cdots, X_{sk}) = 0 \forall s = 1, 2, \cdots, T$
implication: the error term in any time period t is uncorrelated with each of the regressors in all time periods,

past, present, and future.

$$Corr(u_t, x_{11}) = \cdots = Corr(u_t, x_{T1}) = Corr(u_t, x_{1k}) = \cdots = Corr(u_t, x_{Tk}) = 0$$

It is violated when the regression model contains a lag of the dependent variable as a regressor.

- variance-covariance matrix

For a random variable $\mathbf{v}$ and its expectation

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \qquad E(\mathbf{v}) = \underset{n\times 1}{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}$$

variance σ_i^2 covariance σ_{ij}^2
$Var(\mathbf{A}'\mathbf{v}) = \mathbf{A}'Var(\mathbf{v})\mathbf{A}$

The variance-covariance matrix for $\mathbf{v}$

$$Var(\mathbf{v}) = E(\mathbf{v}-\mu)(\mathbf{v}-\mu)' = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix}$$

- The variance of OLS estimator

$$Var(\hat{\beta}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

Proof:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta+\mathbf{u}) = \beta + \underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'}_{\mathbf{A}'}\mathbf{u}$$

$$\begin{aligned} Var(\hat{\beta}|\mathbf{X}) &= Var(\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}|\mathbf{A}) \\ &= Var(\mathbf{A}'\mathbf{u}|\mathbf{A}) \\ &= \mathbf{A}'Var(\mathbf{u})\mathbf{A} \\ E.5 \ \Rightarrow Var(\hat{\beta}) &= \mathbf{A}'\sigma^2\mathbf{I}_n\mathbf{A} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

To estimate σ^2, we can use the unbiased estiomator

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

Think about dimension: $\hat{y}$ is in the column space of X, so it is in a subspace with simension $k+1$

$\hat{\mathbf{u}}$ us orthogonal to column space of $\mathbf{X}$, so it is in a subspace with dimension $n-(k+1) = n-k-1$. So even though there are n coordinates in $\hat{\mathbf{u}}$, only $n-k-1$ of those are free (it has $n-k-1$ *degree of freedom*)

- *B.L.U.E.*

- What is BLUE

  - Best(smallest variance) Linear Unbiased Estimator

- How to prove BLUE
  - Using E.1-E.4, based on Gauss-Markov Theorem, $\hat{\beta}$ is BLUE of β
- ***Interpretation***

$$Murder = \beta_0 + \beta_1 Assault + \beta_2 Population + u$$

$$\widehat{Murder} = \underset{(1.74)}{3.21} + \underset{0.005}{0.044} Assult - \underset{0.027}{0.045} UrbanPop$$

```
##
## Call:
## lm(formula = Murder ~ Assault + UrbanPop, data = USArrests)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5530 -1.7093 -0.3677  1.2284  7.5985
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.207153   1.740790   1.842   0.0717 .
## Assault      0.043910   0.004579   9.590 1.22e-12 ***
## UrbanPop    -0.044510   0.026363  -1.688   0.0980 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.58 on 47 degrees of freedom
## Multiple R-squared:  0.6634, Adjusted R-squared:  0.6491
## F-statistic: 46.32 on 2 and 47 DF,  p-value: 7.704e-12
```

- Rescaling (e.g. change the unit)

- do not create any new information, so it only changes OLS results in predictable and non-substantive ways.

$$\begin{aligned} \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 &\Rightarrow \hat{y} = \hat{\beta}_0^* + \hat{\beta}_1^* x_1^* + \hat{\beta}_2^* x_2 \\ 1.\quad x_1 \rightarrow c x_1 \quad x_1^* = c x_1 & \\ &\Rightarrow \hat{\beta}_0^* = \hat{\beta}_0 \quad \hat{\beta}_1^* = \hat{\beta}_1 / c \quad \hat{\beta}_2^* = \hat{\beta}_2 \\ 2.\quad x_1 \rightarrow a + c x_1 \quad x_1^* = a + c x_1 & \\ &\Rightarrow \hat{\beta}_0^* = \hat{\beta}_0 - \hat{\beta}_1 / c \quad \hat{\beta}_1^* = \hat{\beta}_1 / c \quad \hat{\beta}_2^* = \hat{\beta}_2 \end{aligned}$$

Since $\hat{y}$ does not change residuals, SST, SSE, SSR stay the same, so R^2 will not change.

$$\begin{aligned} y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{u} &\Rightarrow y = \hat{\beta}_0^* + \hat{\beta}_1^* x_1^* + \hat{\beta}_2^* x_2 + \hat{u}^* \\ y \rightarrow c y \quad y^* = c y & \\ &\Rightarrow \hat{\beta}_0^* = c \hat{\beta}_0 \quad \hat{\beta}_1^* = c \hat{\beta}_1 \quad \hat{\beta}_2^* = c \hat{\beta}_2 \quad \hat{u}^* = c \hat{u} \end{aligned}$$

SST, SSE, SSR all change (multiplied by) same amount, so R^2 will not change.

Week 5

$$E(\hat{\beta}|\mathbf{X}) = \beta \qquad Var(\hat{\beta}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

$$\mathbf{u}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I_n})$$

- Assumption MLR.6 or E.5 (Normality): Conditional on X the population errors are normally distributed.

$$Normal\ Distribution\ + Randomness \Rightarrow i.i.d.(independent\ identical\ distribution)$$

i.e. Conditional on explanatory variables, population errors $\mathbf{u}_i$ are i.i.d. $N(0, \sigma^2)$

- MLR.1 - MLR.6 are **Classical Linear Model (CLM)** assumptions.

Under CLM assumptions

$$\hat{\beta}|\mathbf{X} \sim N(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$$

$$\hat{\beta}_j|X \sim N(\beta_j, Var(\hat{\beta}_j))$$

$$Var(\hat{\beta}_j) = \sigma^2\{(\mathbf{X}'\mathbf{X})^{-1}\}_{jj}$$

$$\frac{\hat{\beta}_j - \beta_j}{sd(\beta_j)} \sim N(0,1)$$

$sd(\hat{\beta}_j)$ depends on σ, which is unknown.

- Using $\hat{\sigma}$ as an estimator of σ, instead of normal distribution, we are getting a **t distribution**.

$$\frac{\hat{\beta}_j - \beta_j}{se(\beta_j)} \sim t_{n-k-1} = t_{df}$$

t distribution has fatter tails than N(0,1).
As df increases, it gets more similar to N(0,1).

- ***T TEST***

t statistic (or *t ratio*)

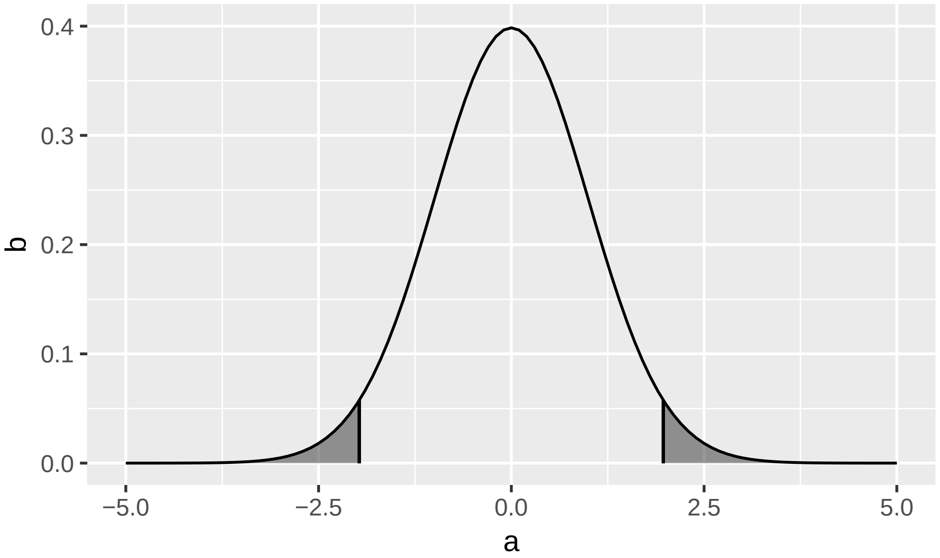
$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

The size or the **significant level** α
the probability that we reject the null when it is true (*Type I error*)

If $H_0 : \beta_j = r$, then $\frac{\hat{\beta}_j - r}{se(\beta_j)} \sim t_{n-k-1}$

A $(1-\alpha)\%$ **confidence interval** is defined as $\hat{\beta}_j \pm c \times se(\hat{\beta}_j)$, where c is $(1-\frac{\alpha}{2})$ percentile of a t_{n-k-1} distribution (Two sided test).

P Value is the probability that the realization falling out of the range between negative t statistic and positive t statistic. (two sided test)



- **Steps of testing a hypothesis**

1. Specify the null. e.g. $H_0 : \beta_1 = 0$
2. Specify the alternative. e.g. $H_1 : \beta_1 > 0$ (one sided) or $H_1 : \beta_1 \neq 0$ (two sided)
3. State the test statistic and its distribution under H_0. e.g. $\frac{\hat{\beta}_j}{se(\beta_j)} \sim t_{n-k-1}$ **under $\mathbf{H_0}$**
4. Specify the level of significance of the test i.e. $\alpha = 0.05$
5. Find the critical value from the distribution of the test statistic with reference to the alternative hypothesis and the desired significant level, and exercise the rejection rule.
6. See if the value of t_{calc} test statistic in your sample t_{calc} is inside or outside the rejection zone and express your conclusion with a sensitive sentence

- **Confidence Intervals**

Another way to use classical statistical testing is to construct a confidence interval using the same critical value as was used for a two-sided test.

A $(1-\alpha)\%$ confidence intercal is defined as

$$\hat{\beta}_j \pm c \times se(\hat{\beta}_j)$$

where c is the $(1-\frac{\alpha}{2})$ percentile of a t_{n-k-1} distribution

The interpretation of a $(1-\alpha)\%$ confidence interval is that the interval will cover the true parameter with probability $(1-\alpha)$

If the confidence interval does not contain the value in null hypothesis, we can reject the null hypothesis at the $\alpha\%$ level.
If the confidence interval does not contain zero, we can deduce that x_j is statistically significant at the $\alpha\%$ level.

- **F-test**

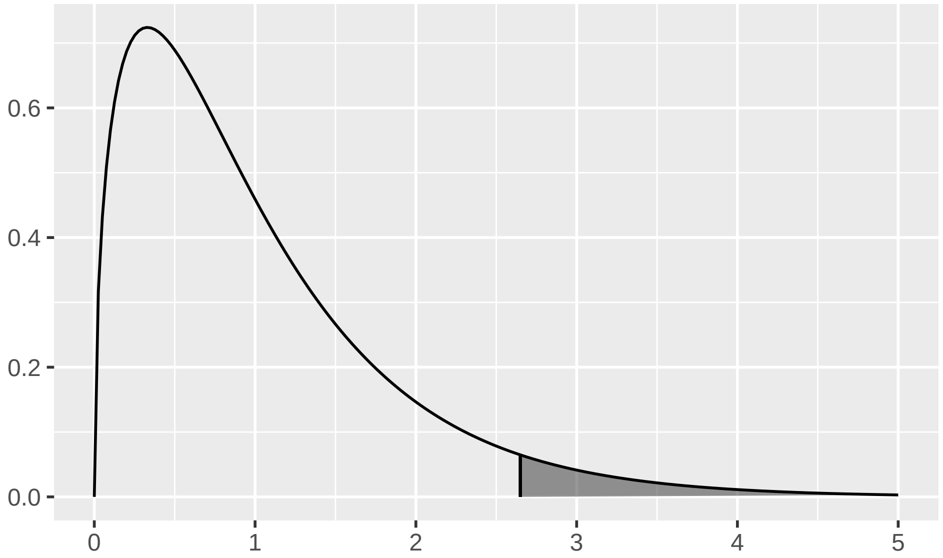
The overall sifnificant: The alternative can only be that at least one of these restriction is not true (i.e. at least one is sifnificant).

We estimatie two equtions: *the unrestricted model* and *the restricted model*

$$F = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)} \sim F_{q,n-k-1} \ under \ H_0$$

q: the numerator df (the number of restriction) $n-k-1$: the denominator of df

F-statistic is always postive ($SSR_r > SSR_{ur}$)



- **A useful formulation of the F-test**

The SST of the restricted and the unrestricted models are the same.

$$SSR_r = (1-R_r^2)SST \qquad SSR_{ur} = (1-R_{ur}^2)SST$$

$$F = \frac{(R_{ur}^2 - R_r^2)/q}{(1-R_{ur}^2)/(n-k-1)}$$

F-test for *overall significant* of a model for the special null hypothesis is that all slop parameter are zero

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} \sim F_{k,n-k-1} \ under \ H_0$$

$$F = \frac{(SSR_r - SSR_{ur})/3}{SSR_{ur}/(n-4)} \sim F_{3,n-4} \ under \ H_0$$

$$H_0 : \beta_0 = 1 \ and \ \beta_2 = \beta_3 = 0$$

$$price_i = \beta_0 + \beta_1 assess_i + \beta_2 area_i + \beta_3 bed_i + u_i$$

$$price_i = \beta_0 + assess_i + u_i$$
$$\Rightarrow price_i - assess_i = \beta_0 + u_i$$

- **Reparameterisation**
  one-sided single restriction, more than one parameter

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

$$H_0 : \beta_1 = \beta_2$$
$$H_1 : \beta_1 > \beta_2$$

$$\text{Define } \delta = \beta_1 - \beta_2 \qquad \hat{\delta} = \hat{\beta}_1 - \hat{\beta}_2$$

$$H_0 : \delta = 0 \quad H_1 : \delta > 0$$

$$Var(\hat{\delta}) = Var(\hat{\beta}_1) + Var(\hat{\beta}_2) - 2Cov(\hat{\beta}_1, \hat{\beta}_2)$$

Under CLM assumptions $\hat{\beta}$ condotional on x is normally distribution

$$\beta_1 = \delta + \beta_2$$

$$y = \beta_0 + (\delta + \beta_2)x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

$$\Rightarrow y = \beta_0 + \delta x_1 + \beta_2(x_1 + x_2) + \beta_3 x_3 + u$$

Week 6

The linear regression model only need to be linear in parameters.

$$\log y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$
$$y = \beta_0 + \beta_1 \log x_1 + \beta_2 x_2 + u$$
$$\log y = \beta_0 + \beta_1 \log x_1 + \beta_2 x_2 + u$$

$Model$	$Dep\ Var$	$Indep\ Var$	$Interpretation$
$level - level$	y	x	$\Delta y = \beta_1 \Delta x$
$level - log$	y	$\log x$	$\Delta y = \frac{\beta_1}{100}(\%\Delta x)$
$log - level$	$logy$	x	$\%\Delta y = 100\beta(\Delta x)$
$log - log$	$logy$	$\log x$	$\%\Delta y = \beta(\%\Delta x)$

A strictly positive ranged variable can be logged $postively\ \ skewed \xrightarrow{log} less\ \ skewed$
Explanatory variables measured in years are not logged.
Variables that are already in percentages are not logged.
If a variable is positively skewed (like income or wealth), taking logarithms makes its distribution less skewed.

- Choose $\log y$ or y

$$\log y = \widehat{\log y} + \widehat{u}$$
$$y = e^{\widehat{\log y} + \widehat{u}} = e^{\widehat{\log y}} \times e^{\widehat{u}}$$

While $\widehat{u}$ has mean zero, the expected value of $e^{\widehat{u}}$ is not equal to 1 (expectation is applied only up to linear transformation). It is a constant bigger than 1, called α
i.e. $\hat{y}_{from\ log} = \alpha e^{\widehat{\log y}}$
To get α, we regress y on $e^{\widehat{\log y}}$ with no constant.
If $\alpha < 1$, we take $\alpha = 1$

To chose between model of $\log y$ and y, we can select log or level model based on which models prediction has a higher corrletion with y (the true value in the sample).

i.e.

$$\widehat{Corr}(y, \hat{y}) \quad v.s. \quad \widehat{Corr}(y, \hat{y}_{from\ log})$$

- Quadratic terms

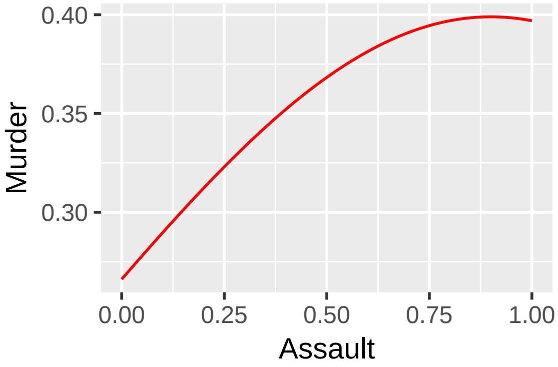
$$\widehat{Murder} = 3.207 + 0.0439 Assault - 0.0445 Population$$

$$\widehat{Murder} = 1.514 + 0.07835 Assault - 0.000094 Assault^2 - 0.0568 Population$$

$$Murder = \beta_0 + \beta_1 Assault + \beta_2 Assault^2 + \beta_3 Population + u$$

$$\frac{\Delta Murder}{\Delta Assault} = \beta_1 + 2\beta_2 Assault$$

dependes on the number of assault alredy happened.



```
Maximum or minimum?
```

- **Transforming non-stationary series to stationary seies**

Simple return : from time $t-1$ to t

Simple gross return

$$1 + R_t = \frac{P_t}{P_{t-1}}$$

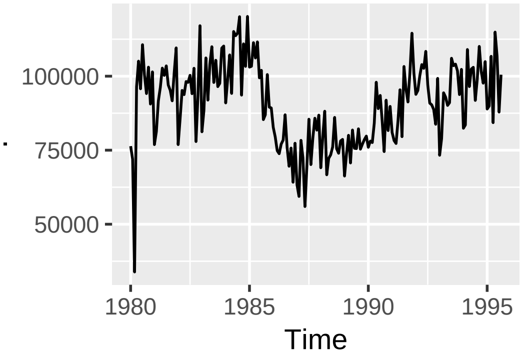
Simple net return

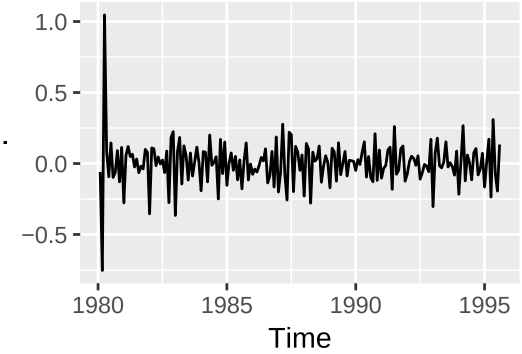
$$R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}}$$

Log return : The natural logarithm of the simple gross return is called the log return

$$r_t = \ln\left(1 + R_t\right) = \ln\frac{P_t}{P_{t-1}} = \ln P_t - \ln P_{t-1}$$

For small R_t , $r_t = \ln\left(1 + R_t\right) \approx R_t$





- Prescriptive analytics
- Investigating **causal effect** of an x on a y, while adding all other possible variables as control variables.
- Inferences of all those other variables are minor objectives.
- Predictive analytics
- **Not** investigating causal effect.
- Simply trying to predic y based on some xs while keeping the model **simple**.
- *Parsimony* ***KISS Principle***: Keep it simple, stupid.

- **Model Selection Criteria**

General Form

$$I(k) = c \quad + \quad \underset{decreasing\ in\ k}{\ln\left[SSR(k)\right]} \quad + \quad \underset{increasing\ in\ k}{(k+1)\frac{\alpha}{T}}$$

1. $Adjusted\ R^2 \ \ (\bar{R^2})$

$\bar{R^2} = 1 - \frac{SSR/n-k-1}{SST/n-1}$

2. $AkaikeInformationCaiteria \ \ (AIC)$

$AIC = c_1 + \ln(SSR) + \frac{2k}{n}$

3. $Hannan-QuinnCriterion \ \ (HQ)$

$HQ = c_2 + \ln(SSR) + \frac{2k\ln(\ln(n))}{n}$

4. $Schwarz\ or\ Bayesian\ Information\ Criterion \ \ (SIC\ or\ BIC)$

$BIC = c_3 + \ln(SSR) + \frac{k\ln(n)}{n}$

order of penalties

$$P(BIC) > P(HQ) > P(AIC) > P(\bar{R^2})$$

- **Predictions and Prediction Intervals**

Two Sources of Error

- 1.Estimation uncertainty: caused by not knowing the value of the true parameters
- 2.u: not predictable by our predictor, even if we know the true value of β
- std.error → Estimation uncertainty

$$\widehat{Murder} = 6.41594 + 0.02093Assault$$

$$\widehat{Murder} = \beta_0 + \beta_1(Assault - 170)$$

The relative magnitudes of $Var(u)$ and $Var(\hat{y})$ is 1 to $1/n$, so often ignore estimation uncertainty.

Only consider estimation uncertainty, 95% confidence interval for $E(y_i|x_{i1}, \cdots, x_{ik})$:

$$\hat{y} \pm (cv(t_{n-k-1}, two\ tailed : 0.05) \times se(\hat{y})$$

95% prediction interval for y_i :

$$\hat{y} \pm (cv(t_{n-k-1}, two\ tailed : 0.05) \times se(\hat{e})$$

where

$$se(\hat{e}) = \sqrt{\hat{\sigma}^2 + [se(\hat{y})]^2}$$

Ignoring estimation uncertainty, 95% prediction interval for y_i :

$$\hat{y} \pm (cv(t_{n-k-1}, two\ tailed : 0.05) \times \hat{\sigma}$$

$$\widehat{Murder} = 6.41594 + 0.02093 Assault$$

```
##
## Call:
## lm(formula = Murder ~ Assault, data = USArrests)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.8528 -1.7456 -0.3979  1.3044  7.9256
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.631683   0.854776   0.739    0.464
## Assault     0.041909   0.004507   9.298  2.6e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.629 on 48 degrees of freedom
## Multiple R-squared:  0.643,  Adjusted R-squared:  0.6356
## F-statistic: 86.45 on 1 and 48 DF,  p-value: 2.596e-12
```

$$\widehat{Murder} = \beta_0 + \beta_1(Assault - 1000,000)$$

```
##
## Call:
## lm(formula = Murder ~ I(Assault - 170), data = USArrests)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.8528 -1.7456 -0.3979  1.3044  7.9256
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      7.756149   0.371863  20.858  < 2e-16 ***
## I(Assault - 170) 0.041909   0.004507   9.298  2.6e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.629 on 48 degrees of freedom
## Multiple R-squared:  0.643,  Adjusted R-squared:  0.6356
## F-statistic: 86.45 on 1 and 48 DF,  p-value: 2.596e-12
```

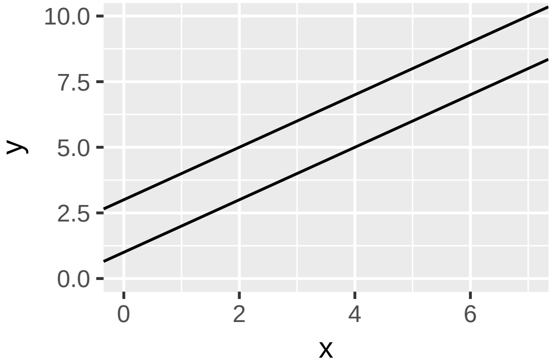
Week 7

- **Dummy variable**

A binary vaiable is a 0-1 variable whose value for observation i is 1 if that observation belongs to a category and 0 otherwise. e.g. $male_i = \begin{cases} 1, & \text{if } i \text{ is male} \\ 0, & \text{otherwise} \end{cases}$

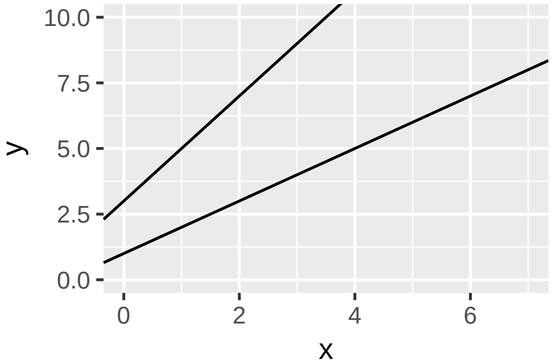
e.g.

$$\begin{aligned} E(wage_i|female_i, educ_i) &= \beta_0 + \delta_0 female_i + \beta_1 educ_i \\ E(wage_i|female_i = 0, educ_i) &= \beta_0 + \beta_1 educ_i \\ E(wage_i|female_i = 1, educ_i) &= (\beta_0 + \delta_0) + \beta_1 educ_i \end{aligned}$$



e.g.

$$\begin{aligned} E(wage_i|female_i, educ_i) &= \beta_0 + \delta_0 female_i + \beta_1 educ_i m + \delta_1 female_i \times educ_i \\ E(wage_i|female_i = 0, educ_i) &= \beta_0 + \beta_1 educ_i \\ E(wage_i|female_i = 1, educ_i) &= (\beta_0 + \delta_0) + (\beta_1 + \delta_1) educ_i \end{aligned}$$



The hypothesis of no difference in expected wage between men and women:

$$H_0 : \delta_0 = \delta_1 = 0 \qquad H_1 : \text{At least one of the two is not zero}$$

- **Multicollinearity**
- Because of high correlation between X_1 (FEMALE) and $X_1 \times X_2$ (FEMALE × EDUC) (for instance), data tells us that although there is strong evidence that expacted wage for men and women is not the same, it is hard to determine if the intercept is different or the slope.
- While perfect collinearity causes problems for the OLS estimator, multicollinearity does not affect any of the properties of OLS. It only makes it different to ascertain the contribution of each one x in a group of multicollinear x's to explaining y.

- **Dummy variable trap** : use all dummy variables in a regression.
- “base” or “benchmark” category - the omitted category
- Changing the benchmark category should not matter qualitatively.

`More than one dummy?`

- **A more accurate estimate of $\Delta log(y)$ given $\Delta x = 1$**

Often use $d\log(y) = \frac{1}{y}dy$ ($100\beta_1$ as the $\%\Delta$)

A better measure of $\%\Delta y$ as x_1 increases by 1 is $100(e^{\beta_1} - 1)$

- Reasoning:

We are looking for $\frac{(\hat{y}_{after} - \hat{y}_{before})}{\hat{y}_{before}}$

In the log-level model, we are getting $\widehat{\log(y)}_{after} - \widehat{\log(y)}_{before} = \beta_1$

Exponentiating and subtracting one gives $\frac{(\hat{y}_{after} - \hat{y}_{before})}{\hat{y}_{before}} = e^{\beta_1} - 1$

For $-0.10 < \beta_1 < 0.10$, this does not make much difference.

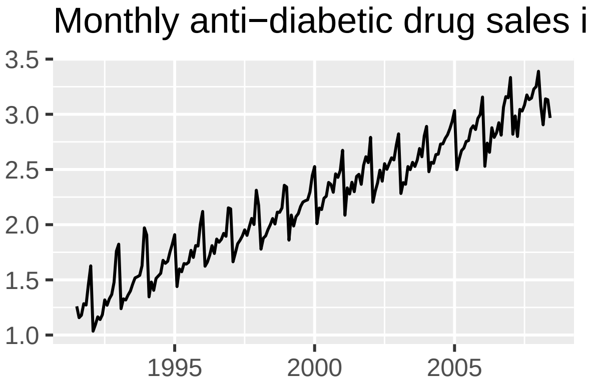
In practice, given estimation uncertainty, such approximation errors are understood and tolerated.

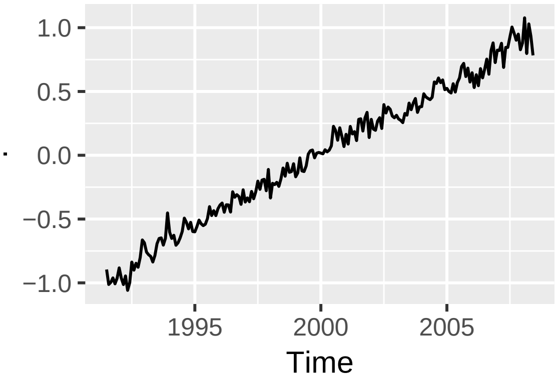
- **Seasonality**

Some time series exhibit certain cyclical/periodic behaviour. This is referred to as seasonality.

The procedure of removing seasonality is called **seasonal adjustment** .

- **Eliminating seasonality**
- Using seasonal differencing at where a pattern exists : $y_t = y_t - y_{t-i}$ where i is the lag of pattern
- Using dummy variables: $Qi_t = 1$ if t is the season with pattern i, and $Qi_t = 0$ if t is not. $y_t = \beta_0 + \beta_1 Qi_t + e_t$





Week 8

- **Doubting Homoskedasticity**

e.g. Variance of food consumption on poor and rich people

e.g. Using average of group instead of individual, variance depends inversely on group size

e.g. In finance, unpredicted news increase the volatility of the market

When HTSK, the estimator is not BLUE with unreliable T and F test. Also, $Var(\hat{\beta}) \neq \sigma^2(\mathbf{X'X})^{-1}$

Since $E(u_i|x_{i1}, \cdots, x_{ik}) = 0$, $Var(u_i|x_{i1}, \cdots, x_{ik}) = E(u_i^2|x_{i1}, \cdots, x_{ik})$

- **Breusch-Pagan test**

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i \ for \ i = 1, \cdots, n$$

$$\begin{aligned} &H_0 : E(u_i^2|x_{i1}, \cdots, x_{ik}) = \sigma^2 \ for \ i = 1, \cdots, n \\ &H_1 : E(u_i^2|x_{i1}, \cdots, x_{ik}) = \delta_0 + \delta_1 z_{i1} + \delta_2 z_{i2} + \cdots + \delta_q z_{iq} \\ &\quad \text{where } z_{i1}, z_{i2}, \cdots z_{iq} \text{ are a subset of } x_{i1}, \cdots, x_{ik} \end{aligned}$$

In fact the z variables can include some variables that do not appear in the conditional mean, but may affect the variance.

1. Estimate the model by OLS as usual. Obtain $\hat{u}_i$ for i=1, $\cdots$, n and square them.
2. Regress $\hat{u}_i^2$ on a constant $z_{i1}, z_{i2}, \cdots z_{iq}$. Denote the R^2 of this auxiliary regression by $R^2_{\hat{u}_i^2}$
3. Under H_0 , the statistic $n \times R^2_{\hat{u}_i^2}$ has a χ^2 distribution with q degrees of freedom in large samples. The satistic is called the Lagrange Multiplier (LM) statistic for HTSK.
4. Given the desired level of significance, we obtain the cv of the test from the χ^2 table, and reject H_0 if the value of the test statistic is larger than the cv.

F-test is also useful ($F_{q, \ n-q-1}$)

- **White Test**

$$\begin{aligned} &H_0 : E(u_i^2|x_{i1}, \cdots, x_{ik}) = \sigma^2 \ for \ i = 1, \cdots, n \\ &H_1 : \text{the variance is a smooth unknown function of } \ x_{i1}, \cdots, x_{ik} \end{aligned}$$

Regress $\hat{u}_i^2$ on a constant, $x_{i1}, \cdots, x_{ik}$, and all pairwise crooss products of $x_{i1}, \cdots, x_{ik}$.

It has the power to dedect this general form of heteroskedasticity in large samples.

Test statistic $n \times R^2_{\hat{u}_i^2}$
Distribution χ^2 with degrees of freedom as the number of explanatory variables.

- **Special case of White Test**
  In the general white test, $k + k(k+1)$ regressors are too many.

Since $\hat{y}$ is a function of all the x s, $\hat{y}^2$ will be a function of the squares and crossproducts of all the x s.

Therefore, $\hat{y}$ and $\hat{y}^2$ can proxy for all of the x s' squares and crossproducts.

Regress the residuals squared on $\hat{y}$ and $\hat{y}^2$

- **Solution for HTSK 1 Robust Standard Errors**
  Live with the second best. Find the usable standard error for t, F test.

$$Var(\hat{\beta}|\mathbf{X}) = (\mathbf{X'X})^{-1}[\mathbf{X'}Var(\mathbf{u}|\mathbf{x})\mathbf{X}](\mathbf{X'X})^{-1}$$

With Homo

$$Var(\mathbf{u}|\mathbf{X}) = \sigma^2\mathbf{I_n} \Rightarrow Var(\hat{\beta}|\mathbf{X}) = \sigma^2(\mathbf{X'X})^{-1}$$

With HTSK

$$Var(\mathbf{u}|\mathbf{X}) = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

$$Var(\hat{\beta}|\mathbf{X}) = (\mathbf{X'X})^{-1}\left[\mathbf{X'}\begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}\mathbf{X}\right](\mathbf{X'X})^{-1}$$

$$\widehat{Var}(\hat{\beta}|\mathbf{X}) = (\mathbf{X'X})^{-1}\left[\mathbf{X'}\begin{pmatrix} \hat{u}_1^2 & 0 & \cdots & 0 \\ 0 & \hat{u}_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \hat{u}_n^2 \end{pmatrix}\mathbf{X}\right](\mathbf{X'X})^{-1}$$

which is a reliable estimator for $Var(\hat{\beta}|\mathbf{X})$ in large samples.

The square root of diagonal elements of this matrix are called White stansard error or Robust standard error, which are reliable for inference.

- **Solution for HTSK 2 Transforming the Model**
- a. Logaristhmic transformation of y solve the problem when population model has $\log(y)$ but we used y
- b. Weighted Least Squares (WLS): used when variance of each error is proportional to a known functuon of a single independent variable
- **Weighted Least Squares (WLS)**

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i \; for \; i = 1, \cdots, n$$

$$Var(u_i|x_{i1}, x_{i2}, \cdots, x_{ik}) = \sigma^2 h_i$$

where h_i is a known function of one of x s, or a function of a variable z as long as

$$E(u_i|x_{i1}, x_{i2}, \cdots, x_{ik}, z_i) = 0$$

e.g. $h_i = x_i$, or $h_i = x_i^2$, or $h_i = \frac{1}{z_i}$

Multiplying both sides of model equation by $w_i = \frac{1}{\sqrt{h_i}}$ eliminates HTSK:
weighted model

$$(w_i y_i) = \beta_0 w_i + \beta_1(w_i x_{i1}) + \beta_2(w_i x_{i2}) + \cdots + \beta_k(w_i x_{ik}) + (w_i u_i) \; for \; i = 1, \cdots, n$$

(no constant term) This estimator is called the *Weighted Least Squares (WLS)* estimator of β

Week 9

- **Inference in Time series**
  Under some assumptions:

  R1 B.L.U.E.

  R2 can use t-test

  R3 can use F-test

- **Assumptions**
- **A1** The model is linear in Parameters $y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + u_t$
- **A2** No perfect collinearity: None of the regressors can be expressed as an exact linear combination of the other regressors.
- **A3** Zero conditional mean: For each time period t, $E(u_t|\mathbf{X}) = 0$
- **A4** Homoskedasticity: For each timem period t, $Var(u_t|\mathbf{X}) = Var(u_t) = \sigma^2$
- **A6** No serial correlation: Conditional on X, the errors in two different time periods are uncorrelated. That is $Corr(u_t, u_s|\mathbf{X}) = 0 \forall t \neq s$
- **A5** Normality of the errors: The errors are independent of X and have identical and independent normal distributions, $u_t \sim n.i.d.(0, \sigma^2)$
- **Theorem(1)**
  When A1, A2, A3 hold, the OLS estimator $\hat{\beta}$ is an unbiased estimator of β , $E(\hat{\beta}) = \beta$
- **Theorem(2)**
  When A1, A2, A3, A4, A6 hold, $\hat{\beta}$ is the best linear unbiased estimator (BLUE) of β (Gauss-Markov Theorem)
- **Theorem(3)**
  When A1 to A6 hold, the OLS estimator of β is normally distributed. Can be tested using t-test and F-test.
- **Theorem(4)**
  When A1 to A6 hold, $Var(\hat{\beta}_j) = \frac{\sigma^2}{[SST_j(1-R_j^2)]} \quad j = 1, 2, \cdots, k$, $\hat{\sigma}^2 = \frac{SSR}{T-k-1}$ is an unbaised estimator of the error variance σ^2 , that is $E(\hat{\sigma}^2) = \sigma^2$

A3 requires **Strictly exogenous** $E(u_t|X_{s1}, X_{s2}, \cdots, X_{sk}) = 0 \forall s = 1, 2, \cdots, T$
implication: the error term in any time period t is uncorrelated with each of the regressors in all time periods, past, present, and future.

$$Corr(u_t, x_{11}) = \cdots = Corr(u_t, x_{T1}) = Corr(u_t, x_{1k}) = \cdots = Corr(u_t, x_{Tk}) = 0$$

Contemporaneous exogeneity: $E(u_t|\mathbf{X}_t) = E(u_t|X_{t1}, X_{t2}, \cdots, X_{tk}) = 0$
implication:

$$Corr(u_t, x_{t1}) = Corr(u_t, x_{t2}) = \cdots = Corr(u_t, x_{tk}) = 0 \qquad t = 1, 2, \cdots, T$$

Violated A3: when the model contains a lag of the dependent variables as a regressors (AR(p), ADL(p,q))

A3 only restrict correlation between error terms and regressors, but not regressors themselves or errors themselves.

A4 's problem, heteroskedasticity is more common in cross-sectional, but can also arise in time series.

A6 requires $Corr(u_t, u_s) = 0 \forall t \neq s$
When violated, the errors are **sutocorrelated** or **serially correlated** , which is very common when omitted

influences are correlated across time.

$$\text{Positive first-order autocorrelation: } Corr(u_t, u_{t-1}) > 0$$

Consequences of violations :

C1 The OLS estimator is no longer efficient.

C2 $t = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$ no longer has a t distribution, even asymptotically

C3 F no longer has an F distribution, even asymptotically.

C2 and C3 are most serious.

- **Testing for Autocorrelation in the error term**
- A simple test

- Breusch-Godfrey test
- **A test** for first-order autocorrelation with strictly exogenous regressors

$$y_t = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + u_t$$

$$u_t = \rho u_{t-1} + e_t \text{ where } |\rho| < 1 \ \& \ e_t \sim i.i.d.(0, \sigma_e^2)$$

Assuming $E(e_t|u_{t-1}, u_{t-2}, \cdots) = 0$ and $Var(e_t|u_{t-1}) = Var(e_t) = \sigma_e^2$

$$\begin{aligned} &H_0 : \rho = 0 \\ &H_1 : \rho \neq 0 \text{ (or } \rho > 0 \text{ in one sided test)} \end{aligned}$$

S1 Estimate $y_t = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + u_t$, obtain $\hat{u}_t$

S2 Estimate $\hat{u}_t = \rho \hat{u}_{t-1} + e_t$, then under H_0 $\frac{\hat{\rho}}{se\hat{\rho}} \overset{asy}{\sim} t(T-k-1)$

S3 Reject H_0 if $|t_{calc}| > t_{crit}$ ($t_{calc} > t_{crit}$ if one sided test)

Limitations : 1. Vaild only if the regressors are strictly exogenous
2. Only for first-order autocorrelation
3. An asymptotic test, may be unreliable in small samples

- **Breusch-Godfrey test**

If the error term in $y_t = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + u_t$ is autocorrelated of order q then $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_q u_{t-q} + e_t$

$$\begin{aligned} &H_0 : \rho_1 = \rho_2 = \cdots = \rho_q = 0 \\ &H_1 : \rho_j \neq 0 \text{ for at least one } \ j = 1, 2, 3, \cdots, q \end{aligned}$$

S1 Estimate $y_t = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + u_t$, obtain $\hat{u}_t$

S2 Estimate $\hat{u}_t = \alpha_1 + \alpha_2 x_{t2} + \cdots + \alpha_k x_{tk} \quad + \rho_1 \hat{u}_{t-1} + \cdots + \rho_q \hat{u}_{t-q} + e_t$, then under H_0 $BG = (T-q) R_{\hat{u}}^2 \overset{asy}{\sim} \chi^2(q)$

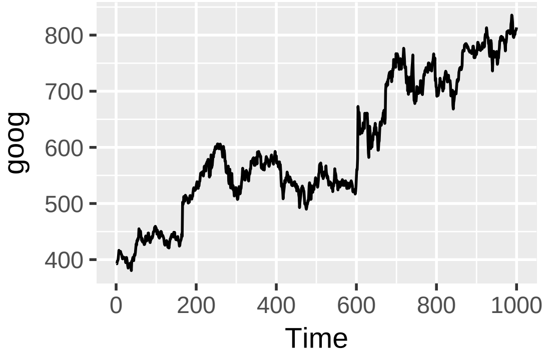
S3 Reject H_0 if $BG_{calc} > BG_{crit}$

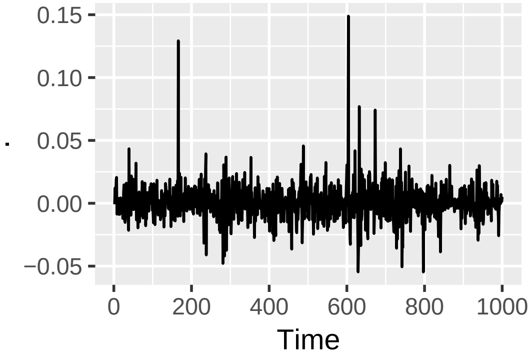
Note : 1. Not require strict exogenous
2. Can test higher order autocorrelation
3. Also an asymptotic test

```
F test?
```

- **Example**

$$\Delta GOOG_t = \log(GOOG_t) - \log(GOOG_{t-1}) \quad \text{for} \quad t = 2, 3, \ldots, 1000$$





$$\Delta GOOG_t = \beta_0 + \beta_1 \Delta GOOG_{t-1} + \beta_2 \Delta GOOG_{t-2} + u_t$$

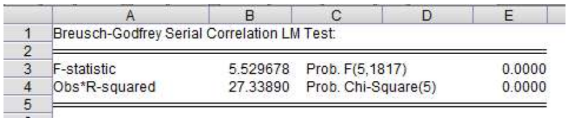
$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \rho_4 u_{t-4} + \rho_5 u_{t-5} + e_t$$

$$H_0 : \rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho_5 = 0$$

```
##
##  Breusch-Godfrey test for serial correlation of order up to 5
##
## data:  .
## LM test = 4.4929, df = 5, p-value = 0.4808
```

```
##
##  Breusch-Godfrey test for serial correlation of order up to 5
##
## data:  .
## LM test = 0.89541, df1 = 5, df2 = 989, p-value = 0.4834
```

Output from Eviews



	A	B	C	D	E
1	Breusch-Godfrey Serial Correlation LM Test:				
2					
3	F-statistic	5.529678	Prob. F(5,1817)		0.0000
4	Obs*R-squared	27.33890	Prob. Chi-Square(5)		0.0000
5					

- **Correcting for autocorrelation**
  1. Using HAC error (heteroskedasticity and autocorrelatuon consistent), also called serial correlation-robust standard error
  2. Changing the specification of our model (e.g. adding additional lag of the dependent variable)
  3. Feasible generalized least squares (FGLS) estimator - Making an assumption about the precise nature of the autocorrelation.
- **HAC error**

$$Var(\hat{\beta}|\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}[\mathbf{X}'Var(\mathbf{u}|\mathbf{x})\mathbf{X}](\mathbf{X}'\mathbf{X})^{-1}$$

With Homo

$$Var(\mathbf{u}|\mathbf{X}) = \sigma^2\mathbf{I_n} \Rightarrow Var(\hat{\beta}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

With HTSK

$$Var(\hat{\beta}|\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\left[\mathbf{X}'\begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}\mathbf{X}\right](\mathbf{X}'\mathbf{X})^{-1}$$

With HTSK and serial correlation

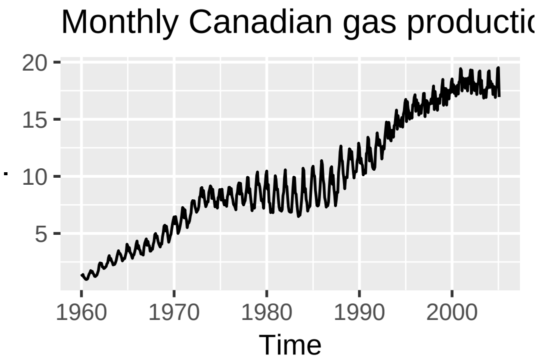
$$Var(\hat{\beta}|\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\left[\mathbf{X}'n\mathbf{\Lambda}_n\mathbf{X}\right](\mathbf{X}'\mathbf{X})^{-1}$$

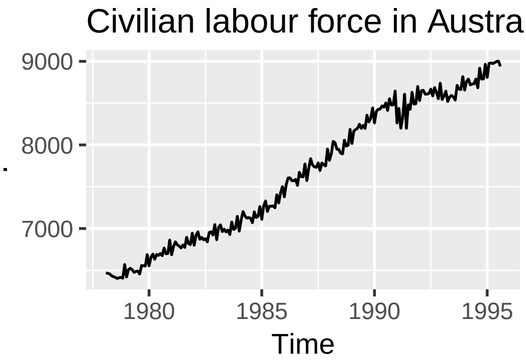
$\mathbf{\Lambda}_n$ is an estimator of the long run variance of $\mathbf{u}$

Week 10

- **Persistence**

Economic and Financial time series typically display **autoregressive behavior**. That is, the value of the time series in the current period is correlated with past values of itself.
e.g. Habit persistence, Institutional arrangements





When working with timeseries data, by collecting a random sample of observations, $(y_1, y_2. \cdots, y_T)$, we are making random draws from each of there T probability distributions. Specific features of these distributions are of interest to us (eg. mwan, variance etc)

If we impose no restrictions on the sequence of these random variable, then we will have T means, $(\mu_1, \mu_2, \cdots, \mu_T)$, T variances, $(\sigma_1^2, \sigma_2^2, \cdots, \sigma_T^2)$ and $T(T-1)/2$ covariances to estimate, which leads to the necessity of introducing **Stationary Time Series**.

- **Stationary Time Series**

A univariate time series is an ordered squance of random variables indexed by time. (Infinite number of realizations) $\{y_t : t = \cdots - 2, -1, 0, 1, 2, \cdots\}$

Weakly Stationary (covariance stationary, second-order stationary)

a) $E(y_i) = \mu < \infty \quad for\ all\ t$

b) $Var(y_t) = E[(y_t - \mu)^2] = \gamma_0 < \infty \quad for\ t$

c) $Cov(y_t) = E[(y_t - \mu)(y_{t-j} - \mu)] = \gamma_j < \infty$

Its first and second moments are both finite and time invariant.
The covariance depends only on the time interval separating them and not on time itself)

- **White Noise**

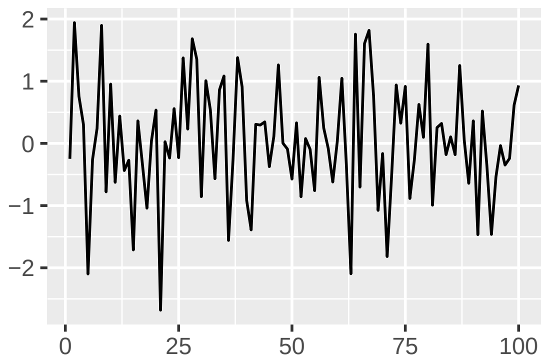
$e \sim WN(0, \sigma^2)$ if

a) $E(e_t) = 0 \forall t$

b) $Var(e_t) = \sigma \forall t$

c) $Cov(e_t, e_{t-j}) = 0 \forall j \neq 0$ (no linear relationship)

If it's also normally distributed Gaussian white noise



- **I.I.D. Independentlly and Identically Distributed**

$e \sim i.i.d(0, \sigma^2)$ if

a) $E(e_t) = 0 \forall t$

b) $Var(e_t) = \sigma \forall t$

c) e_t and e_{t-j} and independent random variables $\forall j$ and $\forall t$ (stronger, norelationship, either linear or nonlinear)

Define $E_t(y_{t+j}) = E(y_{t+j} | y_t, y_{t-1}, y_{t-2}, \cdots)$

Under independent assumption $E_t(y_{t+j}) = E(y_{t+j}) \forall j \geq 1$

- **Lag**

y_{t-j} jth lag of y, the value of y in time period $t - j$ (the value of j periods earlier)

The change in the value of y in period t - the first difference of y - Δy_t - $\Delta y_t = y_t - y_{t-1}$ - $\Delta y_{t-1} = y_{t-1} - y_{t-2}$ - $\Delta y_{t-j} = y_{t-j} - y_{t-j-1}$

Each time we lag a time series one period we lose an observation.

Each time we difference a time series we lose an observation.

- **The stationary autoregresssive model**

$$\text{AR(p) Model: } y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + u_t \qquad \text{where } u_t \sim WN(0, \sigma^2)$$

linear regression model in which the dependent variable is y_t and the repressors are lags of y_t

In its general form this model has p lags of y_t on the right_hand side, and hence we call this a pth-order autoregression or AR(p) model.

AR(1):

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + u_t$$

AR(2):

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + u_t$$

- **The AR(1) model**

$$|\varphi_1| < 1$$

P1 Mean of AR(1) model

$$E(y_t) = \frac{\varphi_0}{1-\varphi_1}$$

P2 Varinance of AR(1) model

$$Var(y_t) = \frac{\sigma^2}{1-\varphi_1^2}$$

P3 Autocovariance and autocorrelation of AR(1) model

$$Cov(y_t, y_{t-j}) = \gamma_j = \frac{\sigma^2}{1-\varphi_1^2}\varphi_1^j,\ \forall j \in \mathbb{N}$$

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \varphi_1^j,\ \forall j \in \mathbb{N}$$

For a stationary AR(1) process the $Corr(y_t, t_{t-h}) \to 0$ as $h \to \infty$

$$\{\rho_1, \rho_2, \rho_3, \cdots\} = \{\varphi_1, \varphi_1^2, \varphi_1^3, \cdots\}$$

The speed is determined by $|\varphi_1|$

- For $0 < \varphi_1 < 1$, they decay monotonically

- For $-1 < \varphi_1 < 0$, they oscilate
- **The AR(2) model**

$$y_t = \varphi_0 + \varphi_1 y_t - 1 + \varphi_2 y_t - 2 + u_t \quad \text{where } u_t \sim WN(0, \sigma^2)$$

P1 Mean of AR(2) model

$$E(y_t) = \frac{\varphi_0}{1-\varphi_1-\varphi_2}$$

P2 The correlation function of AR(2) model

$$\rho_1 = \frac{\varphi_1}{1-\varphi_2},\ and\ \rho_k = \varphi_1\rho_{k-1} + \varphi_2\rho_{k-2},\ k \geq 2$$

- **Random walk**

$$y_t = y_{t-1} + u_t \text{ where } u_t \sim WN(0, \sigma^2)$$

A random walk (or unit root non-stationary time series) is a covariance non-stationary model.

Strong memory.

Assume that the starting value of the random walk is at $t = 0$ where $y_0 = 0$.

$$y_t = y_{t-1} + u_t = y_{t-2} + u_{t-1} + u_t = u_1 + u_2 + \cdots + u_{t-1} + u_t$$

Mean:

$$E(y_t) = E(u_1 + u_2 + \cdots + u_{t-1} + u_t) = 0$$

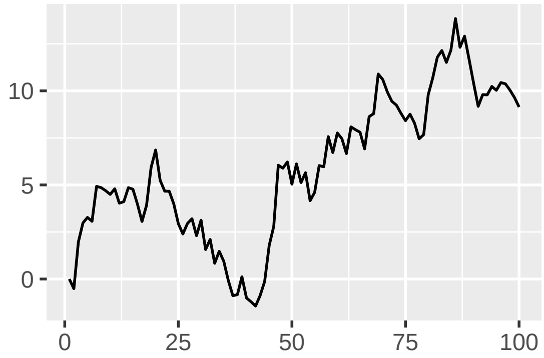
Variance:

$$\begin{aligned} Var(y_t) &= E(u_1 + u_2 + \cdots + u_{t-1} + u_t)^2 \\ &= E(u_1)^2 + E(u_2)^2 + \cdots + E(u_t)^2 \\ &= \sigma^2 + \sigma^2 + \cdots + \sigma^2 = t\sigma^2 \end{aligned}$$

- **Properties of a unit root model**

1. A unit root process is not predictable

2. A unit root process is not mean-reverting (i.e. over time it drifts away from the mean)



- **Random walk with a drift**

$$y_t = \varphi_0 + y_{t-1} + u_t$$

$$y_t = t\varphi_0 + y_0 + u_t + \cdots + u_2 + u_1$$

a time trend ($\varphi_0 t$)
a pure random walk ($u_t + \cdots + u_2 + u_1$)

- **Unit root testing**

$$y_t = \varphi_1 y_{t-1} + u_t$$

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + u_t$$

where $e_t \sim IID(0, \sigma^2)$

$$\begin{aligned} H_0(\text{random walk}) &: \varphi_1 = 1 \\ H_1(\text{stationary}) &: |\varphi_1| < 1 \end{aligned}$$

- **Dicket-Fuller test**

$$DF = \frac{\hat{\varphi}_1 - 1}{se(\hat{\varphi}_1)}$$

Using both forms, the DF statistic under the null has non-standard distributions.

Using the unit root with a drift model and when $\varphi_0 \neq 0$ and $\varphi_1 = 1$ then DF is asymptotically normal under the null.

The critical values required form testing were obtained by Phillips (1987), Fuller (1976) by simulation.

An augmented version of this statistic (ADF) assumes the same t-ratio as before but is based on the following model specification for y_t :

$$y_t = \varphi_0 + \beta t + \varphi_1 y_{t-1} + \sum_{i=1}^{p-1} \gamma_i \Delta y_{t-i} + u_t$$

* where: 1. $c_t = \varphi_0 + \beta t$ is a deterministic trend, and
2. $\Delta y_j = y_j - y_{j-1}$ is the differenced series of y_j

- **Information criteria**

Minimizing

$$I(k) = c \quad + \underset{decreasing\ in\ k}{\ln\left[SSR(k)\right]} + \underset{increasing\ in\ k}{(k+1)\frac{\alpha}{T}}$$

e.g. Akaike's information criterion (AIC): $\alpha = 2$
Schwarz-Bayes information criterion (BIC or SIC): $\alpha = \ln(T)$ ($\frac{\ln T}{T} > \frac{2}{T}$ for $T > 8$, penalizes additional lags more severely)

Note: When using lag length selections criteria to choose between the two methods, **both** models should be estimated using the data with the same number of observations. Otherwise the model with more observations will be advantaged. (be aware of the cases where *extra lag means less observation*)

Week 11

- **Consistency**

Sometimes the random sampling assumption is not satisfied ($E(\mathbf{u}|\mathbf{X}) \neq 0$) but errors are uncorrelated with regressors ($E(u_i|x_{i1}, x_{i2}, \cdots, x_{ik}) = 0$) e.g. time series data.

OLS will not be unbiased, but consistant.

- **Asymptotic Normality**

Sometimes errors are not normally distributed.

The distribution of the OLS estimator will be approximatedly Normal in large samples.

- **The Law of Large Numbers**

Sample averages converge to population means as $n \to \infty$

Reasoning : A single variable y with $E(y) = \mu$ and $Var(y) = \sigma^2$

Taking n observations, sample average $\bar{y}$, $E(\bar{y}) = \mu$ and $Var(\bar{y}) = \sigma^2/n$

As $n \to \infty$, $Var(\bar{y}) \to 0$, the chance of $\bar{y}$ being anything other than μ goes to 0.

Proof for $Var(\bar{y}) = \sigma^2/n$:

$$\begin{aligned} Var(\bar{y}) &= Var(\frac{1}{n}\sum_{i=1}^{n} y_i) = \frac{1}{n^2} Var(\sum_{i=1}^{n} y_i) \\ &\overset{i.i.d.}{=} \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n} \end{aligned}$$

- Convergence

$\bar{y}$ converges in probability to μ i.e. $p\lim(\bar{y}) = \mu$ or $\bar{y} \xrightarrow{p} \mu$

If an estimator converges in probability to the population parameter that it estimates, we say that the estimator is **consistent**.

The sample mean is a consistent estimator of the population mean (LLN).

- $E(\bar{y}^2) \neq \mu^2$, but $p\lim(\bar{y}^2) = \mu^2$
- $E(\frac{1}{\bar{y}}) =?$, but $p\lim(\frac{1}{\bar{y}}) = \frac{1}{\mu}$ provided $\mu \neq 0$
- $E(\frac{1}{\hat{\sigma}_y^2}) =?$, but $p\lim(\frac{1}{\hat{\sigma}_y^2}) = \frac{1}{Var(y)}$

(go through non-linear combinations and smooth functions as well)

Conclusion: Even if we have non-random sample, i.e. $E(\mathbf{u}|\mathbf{X}) \neq 0$, as long as $E(u_i) = 0$ and u_i is uncorrelated with x_{i1} to x_{ik} , then the OLS estimator is consistent.

- **Central Limit Theorem**

The sum (or the average) of n Normal random variables is Normal.

However,

- The sum (or the average) of n Uniform random variables is Not Uniform.
- The sum (or the average) of n F random variables is Not F.
- The sum (or the average) of n Bernouli random variables (1 with probability p and 0 with probability 1-p) is Not Bernouli.

If n is large,

- the sum (or the average) of n Uniform / F / Bernouli random variables is approximately Normal.
- the sum (or the average) of n random variables from any distribution with a finite variance is approximately Normal.

Conclusion: For any error distribution, as long as the sample size is large, the OLS estimator is approximately Normal, and we can base our statistical inference on the usual t and F tests (allow to use OLS even if the distribution of the dependent variable is far from Normal, and may even have spikes on some values, such as the spike at zero for the married women's work hours).