EC3304 Econometrics II

Lingjie, November 26, 2020

Probability, statistics and linear regression

Econometrics is about statistical modelling and data analysis for economic usages, concerning theory, methodology and empirics of analysing data
Economists are interested in casual effects and forecasts

Random Variables

types	: continuous, discrete
event	: $(Y=a),(Y<a),(b<Y<a)$
distribution	: f_X, F_X
summary stat	: $E(Y), Var(Y)$
joint distribution	: $(Y \in A, X \in B)$
marginal distribution	: $(Y \in A, X \in x)$
identical distribution	: $F_X = F_Y$
independence	: $P(Y \in A \mid X \in B) = P(Y \in A)$
uncorrelated	: $Cov(Y,X)=0$
random sample $(i.i.d)$	: $(F_{Y_i}, F_{X_i}) = (F_{Y_j}, F_{X_j})$
	: $P(Y_i, X_i \mid Y_j, X_j) = P(Y_i, X_i)$
	: $E(Y_i) = E(Y_j)$
	: $Var(Y_i) = Var(Y_j)$

Inference

basic principle

1. specific statistical model with parameter of interest
2. assume model accuracy
3. use data to learn about parameter

Set estimate: Sampling distribution

Exact sampling distribution
$X \sim Ber(p)$, when $n=2$,
$P(X=0), P(X=1/2), P(X=1)$ etc and plot the prob against value of sample average

Finite sample properties

Unbiasedness

$E(\hat{\theta}) = \sum_{t \in T} tP(\hat{\theta}=t) = \theta$

Efficiency

Evaluate efficiency by loss function, such as MSE

$$\begin{aligned} MSE = &\ E(\hat{\theta}-\theta)^2 = E(\hat{\theta}-E(\hat{\theta})+E(\hat{\theta})-\theta)^2 \\ = &\ Var(\hat{\theta}) + (E(\hat{\theta})-\theta)^2 \end{aligned}$$

Large sample properties

Consistency

$\hat{\theta}$ is consistent estimator if
$\lim_{n\to\infty} P(|\hat{\theta}-\theta| \le c) = 1\ \forall\ c>0$
$\hat{\theta} \to \theta$ converge in probability to θ

Sample average vs Expectation

average: $\frac{\sum_{i\in n} X_i}{n}$, Expectation: $\sum_{x\in S} xP(X=x)$

Consistency and LLN

convenient way to prove consistency
if $\{X_i\}_{i=1}^n$ is $i.i.d$ with $Var(X_i) < \infty$
then $\bar{X} \to E(X_i)$ converge in probability
since $E(\bar{X}) = E(X_i)$ and
$Var(\bar{X}) = \frac{Var(X_i)}{n} \to 0$ as $n \to \infty$

Asymptotic Normality and CLT

when n is large but not infinite, standardised sample average is approximated by $N(0,1)$
if $\{X_i\}_{i=1}^n$ is $i.i.d$ with $Var(X_i) < \infty$
then $\frac{\bar{X}-E(\bar{X})}{\sqrt{Var(\bar{X})}} \approx N(0,1)$ for large n

Confidence interval

by CLT, $P(\frac{\hat{X}-X}{\hat{\sigma}} \le c) \approx \Phi(c)$
CI: $(\hat{X} \pm \sigma_{\hat{X}} Z_{\{(1-\alpha)/2\}})$, $Z_{0.025} = 1.96$ for 95% CI

Hypothesis test

test $H_0: X=0$ vs $H_1: X \neq 0$
reject H_0 if $|\frac{\hat{X}-X}{\hat{\sigma}}| > 1.96$ or $2\Phi(-|\frac{\hat{X}-X}{\hat{\sigma}}|) < 5\%$

Linear Regression

including $\{Y, log(Y)\} \times \{X, log(X), X^k\}$

Properties of OLS Estimators

model: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$
interpretation of β_1: $\frac{d}{dX} E(Y \mid X_1 = x_1)$

Assumptions (LSA)

model: $Y_i = \beta_0 + \beta_1 X_{2i} + \beta_2 X_{2i} + u_i$

1. $\{(Y_i, X_{1i}, \cdots, X_{2i})\}_{i=1}^n$ is i.i.d
2. $E(u_i \mid X) = 0 \Leftrightarrow E(Y_i \mid X) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$
3. finite $E(Y_i^4), E(X_i^4)$, no outliers
4. No perfect multicollinearity (only essential for multi-linear models)

Least Squares

$E(Y \mid X) = \arg\min_{m(\cdot)\in\mathbb{M}} E(Y - m(X))^2$
For (β_0, β_1): $\min_{\beta_0,\beta_1} E(Y-\beta_0-\beta_1 X)^2$

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

$\hat{\beta}_1 = \frac{\hat{Cov}(X_i,Y_i)}{\hat{Var}(X_i)} = \frac{E[X-E(X)]-E[Y-E(Y)]}{E[X-E(X)]^2}$

Unbiasedness & Large Sample Properties

$\hat{\beta}_1 = \frac{\hat{Cov}(X_i, \beta_0+\beta_1 X_i + u_i)}{\hat{Var}(X_i)} = \beta_1 + \frac{\hat{Cov}(X_i,u_i)}{\hat{Var}(X_i)}$
by CLT:

if LSA (2) true,	$E(u_i \mid X_i) = 0 \Rightarrow Cov(X_i,u_i) = 0$
	$\hat{\beta}_1 \to \beta_1$
if large but finite n,	$\hat{\beta}_1 \approx N(\beta_1, \sigma^2_{\hat{\beta}_1})$
	$\sigma^2_{\hat{\beta}_1} = \frac{1}{n}\frac{Var((X_i-\mu_X)u_i)}{[Var(X_i)]^2}$

if LSA(2) failed, then $\hat{\beta}_1 \to \beta_1 + \frac{Cov(X_i,u_i)}{Var(X_i)} \neq \beta_1$

Homoskedasticity

heteroskedastic: $Var(u_i \mid X_i) = \sigma^2_{u_i}(X_i)$, fn of X
homoskedastic: $Var(u_i \mid X_i) = \sigma^2_{u_i}$, constant

if LSA holds, by CLT: $\hat{\beta}_1 \approx N(\beta_1, \sigma^2_{\hat{\beta}_1})$

heteroskedastic $\sigma^2_{\hat{\beta}_1} = \frac{1}{n}\frac{Var((X_i-\mu_X)u_i)}{(Var(X_i))^2}$
homoskedastic $\sigma^2_{\hat{\beta}_1} = \frac{1}{n}\frac{Var(u_i)}{Var(X_i)}$

Gauss Markov

in addition to LSA, assume homoskedasticity, then OLS is **B**est **L**inear **U**nbiased **E**stimator

Perfect Multicollinearity

$$X = \begin{pmatrix} 1 & X_{11} & X_{k1} \\ \vdots & \vdots & \vdots \\ 1 & X_{1n} & X_{kn} \end{pmatrix}, \text{ PM when } Rank(X) \neq k$$

when little or no variation in $\{X_i\}_{i=1}^n$
$\sigma^2_{\hat{\beta}_1} = \frac{1}{n}(\frac{1}{1-corr(X_{1i},X_{2i})^2})\frac{Var(u_i)}{Var(X_i)}$ is large

Dummy variable trap

includes all one hot encoded dummies and not drop one dummy or β_0

Imperfect multicollinearity

covariates have high correlation but not perfectly collinear

R^2

keywords: Explained/Total Sum of Squares, Sum of Squares Errors, adjusted R^2

$$R^2 = 1 - \frac{SSR}{TSS} \qquad \text{ESS} = \sum(\hat{Y}_i - E(Y))^2$$
$$\bar{R}^2 = 1 - (\frac{n-1}{n-k-1})\frac{SSR}{TSS} \qquad \text{SSR} = \sum(Y_i - \hat{Y}_i)^2$$
$$\text{TSS} = \sum(Y_i - E(Y))^2$$

Omitted Variable Bias

let $Y = \beta_0 + \beta_1 X + \beta_2 Z + u_i$
OVB arise when $Cov(X,Z) \neq 0$ and $Cov(Y,Z) \neq 0$
but we assume $Y = \beta_0 + \beta_1 X + \epsilon_i$
then $\hat{\beta}_1 = \beta_1 + \beta_2 \frac{Cov(X,Z)}{Var(X)} \neq \beta_1$

Panel Data (Longitudinal data)

useful to control for time invariant factors (OVB)
balanced panel when all variables are observed for all entities and all time periods
unbalanced panel when missing observations are present, not a worry if missing mechanism is exogenous

Data Structure

- Repeated cross-sections: different groups, heterogeneous data and more is good
- Panel data: same group over time, correlation between entity (or cluster) over time
- Time series: same entity over time, correlation between obs will be key feature

Least Squares Assumptions (LSA-FE)

model: $Y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_i + u_{it}$
A time invariant covariates α_i that is not able to be observe through data

1. $\{(Y_i, X_{i1}, \cdots, X_{iT})\}_{i=1}^n$ or $\{(\alpha_i, X_{i1}, \cdots, X_{iT}, u_{it}, \cdots, u_{iT})\}_{i=1}^n$ is i.i.d
2. $E(u_i \mid X_{i1}, \cdots, X_{iT}, \alpha_i) = 0$
3. finite $E(Y_i^4), E(X_i^4)$, no outliers
4. No perfect multicollinearity (only essential for multi-linear models)

Random Effects (RE) Model

RE model assumes $E(\alpha_i \mid X_{i1}, \cdots, X_{iT}) = 0$,
then $\frac{\delta}{\delta X_{it}} E(Y_{it} \mid X_{i1}, \cdots, X_{iT}) = \beta_1$
we can run OLS straight since α_i will not affect our result.
However, since composite error $\epsilon_{it} = \alpha_i + u_{it}$ is correlated across i and possibly t
$Var(\epsilon_{it}) = \sigma_\alpha^2 + \sigma_u^2$ and
$Cov(\epsilon_{it}, \epsilon_{is}) = Cov(\alpha_i, \alpha_i) = \sigma_\alpha^2$
assuming $\{u_{it}\}$ to be uncorrelated. Generalised least squares (GLS) is needed.

Fixed Effects (FE) Model

FE model assumes $E(\alpha_i \mid X_{i1}, \cdots, X_{iT}) \neq 0$,
then $\frac{\delta}{\delta X_{it}} E(Y_{it} \mid X_{i1}, \cdots, X_{iT}, \alpha_i) = \beta_1$
therefore we can't run OLS straight since α_i is not observed

n-1 binary regressor regression model

$$\text{Since} \quad \begin{aligned} Y_{1,t} &= (\beta_0 + \beta_2 Z_1) + \beta_1 X_{1,t} + u_{1,t} \\ Y_{2,t} &= (\beta_0 + \beta_2 Z_2) + \beta_1 X_{2,t} + u_{2,t} \\ Y_{3,t} &= (\beta_0 + \beta_2 Z_3) + \beta_1 X_{3,t} + u_{3,t} \end{aligned}$$

$Y_{i,t} = \beta_0 + \beta_1 X_{i,t} + \gamma_{Z_1} DZ_1 + \gamma_{Z_2} DZ_2 + u_{it}$
where $DZ_1 = 1$ for Z_1 and $DZ_2 = 1$ for Z_2 are time dummy variables. DZ_3 is omitted to avoid perfect multicollinearity

Fixed Effect regression model

another way to represent omitted time invariant factors
$Y_{it} = \beta_0 + \beta_1 X_{it} + \alpha_i + u_{it}$
where $\alpha_1 = \beta_0$ and $\alpha_i = \beta_0 + \gamma_i$ for $i > 1$, when related to n-1 binary regressor

First Differencing estimator, FD

$$\begin{aligned} \text{model} &: \Delta Y_{it} = \beta_1 \Delta X_{it} + \Delta u_{it} \\ \text{method} &: Y_{i(t+1)} - Y_{i(t)} \ \forall\, t \in [1, T] \\ \text{since} &: E(\Delta u_{it} \mid \Delta X_{i2}, \cdots, \Delta X_{iT}) = 0 \ \forall\, t > 1 \\ &: \Rightarrow E(\Delta Y_{it} \mid \Delta X_{it}) = \beta_1 \Delta X_{it} \\ \text{OLS} &: \min Q_{FD}(b_1) = \textstyle\sum_{t=2}^T \sum_{i=1}^n (\Delta Y_{it} - b_1 \Delta X_{it})^2 \end{aligned}$$

Fixed Effects estimator FE

$$\begin{aligned} \text{model} &: \tilde{Y}_{it} = \beta_1 \tilde{X}_{it} + \tilde{u}_{it} \\ \text{method} &: \tilde{Y}_{it} = Y_{it} - \bar{Y}_i \\ \text{since} &: E(\tilde{u}_{it} \mid \tilde{X}_{i2}, \cdots, \tilde{X}_{iT}) = 0 \ \forall\, t > 1 \\ &: \Rightarrow E(\tilde{Y}_{it} \mid \tilde{X}_{it}) = \beta_1 \tilde{X}_{it} \\ \text{OLS} &: \min Q_{FE}(b_1) = \textstyle\sum_{t=2}^T \sum_{i=1}^n (\tilde{Y}_{it} - b_1 \tilde{X}_{it})^2 \end{aligned}$$

Regression Errors

$$\begin{aligned} \text{RE} &: \epsilon_{it} &&= \alpha_i + u_{it} \\ \text{FE(FD)} &: \Delta u_{it} &&= u_{it} - u_{it-1} \\ \text{FE(FE)} &: \tilde{u}_{it} &&= u_{it} - \bar{u}_i \end{aligned}$$

note how even if $\{u_{it}\}$ is i.i.d., regression errors can still be correlated

Clustered Variance

To allow the correlation in regression errors, we consider the following sample average

$$\begin{aligned} \text{RE} &= \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \epsilon_{it} \\ &= \frac{1}{n} \left(\frac{1}{T} \sum_{t=1}^T \epsilon_{it} \right) = \frac{1}{n} \sum_{i=1}^n \bar{\epsilon}_i \\ \text{FE(FD)} &= \frac{1}{n(T-1)} \sum_{i=1}^n \sum_{t=2}^T \Delta u_{it} \\ &= \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{T-1} \sum_{t=2}^T \Delta u_{it} \right) \\ \text{FE(FE)} &= \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \tilde{u}_{it} \\ &= \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{T} \sum_{t=1}^T \tilde{u}_{it} \right) \end{aligned}$$

Binary dependent variables

Let $Y \sim Ber(p)$, $E(Y \mid X) = P(Y = 1 \mid X)$

linear probability model	: $P(Y = 1 \mid X) = \beta_0 + \beta_1 X$
generalised linear model	: $P(Y = 1 \mid X) = F(\beta_0 + \beta_1 X)$

Binary Choice Assumptions (BCA)

1. **latent variable**
   $Y_i^* = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_{ki} + u_i$
   s.t. $-u_i \mid X_{1i}, \cdots, X_{ki}$ has known cdf $F(\cdot)$
2. Y_i^* is not observed but **limited dependent variable** $Y_i = 1_{\{Y_i^* \geq 0\}}$ is observed
3. $\{(Y_i, X_{1i}, \cdots, X_{ki})\}_{i=1}^n$ is i.i.d. across i
4. no perfect multicollinearity + no outliers

the latent variable can be interpreted as utility that determines the outcome of decision. e.g. if utility gain in smoking > 0.5 then individual smoke

Linear Probability Model (LPM)

model:	$Y_i = \beta_0 + \beta_1 X_i + u_i$
where	$Y_i \sim Ber(p)$ and $E(u_i \mid X_i) = 0$
then	$E(Y_i \mid X_i) = P(Y_i = 1 \mid X_i)$
β_1	$= \frac{\partial}{\partial x} P(Y_i = 1 \mid X_i = x)$

and $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ is the predicted probability that $Y_i = 1$ given X_i

Remarks about LPM

1. if $Y_i \sim Ber(p)$, heteroskedastic robust SE must be use
   $Var(Y_i \mid X_i) = \theta(1 - \theta)$
   $= P(Y_i = 1 \mid X_i) - (1 - P(Y_i = 0 \mid X_i))$
   $= \beta_0 + \beta_1 X_i$, depends on x
2. LPM is not suitable to use as probabilities fall outside $[0, 1]$ and $\frac{\partial}{\partial X}$ is constant

Probit Model

$P(Y = 1 \mid X) = \Phi(\beta_0 + \beta_1 X + \beta_2 Z)$
where $\beta_0 + \beta_1 X + \beta_2 Z$ is called index
and $\Phi(z) = P(N(0, 1) \leq z)$
we can use index as z-value to compute predicted probabilities for any x

Partial effect

$\frac{d}{dX} \hat{P}(Y = 1 \mid X) = \hat{\beta}_1 \Phi'(\hat{\beta}_0 + \hat{\beta}_1 X)$

Percentage point

$100 \times (\hat{P}(Y = 1 \mid X = x') - \hat{P}(Y = 1 \mid X = x))$

Percentage change

$\frac{\hat{P}(Y=1|X=x') - \hat{P}(Y=1|X=x)}{\hat{P}(Y=1|X=x)}$

Logit Model

$P(Y = 1 \mid X) = \frac{1}{1+e^{-(\beta_0+\beta_1 X)}}$
where Logit function: $F(z) = (1 + e^{-z})^{-1}$

Comparing Probit and Logit Estimates

- Probit and Logit are qualitatively the same
- but dont compare magnitude of parameter
- test should done on individual model
- instead look at predicted probabilities
- measure of model fit by Pseudo-R^2

Maximum Likelihood Estimation

likelihood of sample $\{Y_i\}_{i=1}^n$

w/o covariates	: $L(\theta) = \theta^{\sum_{i=1}^n Y_i}(1 - \theta)^{n - \sum_{i=1}^n Y_i}$
w covariates	: $L(\beta_0, \beta_1)$

$= F(\beta_0 + \beta_1 x_i 1)^{\sum_{i=1}^n Y_i}(1 - F(\beta_0 + \beta_1 x_i 1))^{n - \sum_{i=1}^n Y_i}$

Pseudo-R^2

pseudo-$R^2 = 1 - \frac{\ell_{(\text{Probit } or \text{ Logit})}}{\ell_{\text{bernoulli}}}$
where $\ell_{\text{bernoulli}}$ is max log-MLE without covariates
where $\ell_{\text{probit/logit}}$ is max log-MLE with covariates

STATA command

set pannel data	*xtset panel time*
FE with cluster errors	*xtreg Y X, fe vce(cluster panel)*
probit	*probit Y X, r*
logit	*logic Y X, r*

Instrumental Variable

General solution to endogeneity problem caused by omitted variable bias **(OVS)**
Satisfying conditions

1. Exclusion, Z_i is different to X_i
2. Relevance, Partial $Corr(Z_i, X_i) \neq 0$
3. Exogeneity, $Cov(Z_i, u_i) = 0$

Since $0 = Cov(Z_i, u_i) \Rightarrow \beta_{1,IV} = \frac{Cov(Z_i, Y_i)}{Cov(Z_i, X_i)}$
Contrast to $\beta_{1,OLS} = \frac{Cov(X_i, Y_i)}{Var(X_i)}$

$\beta_{0,OLS} = \bar{Y} - \bar{X}\beta_{0,OLS}$
$\beta_{0,IV} = \bar{Y} - \bar{X}\beta_{0,IV}$

$\hat{\beta}_{1,IV} = \frac{\hat{C}ov(Z_i, Y_i)}{\hat{C}ov(Z_i, X_i)} = \beta_1 + \frac{\hat{C}ov(Z_i, u_i)}{\hat{C}ov(Z_i, X_i)}$

IV vs OLS

Both OLS and IV estimators are biased in finite sample
$\hat{\beta}_{1,OLS} \to \beta_1 + Corr(X_i, u_i) \frac{\sigma_u}{\sigma_X}$
$\hat{\beta}_{1,IV} \to \beta_1 + \frac{Corr(Z_i, u_i)}{Corr(Z_i, X_i)} \frac{\sigma_u}{\sigma_X}$

weak instrument: $Corr(Z_i, X_i)$ is small, then bias will be huge. F-stat < 10 is consider weak IV

Partial correlation, Structural equation, Reduced form equation

structural:	$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_{1i} + \cdots + \beta_{r+1} W_{ri} + u_i$
reduced form:	$X_i = X_i^* + v_i$
where	$X_i^* = \pi_0 + \pi_1 W_{1i} + \cdots + \pi_r W_{ri} + \pi_{r+1} Z_i$

Structural equation

- Meaningful parameters (causal)
- regressors may be endogenous
- consistently estimate by TSLS

Reduced form equation

- Parameters are correlations
- regressors must be exogenous
- consistently estimate by OLS

Threats to internal validity

exogenous variable: variable uncorrelated with u_i
endogenous variable: variable correlated with u_i

- Omitted variable that is relevant and correlated to the observed regressor
- Measurement errors
- Simultaneous causality

Measurement Errors

$X_i = X_i^* + v_i$, Observed data is actual data + error

true model:	$Y_i = \beta_0 + \beta_1 X_i^* + \epsilon_i$
observed model:	$Y_i = \beta_0 + \beta_1 X_i + u_i$
where	$u_i = \epsilon_i - \beta_1 v_i$

$\Rightarrow Cov(u_i, X_i) \neq 0$

Simultaneous Equations

$$\begin{aligned} Y_{1i} &= \alpha_0 + \alpha_1 Y_{2i} + u_{1i} \\ Y_{2i} &= \beta_0 + \beta_1 Y_{2i} + u_{2i} \end{aligned}$$

$\Rightarrow Cov(u_{1i}, Y_{2i}) \neq 0, Cov(u_{2i}, Y_{1i}) \neq 0$
Supply shifter: does not directly affect demand equation other than through the equilibrium price

Consistency of IV estimator

1. $\{(Y_i, X_i, Z_i)\}_{i=1}^n$ is i.i.d.
2. $Y_i = \beta_0 + \beta_1 X_i + u_i$, such that $E(u_i) = 0$
3. $E(Y_i^4), E(X_i^4), E(Z_i^4)$ are finite
4. Z_i is a valid instrument

Then IV estimator is consistent (while biased in finite sample) and has asymptotic normal distribution

OLS estimator will be biased and inconsistent

Testing for relevance

$X_i = \pi_0 + \pi_1 Z_i + \pi_2 W_i + v_i$
Since $\hat{\pi}_1 \to \frac{Cov(Z_i, X_i)}{Var(Z_i)} = Corr(Z_i, X_i) Var(X_i)$
Then we just have to test $H_0 : \pi_1 = 0$

Two Stage Least Squares (TSLS)

Stages:

1. OLS on the reduced form equation
   Regress X_i on $W_{1i}, \cdots, W_{ri}, Z_{1i}, Z_{2i}$
2. OLS on the structural equation with $\hat{X}_i$
   Regress Y_i on $\hat{X}_i, W_{1i}, \cdots, W_{ri}$

where $\hat{X}_i$ is the fitted value from Stage 1

Time Series

Definitions

time series (process)	:	data collected on the same observational unit at multiple time periods
forecasting	:	focus on model fit and external validity
j^{th} lag	:	Y_{t-j}
first difference	:	$\Delta Y_t = Y_t - Y_{t-1}$
first autocovariance	:	$Cov(Y_t, Y_{t-1})$
first autocorrelation	:	$Corr(Y_t, Y_{t-1})$
	:	$\rho_1 = \frac{Cov(Y_t, Y_{t-1})}{\sqrt{Var(Y_t) Var(Y_{t-1})}}$

Least Squared Assumption for TS

1. $E(u_t | I_t) = 0$
   where $I_t = (Y_{t-1}, Y_{t-2}, \cdots, X_{t-1}, X_{t-2}, \cdots)$
2. $\{(Y_t, X_t)\}_{t=1}^T$ is stationary and weakly dependent process
3. $E(Y_t^4)$ and $E(X_t^4)$ are finite
4. There is no perfect multicollinearity

$\{u_t\}_{t=1}^T$ denote a noise process with mean zero and variance σ_u^2
Stationary and weakly dependent assumptions are required to relax the i.i.d assumption to ensure LLN and CLT can be applied to time series data

Stationary

Condition:

- Y_t is the same for all t
- $\{Y_{t_1}, \cdots, Y_{t_n}\}, \{Y_{t_1+s}, \cdots, Y_{t_n+s}\}$ are the same $\forall\ s, \{t_1, \cdots, t_n\}$

Properties:

- $E(Y_t)$ is the same $\forall\ t$
- $Var(Y_t)$ is the same $\forall\ t$
- $Corr(Y_t, Y_{t+s}) = \rho(s)\ \forall\ t, s$

if $\beta_p < 1\ \forall\ p$ then AR(p) is a stationary process

Weakly Dependent Processes

$\lim_{s \to \infty} Corr(Y_t, Y_{t+s}) = 0$
$\Rightarrow Y_t, Y_{t+s}$ are asymptotically uncorrelated

Roots

$L :=$ lag operators
$LY_t = Y_{t-1}, L^2 Y_t = Y_{t-2}$ etc

Then $Y_t = \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} + u_t$
$\iff (1 - \beta_1 L - \cdots - \beta_p L^p) Y_t = \beta_0 + u_t$
$\Rightarrow (1 - \beta_1 L - \cdots - \beta_p L^p) = (1 - \lambda_1 L) \cdots (1 - \lambda_p L)$

$\{\lambda_i\}_{i=1}^p$ are the roots and stationarity and weak dependence require all $\lambda_i < 1$
unit root is when $\lambda_i = 1$ for any i

Autoregressive Model, $AR(p)$

$Y_t = \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} u_t$
$\hat{\beta}_1 \to \frac{Cov(Y_t, Y_{t-1})}{Var(Y_t)}$

Forcasting

Focus on out-of-sample prediction
$Y_{T+1} = \beta_0 + \beta_1 Y_T + u_{T+1} = Y_{T+1|T} + u_{T+1}$

Accuracy of Forecast: MSFE

mean squared forecast error (MSFE)
$MSFE = E\left[(Y_{T+1} - \hat{Y}_{T+1|T})^2\right]$
95% forecast interval: $(\hat{Y}_{T+1|T} \pm 1.96 RM\hat{S}FE)$
68% forecast interval: $(\hat{Y}_{T+1|T} \pm RM\hat{S}FE)$

Forecasting Errors

(one-period ahead) forecasting error $Y_{T+1} - \hat{Y}_{T+1|T}$

BIC(p), AIC(p)

Helps to decide on optimal period. We can use adjusted R^2 and t/F statistics as well

Bayes information criterion and Akaike information criterion

$$BIC(p) = ln\left[\frac{SSR(p)}{T}\right] + (p+1)\frac{ln(T)}{T}$$

$$AIC(p) = ln\left[\frac{SSR(p)}{T}\right] + (p+1)\frac{2}{T}$$

where T is the sample size and p is the number of lags

ADL(p, q)

Autoregressive distributed lag includes predictors in time series analysis

$$\begin{aligned} Y_t = \quad & \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} \\ & + \delta_1 X_{t-1} + \cdots + \delta_q X_{t-q} + u_t \end{aligned}$$

popular choice is to set $p = q$

Granger Causality test

test $H_0 : \delta_1 = \cdots = \delta_q = 0$

Non-stationary Process

deterministic trend

$Y_t = \alpha + \beta t + u_t$
where $E(Y_t) = \alpha + \beta t$ depends on t

stochastic trend

$$\begin{aligned} Y_t &= \beta_0 + Y_{t-1} + u_t \\ &= \beta_0 t + Y_0 + \textstyle\sum_{s=1}^{T} u_s \end{aligned}$$

where $Var(Y_t) = Var(Y_{t-1}) + \sigma_u^2$ and $Var(Y_t)$ increases in t
drift: presence of β_0

issues with non-stationary process

- Estimators can be biased
- Normal approximation is invalid
- Spurious regression: intuitively unrelated variables may appear highly correlated

Differencing

Case 1: $Y_t = \alpha + \beta t + u_t$

$$\begin{aligned} \Delta Y_t &= Y_t - Y_{t-1} \\ &= (\alpha + \beta t + u_t) - (\alpha + \beta(t-1) + u_{t-1}) \\ &= \beta + u_t - u_{t-1} \end{aligned}$$

$u_t - u_{t-1}$ has a stationary distribution

Case 2: $Y_t = \beta_0 + Y_{t-1} + u_t$

$$\begin{aligned} \Delta Y_t &= Y_t - Y_{t-1} \\ &= \beta_0 + u_t \end{aligned}$$

so ΔY_t is just noise plus a constant