

ECMT2160: ECONOMETRIC ANALYSIS

LINEAR REGRESSION ANALYSIS WITH TIME SERIES DATA

Yiran Jing

Semester 2, 2018

Summary

In this course, we focus on OLS method in time series data and thus, firstly, we must judge 'correct' model based on **Gauss-Markov Assumptions**, which are similar with cross sectional data, but actually more challenge and harder to achieve.

Unlike cross sectional data, time series have some special features: trend, cyclic, seasonality. We only consider simple trend pattern (linear and exponential trend) in this course.

To make model reliable, we emphasis **stationary** and **weak dependency**. As they are important to ensure stable properties of stochastic process over time. Also, CLT and LLN need weak dependence and stationary hold.

For the model part, we study **static model** first, which is not used for forecast, but good at trade-off interpretation, and also the basic idea of hypothesis tests (e.g. serial correlation test) and some advanced model (e.g. error correction model). **Lag distributed model** is another basic model, based on the intuition that time series always have 'lag effect'. For forecast purpose, we study **AR** and **VAR model**.

Given $I(1)$ series, We talk about two potential issues in time series *linear* regression: **Spurious** and **cointegration**. Spurious regression is actually no relationship but we falsely believe they have, while cointegration is related to long run relationship but inequilibrium in short run. From this point, we can realize how important to ensure stationary and weak dependent for modelling. Otherwise, we can easily make mistake and cannot figure out.

we discuss **hypothesis tests** for unit root, serial correlation, cointegration, Granger causality. But actually hypothesis test plays minor role in statistical inference, as they are based on model, but all models are wrong. And statistical significant has no relationship with practical significant.

Contents

Summary

1	Introduction of Time Series Analysis	1
1.1	Time Series Data	1
1.2	Time Series Analysis	1
1.2.1	Introduction of Time series patterns	1
1.3	Gauss-Markov Assumption	2
1.4	Issues In Using OLS with Time Series Data	5
1.4.1	Stationary Time Series	5
1.4.2	Weakly Dependent Time Series	6
1.4.3	Asymptotic Properities of OLS	7
1.4.4	Highly Persistent Time Series in Regression Analysis	7
2	Basic Models	8
2.1	Static Model	9
2.2	Distributed Lag Models	10
2.2.1	Finite Distributed Lag (FDL) Models	10
2.2.2	Infinite Distributed Lag (IDL) Models	11
2.2.3	Geometric Distributed Lag (GDL) Models	12
3	Trending data	13
3.1	Deterministic Trend	14
3.1.1	Fitting Linear Trending in Regression	14
3.1.2	Fitting Exponential Trending in Regression	15
3.2	Stotistic Trend	16
3.2.1	Regression issue under high persistent series	16
3.3	Spurious regression	17
4	AutoReressive Model	18
4.1	AR in $I(0)$ case	19
4.2	AR fail in $I(1)$ case	20
4.2.1	AR(1) model for unit root test	20
4.2.2	Dickey Fuller Test for a unit root	21
5	Vector Auto-Regressive Model	22
5.1	Structural AR model	22
5.2	VAR model	22
5.3	Granger Causality method for Prediction:	23
6	Serial correlation of error term	25
6.1	Testing for Serial Correlation	25
6.1.1	T-Test for AR(1) serial correlation under strict exogeneity:	26
6.1.2	F-Test for AR(p) serial correlation under strict exogeneity:	27
6.1.3	Breusche-Godfrey Test under weak exogeneity:	27
6.2	Correcting Serial Correlation	27
6.2.1	AR(1) serial correlation with static model:	27
6.2.2	Cochrane-Orcutt Correcting Serial Correlation (Feasible GLS)	28
6.2.3	AR(1) serial correlation in FDL(1) model:	29

7	Cointegration and Error Correction Model	31
7.1	Cointegration	31
7.2	EngleSX Granger test for Cointegration:	31
7.3	Error Correction Model for Cointegration	32
8	Forecast	34
9	Appendix: Basic Knowledge	35

1 Introduction of Time Series Analysis

1.1 Time Series Data

A sequence of RV indexed by Time.(fixed intervals, stored chronologically) (*Stochastic Process*). Intuitively, just data with autocorrelation/ corrected over time! and thus it is important information to help us to forecast!

$$\{Y_t : t = 1, 2, 3 \dots T\} = \{Y_1, Y_2, \dots, Y_T\} \quad (1)$$

The sample size for a time series data set is the number of time periods T.

Remark:

- Big difference with cross sectional data is: **data ordering**, time series must **fix order**, while cross sectional can shuffle index anyway we like.
- Also, in time series data, at any given period, we only observe **one realization/ outcome** of a RV. (cross-sectional + time series = panel data).
- Another difference between cross-sectional and time series data is that the **randomness** in time series data is diff from cross sectional case. Certainly, outcomes is random variable time series case.

1.2 Time Series Analysis

Time series analysis involves developing econometric methods capable of estimating and interpreting model parameters, testing hypotheses concerning economic phenomena, and making forecasts of economic variables.

Dynamic Linear Models (DLMs)/State Space Model define a very general class of non-stationary time series models.

1.2.1 Introduction of Time series patterns

We interpret a time series as:

$$Y_t = f(T_t, S_t, C_t, E_t)$$

- Trend component: Time series models may incorporate *trends*, as economic data are often trending (increasing or decreasing for extended periods). In this course, we consider two common trend pattern: linear and exponential trend.
- Seasonality: For time series data at the monthly and quarterly frequencies, seasonality can be an issue and needs to be addressed. Time series data can have complex Seasonality pattern.
- Cyclic Component: A cyclic pattern exists when there are medium or long run fluctuations in the time series that are not of a fixed period.
- Irregular: Short term fluctuations and noise. Generally after removing other components, white noise already. AND sometimes, we can based on irregular plot to decide if cleaning data needed.

white noise: The time series is **uncorrelated over time** is known as white noise. (i.e. indexed by time and also un autocorrelated). If a time series is white noise, it is a sequence of random number and cannot be predicted. And thus, if the forecast errors are not white noise, it suggests we can improve model further. Note that white noise can be homoskedasticity or heteroskedasticity. And we can use robust method if heteroskedasticity.

1.3 Gauss-Markov Assumption

OLS on time series data can be **unbiased**, but we need **quite strict assumptions** for this to work. We can use a lot of assumptions that were similar with cross sectional OLS case. Generally, the statistical properties of OLS in cross-sectional data cases are easily to be violated in time series data cases.

- **Linearity:** Same as cross sectional case.

$$y_t = \beta x_t + \epsilon_t \quad (2)$$

where $x_t = (x_{t1}, x_{t2}, \dots, x_{tn})$ and β is the associated set of parameters. We assume that y_t is a linear combination of β terms and an error term.

- **No Perfect Collinearity:** No independent variable is constant and no explanatory variable is an exact linear function of the others. This assumption **rules out perfect correlation** between the explanatory variables. $Corr(x_{t,j}, x_{t+1,j}) \neq 1, Corr(x_{t,j}, x_{t+1,j}) \neq 0$.
The consequence of **high correlation among the explanatory variables over time** : still unbiased and consistent, but it can **cause difficulty in precisely estimating parameters** (i.e. Cause the variance of the estimators to be high and affect standard inferences.).

- **Strict Exogeneity:** The regressors are contemporaneously exogenous, that is:

$$E(\epsilon_t | \mathbf{X}) = 0, \forall t \quad (3)$$

where

$$X = (x_1, x_2, \dots, x_T)', \text{matrix of regressors}$$

X is vectors in multi-variate regression.

Strictly exogeneity: there is no correlation or relationship between the error term and the explanatory variables for **all** the time.

BUT anything that results in the unobservables at time t , correlated with any explanatory variables at any time period (*omitted variable*) is a violation of strict exogeneity assumption. Thus, we consider a less restrictive form **weak exogeneity**.

$$E(\epsilon_t | x_t, x_{t-1}, x_{t-2}, \dots) = 0 \quad (4)$$

This means that conditional expectation of error term is not zero for all time periods, just uncorrelated with **the current and past** explanatory variables, say nothing about the future. Also, we define **contemporaneous exogeneity**:

$$E(\epsilon_t | x_t) = 0 \quad (5)$$

This imply NO correlation between the explanatory variables and the error term for within that time period only. The good news is the contemporaneous exogeneity is **enough for consistency** in OLS. But fail in unbiased.

Remark:

- Under the assumptions 1,2 3 of linearity, **no perfect collinearity, and strict exogeneity**, the OLS estimator is **unbiased**. $E(\hat{\beta}) = \beta$
- under the assumptions 1,2 3, the error terms ϵ_t are allowed to be correlated over time. $\text{corr}(\epsilon_t, \epsilon_{t+1}) \neq 0$, so the correlation between explanatory variables over time does not affect unbiasedness.
- Limitation of unbiased (based on assumption 1,2 3): unbiasedness say nothing about how *precise* the OLS are, and does not give way to test hypotheses or construct CI (We dont know anything regarding certainty). Therefore, we need assumption 4-6 to be satisified in order for us to test hypothesis and construct confidence intervals so that we can derive the estimators distribution.

- **Homoskedasticity:** The conditional variance of error term is constant over time.

$$V(\epsilon_t|\mathbf{X}) = \sigma^2, \forall t \quad (6)$$

- **No Autocorrelation:** The error terms are uncorrelated over time.

$$\text{Cov}(\epsilon_t, \epsilon_s|\mathbf{X}) = 0, \forall t \neq s \quad (7)$$

We are saying here that there is no covariance between error terms across all time periods. But in time series data, it is very easily violated, so we discuss it more detailed and give method to correct it in chapter 6.

With assumption 4 and 5, we can now derive the **spherical disturbances** assumption:

$$X(\epsilon\epsilon'|\mathbf{X}) = \sigma^2\mathbf{I}, \forall t \neq s \quad (8)$$

where \mathbf{I} is the T-dimensional identity matrix and $\sigma^2\mathbf{I}$ is T*T dimension. Since it is an identity matrix, that means we dont have any covariance between error terms across time but we do have variance terms for the error terms.

Three different kinds of correlations:

Note that we have 3 different kinds of correlations that may differ across time series regression and need to consider.

1. **correlation between a predictor $\{x_t\}$ itself across time:** Correlation between $x_{t,j}$ and $x_{s,j}$. This is **not an issue** and always happen, unless there is a perfect linear relationship.
2. **Correlation between $x_{t,j}$ and ϵ_s :** This violates the strict exogeneity assumption and this leads to OLS being **biased** as a result. However, we can still have weakly exogeneity assumption.
3. **Serial correlation between ϵ_t and ϵ_s .** Correlation between error term in different period. This violates no autocorrelation assumption 5 but only affects the efficiency of the OLS estimator. Does not cause bias.

It is possible to have the strict exogeneity satisfied, while the no autocorrelation is violated. And generally, serial correlation in $x_{t,j}$ is allowed and even expected!

Autocorrelation:

$$\rho = \frac{\sigma_{x_t, x_s}}{\sigma_{x_t} \sigma_{x_s}}, \forall t \neq s \quad (9)$$

Assume $s = t - k$, ($k = 1, 2, 3, \dots$), then define **Autocorrelation of the order k**:

$$\rho_k = \frac{\sigma_{x_t, x_{t-k}}}{\sigma_{x_t} \sigma_{x_{t-k}}}, \forall k \quad (10)$$

Note that ρ_t, ρ_{t-k} come from the same stochastic process.

Under the assumption of **stationary**, we have *constant variance*:

$$\sigma_t = \sigma_{t-k} = \sigma_0, \forall t \quad (11)$$

and thus covariance only related to k:

$$\sigma_{x_t, x_{t-k}} = \sigma_{x_s, x_{s-k}} = \sigma_k, \forall s, t \quad (12)$$

autocorrelation of stationary series:

$$\rho_k = \frac{\sigma_k}{\sigma_0^2} \quad (13)$$

where

$$\sigma_k = E[(x_t - \bar{x})(x_{t-k} - \bar{x})], \sigma_0 = E[(x_t - \bar{x})^2]$$

Note that we only have 1 \bar{x} since the mean is the same due to stationary data. And in Weakly stationary case, the formula of order k autocorrelation is a function of k, nothing to do with t.

Gauss-Markov assumption: Under Assumption 1 -5, the OLS estimator is the **Best Linear Unbiased estimator (BLUE)**.

- **Normality:** Cannot perform exact inference unless add normality assumption: ϵ_t is iid as normal random variables with zero mean and constant variance.

$$\epsilon_t \sim N(0, \sigma^2), \forall t$$

The *Gauss Markov Assumptions* and *normality assumption* gives us the **Classical Linear Model assumptions (CLM)** for time series. We need this so that we can perform exact inferences. Then, all of the statistical inference procedures can be applied here (same as the cross sectional case):

1. **t statistic** have exact t_{T-k-1} distributions under the null hypothesis (k is the number of regressors).
2. **F statistic** have exact F distributions under the null hypothesis.

Remark:

- With contemporaneous exogeneity, this is sufficient for OLS estimators to be consistent. But not enough to ensure unbiasedness.

- **Serial correlation** (fail in assumption 5) tends to be a problem even when strict exogeneity holds. namely, static and finite distributed lag models serial correlation is often a problem.
- When **spherical disturbances** assumption holds, OLS is BLUE and usual OLS variance formulas apply.
- we can add *normality* and then exact inference. However, the full set of CLM assumptions tends to be highly unrealistic for many real cases.

1.4 Issues In Using OLS with Time Series Data

Summary

1. **Large Sample Approximation:** Same as cross-sectional analysis, in time series analysis, if Error term not from normal dist, we must rely on CLT to justify the usual OLS test statistic and Confidence Interval.
2. **Strict Exogenous Assumption:** might be violated in static model and distributed lag models. And must be violated in models with lagged dependent variables
3. **Random Sampling:** Big Challenge in time series cases as observations always correlated over time.
4. **CLT possible in weakly dependent case:** It is lucky that the CLT holds for the weakly dependent time series process. The key is whether the correlation between variables at different time periods tends to zero quickly enough. The time series that are not weakly dependent, (examples in 11.3), do not generally satisfy the CLT.
5. **CLT possible in stationary case:** The CLT for time series data requires stationarity and some form of weak dependence: thus, the stationary and weak dependence are ideas for use in multiple regression analysis.

1.4.1 Stationary Time Series

Stationary is an aspect of a **sequence** result, not for single point. We use stationary process means the strict form.

A (strict) stationary time series is a sequence of a random variable collected over time, whose probability distribution are stable. Stationary process can be highly correlated, and it does require that the nature of any correlation for fixed time horizon h is same over time.

- **Non-Stationary:** At minimum, it means change over time!
- **Strict Stationary:** Fixed the time horizon h , the **shift joint distribution** does not change.
- **Weak stationary/ Covariance-stationary:** When sequence x_t , with finite second moment, i.e, $E(x_t) < \infty, \forall t$, and satisfy:

$$E(Y_t) = \mu \quad (14)$$

$$V(Y_t) = \sigma^2 \quad (15)$$

$$\text{Cov}(Y_t, Y_{t+h}) = g(h), \quad (16)$$

a function of h only.

The weakly stationary process focus only on the first and second moments of a stochastic process. i.e. it only need the **mean and variance exist and constant over time**.

Remark:

- **Strict stationary \neq weakly dependent**
Strict stationary \neq weak stationary
Weak stationary \neq weakly dependent .
 When a strict stationary has finite second moment, it is Weakly Stationary.
- Simple strict stationary do NOT imply weakly stationary. For example, a strong stationary process related to the stable Cauchy joint pdf when shifted over time, but it fail in covariance stationary, as Cauchy dist has undefined mean and variance (infinity first and second moments).
- When a (strict) stationary process has a **finite second moment**, then it must be weakly stationary, but the converse is not true.
- A unit root time series is non-stationary, but **non-stationary series might not a unit root** as there are lots of other ways a time series can be non-stationary, such as being trend stationary.
- Stationarity is a property of a process sequence, not time series themselves. These processes generate time series.

Stationary in LLN and CLT:

Stationary simplifies statements of the LLN and CLT. As only when the features of variable are stable over time, we can have chance to learn some things from regression. Otherwise, if variable is just an arbitrary fact, it is impossible for us to conclude/predict anything, as the history cannot be useful in the future. In the multiple regression model for time series, we are assuming a certain form of stationary in that the β_j does not change over time, and based on this, the CLT is possible to be hold.

1.4.2 Weakly Dependent Time Series

Another main concept for CLT holds! Dependency places restrictions on how strongly related the RV X_t and X_{t+h} as h increases without bound.

A covariance stationary time series is weakly dependent if the correlation bet X_t and X_{t+h} goes to zero sufficiently quickly as h goes to infinity. i.e. as the variable goes further over time, the correlation bet them becomes smaller and smaller, and become almost independent as h increase:

$$\text{Corr}(Y_t, Y_{t+h}) \rightarrow 0, h \rightarrow \infty \quad (17)$$

Weak dependent in LLN and CLT:

Essentially, weak dependent replace the assumption of Random sampling in LLN and

CLT. The simplest example of a weakly dependent time series is an iid sequence.

An **asymptotically uncorrelated** example is **Stable AR(1) model**:

$$y_t = \rho y_{t-1} + e_t, t = 1, 2, 3... \quad (18)$$

The starting point is y_0 at $t = 0$, and residual is an iid sequence with mean 0 and constant variance. Also the e_t ind of y_t . With $|\rho| < 1$, then, $\text{corr}(y_t, y_{t+h}) = \rho^h$ goes to zero (convergent geometric series).

By contrast, when $\rho = 1$, **Random walk model here is not weakly dependent**, the geometric series is divergent.

Remark:

The **trending series (non stationary) can be weakly dependent**. i.e. There is no relationship bet stationary and dependent. For example, the simple linear time trend model

$$Y_t = \alpha_0 + \alpha_1 t + e_t$$

is actually weakly dependent.

Difference-stationary process:

Unit root processes like random walks are $I(1)$. **$I(1)$ means that first differencing them will lead to a weakly dependent series (and often stationary process)**. Therefore, a time series that is $I(1)$ is said to be a **difference-stationary process**, i.e. $I(1)$ can get stationary after first difference.

$I(0)$: A stationary, weakly dependent time series process that, when used in regression analysis, satisfies the LLN and the CLT. So, Covariance stationary is $I(0)$. i.e. **all stationary time series are $I(0)$ processes but not all $I(0)$ processes are stationary**. $I(0)$ process is a necessary but not sufficient condition for a time series to be stationary.

When we talk about $I(D)$, it talks about the **number of times we need to difference the time series in order to get a stationary data**. After you differenced it, the new time series you generated by differencing it is $I(0)$, since you don't need to difference it to make it stationary. The original time series is described to be $I(1)$ whilst the new time series is $I(0)$, it's all a matter of perspective on which time series you are describing. Hence, $I(0)$ is a different case so don't worry too much about it.

1.4.3 Asymptotic Properties of OLS

1.4.4 Highly Persistent Time Series in Regression Analysis

Random walk staff here

2 Basic Models

Summary

- Static models closely resemble cross-sectional models (e.g., data ordering doesn't matter here)
Static models typically used to estimate a contemporaneous relationship between two or more variables. Static models do not capture effects that take place with a delay (autoregressive with itself), which is a big difference with AR/VAR model
- Distributed lag models can be used for such purposes (data ordering matters here).
- Both static model and distributed lag model **can not be suitable for forecasting**:
They ignore that past outcomes on y can also help forecast y ;
The future realisations of x need to be known in order to forecast y ; (forecast y_t when x_t has already been known? stupid).
- Static model and DL model are good in terms of **interpreting causal effects for two variables in the same time period**.

2.1 Static Model

A static model is only for the current time period and relates two or more time series with each other. We use static model when we interest in knowing the **contemporaneous trade off** between Y and Z. For example, the *Phillips Curve*:

$$Inf_t = \alpha + \beta Unem_t + \epsilon_t \quad (19)$$

where Inf_t is the annual inflation rate and $unem_t$ is the annual unemployment rate. This form of the Phillips curve assumes a *constant natural rate of unemployment* and *constant inflationary expectations*, β measure a potential **contemporaneous trade off** between inflation and unemployment. (positive trade off when $\beta > 0$)
More generally:

$$y_t = \alpha_t + \beta_1 x_{1,t} + \dots + \beta_k x_{k,t} + \epsilon_t, \quad \epsilon_t \sim iid(0, \sigma^2) \quad (20)$$

whereby β_j looks at only the *contemporaneous relationship* between $\{x_j\}$ and $\{y\}$. For example, the **ceteris paribus effect of an increase in the Y_t on a value of $x_{1,t}$**

Assumption of static model:

1. **white noise of error term:** white noise are where observations are not autocorrelated with each other, homoskedastic, and indexed by time.
2. **Only contemporaneous relationship:** Assume a change in Z_t at time t has *immediate* effect on y_t . So, serial correlation easily to be a problem in static model.

Remark:

- Recall that **serial correlation is the correlation of a variable with itself over time**. Serial correlation for the error terms does not affect the bias or consistencies of least squares, but less efficiency and fail in statistical inference. even in the case of large samples. (See the issue of serial correlation section)
- If **strictly exogeneous**, then unbiased and consistent OLS possible. BUt If only **contemporaneously exogenous**, then OLS is consistent only (might biased) given the time series are weakly dependent.

2.2 Distributed Lag Models

DL model are good for estimating **lagged effects** of, say, policy. Useful when effects are not immediately and the past can affect the future.

2.2.1 Finite Distributed Lag (FDL) Models

Instead of looking at things in a contemporaneous measure, we can add in lag effects when we believe one or more variables to affect y with a lag.

Suppose we think that a change in a variable z today, can affect a variable y up to two periods into the future. Then, the required FDL model is:

$$umbrellassold_t = \alpha_0 + \delta_0 rain_t + \delta_1 rain_{t-1} + \delta_2 rain_{t-2} + \epsilon_t \quad (21)$$

which tells us that umbrellas sold today is affected by amount of rain today, yesterday, and 2 days ago (since if its been raining for past few days, people will start giving in and buying their own umbrellas instead of stealing it from poor souls at Fisher library).

In general, an FDL model of order q is given by:

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + \dots + \delta_q z_{t-q} + \epsilon_t \quad (22)$$

Lag selection (optimal q) by economic judgment suggested theory, or statistical tests: Information criteria (AIC, BIC), or R^2

The coefficient on the contemporaneous z, δ_0 , is called the **impact propensity (IP)**:

1. It tells us the **immediate effect** in y when z increases by one unit.
2. If both variables are in *logarithmic form*, the IP is sometimes called the **short run (or instantaneous) elasticity**.
3. $\delta_0 = 0$, which impose dropping z_t from regression, means no immediately effect.

The sum of all lag coefficients (including the zero-lag) is called the **long run propensity (LRP)** :

$$LRP = \delta_0 + \delta_1 + \delta_2 + \dots + \delta_q \quad (23)$$

1. **Interpretation** of LRP: Suppose the level of z increases **permanently** at a given period in time (e.g., the minimum wage increases by 1.00 per hour and stays there). The LRP is the *ceteris paribus* (holding others constant) change in y after the change in z has passed through all q time periods.
2. If y and z are both given in natural logarithms, the LRP is called the **long run elasticity**.
3. **LRP intuition:** (q = 2) With a permanent effect in z, after one period, y increase by $\delta_0 + \delta_1$, after two period, y increase by $\delta_0 + \delta_1 + \delta_2$. As suppose q = 2, there is no further change after two period. This shows that the sum of the coefficients on current and lagged z $\delta_0 + \delta_1 + \delta_2$ is the long-run change in y given a permanent increase in z and is called the long-run propensity (LRP) or long-run multiplier.

Proof 7: Show ceteris paribus effect of z on y/ temporary increase in z

$y_{t+3} = y_{t-1}$ in the FDL model lag 2, when change is happen at y_t .

$\delta_0 = y_t - y_{t-1}$: impact propensity

$\delta_1 = y_{t+1} - y_{t-1}$: change in y one period after the temporary change in z, $\delta_2 = y_{t+2} - y_{t-1}$.

Proof 8: LRP, Show change of y due to permanent increase in z

After one period, y increase by $\delta_0 + \delta_1$, after two period, y increase by $\delta_0 + \delta_1 + \delta_2$. No further change after two period ($y_{t+2} = y_{t+3} = y_{t+4} = \dots$)

Given an FDL model with lag q, it contain the static model when after coefficients are zero. Sometimes, the primary purpose for estimating a distributed lag model is to test whether z has a lagged effect on y.

For any horizon h, we can define the effect as $\delta_0 + \delta_1 + \delta_2 + \dots + \delta_h$, which is interpreted as the change in the expected outcome h periods after a permanent, one-unit increase in x. The LRP is the cumulative effect after all changes have taken place; it is simply the sum of all of the coefficients.

Furthermore, note that since z_t is serially correlated with itself through time (which is fine and should be expected!), it will be *difficult in getting precise estimates of each δ* . From this, even if we run a **F-test** on $\delta_0 = \dots = \delta_h = 0$ and find that it is statistically significant, we still don't know which lag period is significant with y_t (they may be jointly significant but individually insignificant). If $\delta_0 = \dots = \delta_h = 0$, then it will suggest to us to use a static model since lags are insignificant.

2.2.2 Infinite Distributed Lag (IDL) Models

An infinite distributed lag (IDL) model relates dependent variable to current and all past values of independent variable. In other words, IDL model allow a change in an explanatory variable to **affect all future values** of the dependent variable:

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + \dots + \epsilon_t \quad (24)$$

Similar interpretation as in an FDL, **IP** still δ_0 . And δ_h has same interpretation with DFL mode:

From any given level $z_s = c$, $s < t$, if z temporarily increases y one unit., then δ_h measures the change in the expected value of y after h period. LRP measures the long-run change in the expected value of y, given one unit permanent increase in z.

For a **temporary change** in z, we have

$$E(y_t) = \alpha + \delta_h \quad (25)$$

for IDL to make sense, we need that δ_h close to 0 as time progresses. This means that *there is no long run effect from a temporary change in z.* and

$$E(y_t) = \alpha, \quad h \rightarrow \infty \quad (26)$$

For a **permanent change** in z, we have

$$E(y_t) = \alpha + \delta_0 + \delta_1 + \delta_2 + \dots + \delta_h \quad (27)$$

which converges to some new level as h goes to infinity.

Unlike the FDL model, lag here is infinity. The model, of course, is only an approximation to reality (i.e., no economic process started indefinitely far into the past).

δ_h measures the change in the expected value of y after h periods. The δ_j goes to 0 as the order j goes to infinity. This is not to say that δ_2 is smaller in magnitude than δ_1 ; it only means that **the impact of z_{t-2} on y_t must eventually become small as j gets large**. It make sense that the effect of a z variable on y decreases the further back in time we go! (Which makes sense since inflation in 1900 shoudnt affect inflation in 2017). Therefore, the effects are strongest for more recent and contemporaneous effects. Note that this doesnt mean that 2016s effect has to be larger than 2015s, just that when its really far back, it gets smaller.

We implictly assume **strict exogeneity** for this model whereby even all future values of z_t are uncorrelated with the error term. An issue with this model is that it assumes that future values of z_t are constant and never changing which isnt realistic (since if z_t was interest rates, this would definitely be changing in 20 years time compared to today). Therefore, we use a **weak exogeneity** assumption instead (where the error is only uncorrelated with current and past z).

2.2.3 Geometric Distributed Lag (GDL) Models

Even if we were to claim that the IDL is a useful model, we cannot estimate it without restrictions. The simplest **restriction type** in the geometric (or Koyck) distributed lag (GDL) model.

From this, the new δ_j depends on two parameters:

$$\delta_j = \gamma \rho^j, \quad |\rho| < 1, \quad \forall j = 0, 1, 2, \dots \quad (28)$$

where γ and ρ can be positive related, or negative related. As long as $|\rho| < 1$, the δ_j will goes to zero as n goes to innfinity. (Convergent geometric series). otherwise, divergent. To ensure $|\rho| < 1$ in any case, we shall consider difference sometimes.

The **impact propensity (IP)** is $\delta_0 = \gamma \rho^0 = \gamma$, so the **sign of impact propensity is determined by the sign of γ** ,

$$y_t = \alpha + \gamma z_t + \gamma \rho z_{t-1} + \gamma \rho^2 z_{t-2} + \dots + \epsilon_t, \quad |\rho| < 1 \quad (29)$$

Note that if $\gamma > 0$, and $\rho > 0$, then every coefficient is positve. However, if $\gamma < 0$, then the lag coefficients will alternative in sign. And **LRP also has the same sign with γ !**

$$LRP = \frac{\gamma}{1 - \rho} \quad (30)$$

Proof 12: Application of Koyck

Given $y_t = \alpha + \gamma z_t + \gamma \rho z_{t-1} + \gamma \rho^2 z_{t-2} + \dots + \epsilon_t$,
show it is equivant to FDL model of order 1: $y_t = \alpha_0 + \gamma z_0 + \rho y_{t-1} + \epsilon_t - \rho \epsilon_{t-1}$ However, if we run OLS on this, we get inconsistent estimates as y_{t-1} is correlated with ϵ_{t-1} . (we can use an instrumental variable and estimate the model with that instead).

3 Trending data

Trending process:

A time series whose expected value is an increasing or a decreasing function over time. A natural feature in most time-series data! A special case of **non-stationary** process.

Summary

- when **deterministic trend** in time series, by detrending, we can get **trend-stationary** result, and a process that is stationary once the trend is removed, usually implicit that series is **weakly dependent**.
- when **stochastic trend** in series, by difference term, we can get **differencing stationary**
- Unit root is a kind of non-stationarity.
- High-persistent series must be non-stationarity.
- Series can be trending but not be highly persistent.
- **Spurious regression** could be an issue among deterministic or stochastic trend.

3.1 Deterministic Trend

In order to capture 'true' relationship between Y and Z, we can adding **trend variable**: Detrending y and all explanatory variables by it. Just count data. For example,

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \gamma t + \epsilon_t, \quad t = 1, 2, \dots, T$$

This equation allow as to control for **linear trend** that affect y_t , but may also be related to trend in x_{1t} and x_{2t} .

After adding trend t, if we had all the assumptions of the classical linear models satisfied, then we can apply test statistics and confidence intervals in the usual way.

If assumptions of linearity, no perfect collinearity, and strict exogeneity holds, then without trend variable t can cause bias in estimating β_1 and β_2 . (we can think trend variable as an **omitted variable bias**).

However, when **dependent variable is trending**, we shall care about **goodness of fit**, i.e. we can still have a high R-square which is very misleading and thus we should not credit the fit result due to y_t trending up or down for reasons we have not explained.

The time trend t represents our ignorance about omitted factors causing y_t to trend up or down. However, we should still be happy with it since we can still fit y to a trend.

3.1.1 Fitting Linear Trending in Regression

We tend to use linear specification to account for linear trend.

$$y_t = \alpha_0 + \alpha_1 t + \epsilon_t, \quad \epsilon_t \sim iid(0, \sigma^2)$$

Then, the expectation is a linear function over time

$$E(y_t) = \alpha_0 + \alpha_1 t$$

, and ϵ_t represent the **deviation of the trend**.

By $y_{t-1} = \alpha_0 + \alpha_1(t-1) + \epsilon_{t-1}$, We also can show that the **average change in y_t is α_1** , $\forall t$:

$$E(\Delta y_t) = \alpha_1$$

So from the original equation: $y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \gamma t + \epsilon_t$, the γ tells us what is the expected trend change in y.

Also, we can do detrend transformation: step1: Fitting linear trend model with *trend component* by:

$$y_t = \hat{\alpha}_0 + \hat{\alpha}_1 t + \varepsilon_t$$

step2: get **detrend component**:

$$y_t^* = y_t - \hat{\alpha}_1 t = \hat{\alpha}_0 + \varepsilon_t$$

then, the OLS coefficient on regressor in this detrend series case shall same as the ones in the model with a trend t.

3.1.2 Fitting Exponential Trending in Regression

Some data are better approximated by exponential trends. Therefore we can use exponential trends whereby for strictly positive variables, we can capture an exponential trend by:

$$y_t = e^{\beta_0 + \beta_1 t + \epsilon_t}$$

,

and the **natural log follow the linear trend** as $\log(y_t) = \beta_0 + \beta_1 t + \epsilon_t$. and

$$\Delta \ln(y_t) = \ln(y_t) - \ln(y_{t-1})$$

$$E(\Delta \ln(y_t)) = \beta_1$$

But the **change in log approximates the growth rate**:

$$\beta_1 \approx \frac{\Delta y_t}{y_{t-1}}$$

Note that with exponential trend, the **rate of growth is constant**. From this, β_1 tells us the **average growth rate over time**.

3.2 Stotistic Trend

An alternative to represent the trending series is to use a **random walk process** whereby:

$$y_t = y_{t-1} + \epsilon_t$$

then,

$$y_t = y_0 + \epsilon_1 + \epsilon_2 + \dots + \epsilon_T$$

and $\text{corr}(y_{t+h}, y_t)$ never die out, **RW process is high presistent:**

$$E(y_t) = E(y_0)$$

$$V(y_t) = T\delta_\epsilon^2$$

$$\text{cov}(y_t, y_{t+s}) = V(y_t) = t\delta_\epsilon^2$$

$$\text{corr}(y_t, y_{t+s}) = \frac{\text{cov}(y_t, y_{t+s})}{\sqrt{V(y_t)}\sqrt{V(y_{t+s})}} = \sqrt{\frac{t}{t+s}}$$

The varianve of RW process increases linearly with time. And for sufficiently long time series, this implies that the correlation between y_t and y_{t+h} remains high (close to unity) even as s grows. This is known as highly persistent or strongly dependent time series which **does not satisfy the LLN or CLT**.

High persistent series, such as RW, we now have a sequence of random variables which do not disappear even as time moves on. The effects of shocks do not disappear and therefore **non-stationary**.

Do not confused trending and highly persistent. **Series can be trending but not be highly persistent**. However, often the case that highly persistent also contains a clear trend, such as a **random walk with drift**:

$$y_t = \alpha + y_{t-1} + \epsilon_t$$

3.2.1 Regression issue under high persistent series

Highly persistent series such as random walks causes serious problems for regression analysis. If we regress y on x, then we get parameters that are statistically significant even though they are unrelated. Consider

$$y_t = \beta_0 + \beta_1 x_t + \zeta_t$$

where **y and x are random walk process**, s.t,

$$y_t = y_{t-1} + \epsilon_t, \epsilon \sim iid(0, \delta_\epsilon^2)$$

$$x_t = x_{t-1} + \nu_t, \nu \sim iid(0, \delta_\nu^2)$$

assume

$$\text{corr}(\nu_t, \epsilon_t) = 0$$

which mean that y and x are independent process. Then, given also $E(y_t) = 0$, it would suggest that $\beta_0 = \beta_1 = 0$, and then

$$\zeta_t = y_t = \epsilon_1 + \epsilon_2 + \dots + \epsilon_T$$

which is mean zero but varince goes with t.

Remark:

- Under the regress with two independent RW process, more generally I(1), Error term is high persistent

- Also the **OLS estimator** $\hat{\beta}_1$ **does not converge to zero**, which means that OLS is **inconsistent**.
- An even more serious issue is that the t-statistic for $H_0 : \hat{\beta}_1 = 0$ does not have a t-distribution, even in large samples, as we get more data, $\hat{\beta}_1$ does not get closer to 0 or even converge to a specific value. Furthermore, the t-statistic rejects the null hypothesis too often and this issue gets worse as sampling size grows (we are more likely to make a type 1 error, rejecting true null, with more data).

As we fail in statistical inference, we will never figure out that yt is unrelated to xt, and thus **spurious relationship** occur. With RW, the problem arises because xt has too much temporal correlations for the law of large numbers to hold.

3.3 Spurious regression

Spurious regressions can happen in **deterministic trend, stochastic trend or non-cointegrated I(1) series** (but generally trend data), whereby we think there is a significant correlation and observe high R^2 when in fact, the series are independent.

A spurious regression means the relationship specified in the regression does not exist, but the regression output fails to convey this fact. So what happens? OLS estimators are **not consistent, the t-statistics diverge to infinity** as sample size increases.

Provided the series are *weakly dependent* about their time trend, the problem is efficiently solved by including a time trend in the regression model.

Regressions with two or more independent random walks results in spurious regression problem in time series. The two series are independent but they appear to be strongly related since they could just trend in the same manner. From this spurious regression, the R^2 will be quite large and thus invalidated.

4 Autoregressive Model

Summary

If a process is $I(1)$, then it has a unit root (so we need to first-difference it to get a $I(0)$ process). By fitting the $AR(1)$ model, we are effectively estimating the unit root parameter, which, first of all, will be biased; but also, we will have inference issues, as the variance will be non-finite.

So generally, $AR(1)$ model can only be 'correct' model in $I(0)$ case. "Correct" Typically means whether the Gauss-Markov assumptions are (mostly) satisfied.

4.1 AR in I(0) case

AR model is a time series regression, in which both the dependent and independent variables belong to the **same stochastic process**. Then, by **assuming white noise process in error term**, the simple AR(1) is:

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$$

where β is called **persistent parameter**. Note that error term is not autocorrelated and iid unlike response variable.

We shall have **high persistent** time series, when $\beta = 1$, it is Random Walk process. By contrast, when $\beta = 0$, y_t then just follow white noise process, no persistent.

Proof 1: Show persistent in AR(1) model

$$y_t = \alpha \sum_{i=0}^{t-1} \beta^i + \beta^t y_0 + \sum_{i=0}^{t-1} \beta^i \varepsilon_{t-i}$$

Clearly, we can see if the series is weak persistent depends on if $|\beta| < 1$

Proof 2: Show weak persistent AR(1) model converge form

$$y_t = \frac{\alpha}{1-\beta} + \sum_{i=0}^{t-1} \beta^i \varepsilon_{t-i}$$

Proof 3: Show unconditional mean, variance in weak persistent AR(1)

Unconditional Mean: $E(y_t) = \frac{\alpha}{1-\beta} = \mu$

Unconditional VArance: $E(y_t) = \frac{\sigma_\varepsilon^2}{1-\beta^2} = \gamma_0$

Autocorrelation: $\rho_k = \beta^k$

Thus, y_t is stationary, so time in-variant, but RW process is non-stationary as its mean/variance never converge.

The **Autocovariance in AR(1)** by defination is:

$$\gamma_k = cov(y_t, y_{t-k}) = E(y_t, y_{t-k}) - \mu^2$$

Back to AR(1) model:

$$E(y_t y_{t-k}) = \alpha\mu + \beta E(\beta y_{t-1} y_{t-k})$$

Proof 4:

$$\gamma_k = \beta \gamma_{k-1}$$

and thus, the **autocorreation of AR(1)** is:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \beta \frac{\gamma_{k-1}}{\gamma_0} = \beta \rho_{k-1}$$

It follows that:

$$\rho_0 = \frac{\gamma_0}{\gamma_0} = 1, \rho_1 = \beta\rho_0 = \beta, \rho_2 = \beta^2\rho_0 = \beta^2, \rho_k = \beta^k\rho_0 = \beta^k$$

That is, the autocorrelation of AR(1) is **geometric decay**. The smaller $|\beta|$ is, more rapid decay. **ACF plot**

The optimal lag selection is based on **Information criteria**. The model with mini AIC/BIC is preferred. BIC selects might less parameters than AIC, but can also give same selection result. Relative value (not absolute) of the criteria matter. Also need to test for serial correlation of selected model.

4.2 AR fail in I(1) case

Weakly dependent processes are said to be integrated of order zero or I(0). Therefore, we don't need to do anything to them before regressions since they satisfy standard limit theorems.

By contrast, **Unit root** processes like random walks are integrated of order one or I(1). This means that first differencing them will lead to a weakly dependent series (and often stationary process). It can be thought as a **stochastic trend** in time series. Therefore, a time series that is I(1) is said to be a difference stationary process. Unit roots means that **the effects of any shocks do not disappear over time** (which is the opposite of weakly dependent).

unit root is nonstationary but inverse is false, and RW is a special case of unit root, as we can get stationary after first order difference.

To elaborate more on unit roots, we can think of them as a stochastic trend in time series. If we had a time series of:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \epsilon_t$$

the coefficient α_1 is a **root**. We expect this process to always converge back to the value of α_0 when absolute value of $\alpha_1 < 1$. If we set $\alpha_0 = 0$ and $\alpha_1 = 0.5$, if yesterday was 100, then today it's 50, tomorrow 25, and so on until it gets to 0. Here, we can see that this series will **converge back** to α_0 . However, if we had a root that is a unit, or in other words, when $\alpha_1 = 1$, we see that the series will never converge back. From this, we can see that the time series will never recover back to its expected value and therefore the process is very susceptible to shocks and hard to predict.

4.2.1 AR(1) model for unit root test

A time series with an unit root means that any shock does not disappear over time, as the geometric series divergent, it cannot converge back to initial state before the shock. I.e. Non-stationary, as no fixed structure.

Looking at the unit root tell us whether the time series is stationary. If does not have unit root, it is stationary I(0), so based on AR(1) model:

$$y_t = \beta y_{t-1} + \epsilon_t \tag{31}$$

and $H_0: \beta = 1$ (unit root, not stationary), reject H_0 means stationary. But we **cannot use usual t-stat as data is not stationary** (Weak dependent of error term is necessary

for CLT holds).

4.2.2 Dickey Fuller Test for a unit root

Dickery Fuller distribution is an **asymptotic distribution of t-statistic under null of one unit root**.

$$\begin{aligned}y_t &= \rho y_{t-1} + \epsilon_t, \\y_t - y_{t-1} &= \rho y_{t-1} - y_{t-1} + \epsilon_t, \\ \Delta y_t &= (\rho - 1)y_{t-1} + \epsilon_t\end{aligned}$$

Let $\theta = (\rho - 1)$, then $\Delta y_t = \theta y_{t-1} + \epsilon_t$

H0: $\theta = 0$, and we shall use the **Dickery Fuller distribution table** for the critical value. If no time trend:

$$\Delta y_t = \alpha + \gamma y_{t-1} + \epsilon_t$$

We also has Dickery Fuller distribution table **with time trend**, so that a **trend-stationary process is not mistaken for a unit root process**.

$$\Delta y_t = \alpha + \delta t + \gamma y_{t-1} + \epsilon_t$$

Then under the **alternative y_t is a trend-stationary process**.

Dickery Fuller test based on the **assumption that ϵ_t is white noise**(i.e. serial uncorrelated), otherwise the serial correlation must be account for by AR(p) process, before unit root test. This result in the **Argumented Dickey Fuller test**:

$$\Delta y_t = \alpha + \delta t + \gamma y_{t-1} + \theta_1 \Delta y_{t-1} + \dots + \theta_p \Delta y_{t-p} + \epsilon_t$$

The distribution not affected by additional of the lagged differences, and thus the critical values can be applied before.

In sum, we can have **three versions of Dickey fuller test**:

1. $\Delta y_t = \rho y_{t-1} + \epsilon_t$ (Test for Unit root)
2. $\Delta y_t = \rho y_{t-1} + \alpha + \epsilon_t$ (Unit root + draft)
3. $\Delta y_t = \rho y_{t-1} + \alpha + \gamma t + \epsilon_t$ (Unit root + draft + deterministic time trend)

Each version of the test has its own critical value. The intuition behind the test is as follows. If the series y is stationary (or trend stationary), then it has a tendency to return to a constant (or deterministically trending) mean. Therefore, large values will tend to be followed by smaller values (negative changes), and small values by larger values (positive changes).

Remark:

- Serial correlation must be account for (by AR(p) process), before unit root test.
- High-persistent series imply non-stationarity, and the inverse not true.
- Series can has trend, but also weakly persistent. Example is **trend-stationary processes**.
- A random walk with drift is an example of a highly persistent series that is trending.

5 Vector Auto-Regressive Model

5.1 Structural AR model

Recall: Given $\{y_t\}$ and $\{z_t\}$ are two **weakly dependent** time series. (i.e. $I(0)$, but not iid sequence) We define the *static relationship* by:

$$y_t = \alpha + \beta z_t + \varepsilon_t, \varepsilon_t \text{ iid}(0, \sigma^2)$$

Note that if $\{y_t\}$ and $\{z_t\}$ are two $I(1)$ process, then without cointegration, error term still $I(1)$.

Also *FDL model* possible:

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + \dots + \delta_q z_{t-q} + \epsilon_t$$

when lag order is 2, it means that a change in z today, can affect y up to two period in the future.

If error terms are autocorrelated, $\epsilon_t = \rho \epsilon_{t-1} + v_t$, we can add lags of dependent variables in order to remove this autocorrelation by build **Structural AR model**

Proof 10: adjust serial correlation in static model

$$\tilde{y}_t = \tilde{\alpha}_0 + \tilde{\alpha}_1 y_{t-1} + \tilde{\beta}_0 z_t + \tilde{\beta}_1 z_{t-1} + v_t, \quad (32)$$

Where $\tilde{\alpha}_1 = \rho$. Note that we have *immediately effect* when z_t in model and $\beta_0 \neq 0$. In this model, we **assume white noise error term**, also note that it can be written by quasi-difference form as.

Now, the parameters in new model are different with the static model ($\tilde{\beta}_0$ is same as the static model, but now, we do have more estimated parameters). So if our model should be static, we have the issue is that now the parameters we have constructed are different to the static model.

From this, we think a model that assuming **no immediate effect**, which means that z_t disappears. and We can think of a **feedback effect** whereby z can affect y BUT then y can also affect z . We can write 2 equations affecting y and z . We have the same variables on the RHS of both equations but the parameters differ.

5.2 VAR model

Assuming all series are weakly dependent!, which mean that they are not really persistent whereby any shocks disappears reasonably quickly. Since both of them are weakly dependent, they are **$I(0)$ and therefore not suffer from spurious regression**. Also **endogenous variable assumption/ no pure exogenous variables** as the regressors are lags of all the variables. and **no current endogenous variables**

VAR is the *multivariate generalization of an AR model*, whereby instead of 1 dependent variable, we now have **m endogenous dependent variables** (since we now have this feedback effect). The simplest **reduced form VAR** (reduced form because we do

not include regressor at the same time period) example (bivariate VAR(1) model):

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \begin{bmatrix} \alpha_{1,0} \\ \alpha_{2,0} \end{bmatrix} + \begin{bmatrix} \alpha_{1,1} \\ \alpha_{2,1} \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \alpha_{1,2} \\ \alpha_{2,2} \end{bmatrix} \begin{bmatrix} x_{2,t-1} \\ x_{1,t-1} \end{bmatrix} + \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix} \quad (33)$$

Simplify:

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \begin{bmatrix} \alpha_{1,0} \\ \alpha_{2,0} \end{bmatrix} + \begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} \\ \alpha_{2,1} & \alpha_{2,2} \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix} \quad (34)$$

where

$$v_t = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix}, v_t \sim iid(0, \Sigma), \Sigma = \begin{bmatrix} \sigma_1^2 & cov(\nu_1, \nu_2) \\ cov(\nu_1, \nu_2) & \sigma_2^2 \end{bmatrix} \quad (35)$$

Note that ν_1 DIFFER ν_2 , and the **error covariance matrix** allows $cov(\nu_1, \nu_2) \neq 0$. **Each error term in each equation is assumed to be iid relative to own future and past realisation BUT may be correlated between equations.**

More precisely, it is a *reduced form of simultaneous equations*(no z_t) with a very free dynamic specification, such that *no causal linages are imposed a priori*: As y_t equation doesnot have z_t inside of it anymore. So we can now estimate it (**no simultaneous equation bias**). So hence, what we can do with our bivariate VAR(1) model that we have, is run 2 OLS regression and hence we get our parameter estimates. However, we may have the case of autocorrelation between the models. both GLS and OLS estimators will be **unbiased**.

The general form **n-VAR(p)** model:

$$x_t = \alpha + \Pi_1 x_{t-1} + \Pi_2 x_{t-2} + \dots + \Pi_p x_{t-p} + \varepsilon_t, \varepsilon_t \sim iid(0, \Sigma) \quad (36)$$

where x_t is $n \times 1$ vector at time t, Π matrix is $n \times n$, and the off-diagonal elements can be nonzero. Π is $n \times n$ size as for each time period, n variables can have feedback effect with each other.

So if we set the effect from all other variables to 0, then we are left with an AR model. We can have n variables and believe that some variables affect the other variables. Since there is no current endogenous variables on the RHS of equations, the specification in reduced form gives us a model that **does not impose any restrictions** ahead of time. Also, because the set of regressors is the same for each equations, the AVR model can be **efficiently estimated by OLS**.

The lag selection in VAR, we do penalty on $(pn^2 + n)$, as both the number of equation and lag order complex the model.(formula see p8 ppt).

To summarise, in VAR, we select the optimal order of p by information criteria, and select the number of equation (the number of endogenous variables, based on Granger Causality test).

5.3 Granger Causality method for Prediction:

The VAR model allows to test whether any given variable in the system causes another variable.

Motivation: Is VAR better than AR? (Or are multi-variate time series models are better choice than univariate one ?) or Does the new variable Z does help to predict Y? To

address the question of *the usefulness of one or a set of additional variable* for predicting Y or a set of targets.

Structural AR model requires a restriction on the right hand side. We can just use OLS but if variables are different, then we have to use MLE. To identify the system, it means that if we have 2 endogenous variables with 2 equations so direction of causality may be hard to identify so this structural model will allow us to identify the causality via restrictions made.

Based on the premise that the *future cannot cause past*, so we can easily check if past cause future, with and without additional variable. As such, Granger causality **addresses the question of usefulness of one variable (or a set of variables) for predicting another variable (or a set of variables)**. We say Z does Granger causes Y (or Z is useful, in addition to past y, to predict y) if:

$$E(y_t|I_{y,z,t-1}) \neq E(y_t|I_{y,t-1}) \quad (37)$$

, where $I_{y,z}$ is the *information set* containing all **past** value of y and z (and any deterministic components), while I_y does not contain the past infor of z.

From this, If only one lag of z in equation, We do **two side t-test for z_{t-1}** in granger causality method. **H0: z doesnot granger cause Y**. Critical value is from t-distribution with degree of freedom = number of obs - number of estimated parameter (non-intercept) - 1(intercept). If we have no less than two lags of z in equation Z_{t-1}, Z_{t-2}, \dots , then we do **F test** for joint significance of $Z_{t-1}, Z_{t-2} \dots$.

The term causes should be interpreted with caution. The only sense in which z causes y is given in the formula above. In particular, it has NOTHING to say about '**contemporaneous causality**' between y and z, so it does not allow us to determine whether z_t is an *exogenous* or *endogenous* variable in an equation relating y_t to z_t , as the formula above has nothing to the about the relationship of y and z in the same period.

Note that **Granger causality method \neq Eager-granger method**: Granger causality is selected to 'forecast', while Eager-granger is explain if 'cointegration'.

6 Serial correlation of error term

Assumption 3: Exstrictly exogeneity We have the case that there is no correlation or relationship between the error term and the explanatory variables for all of time.

$$E(\epsilon_t|\mathbf{X}) = 0, \forall t \quad (38)$$

where \mathbf{X} passes though all time period, both past and future.

Assumption 5: No Autocorrelation: The error terms are uncorrelated over time.

$$Cov(\epsilon_t, \epsilon_s|\mathbf{X}) = 0, \forall t \neq s \quad (39)$$

Violation of the white noise assumption:

- Recall if linearity, no perfect collinearity and strictly exogenous, then OLS is unbiased.
- The violation of the assumption that errors are uncorrelated over time will not (directly) affect unbiasedness or consistency of least squares. But **less efficient!** Because the Gauss-Markov Theorem requires serially uncorrelated errors, OLS is **no longer BLUE** in the presence of serial correlation.
- Autocorrelated errors (serial correlation) means that the **usual statistical inference unreliable**, even in the case of large samples. No efficient. Using the usual OLS standard error in the presence of serial correlation is invalid.
- Sereal correlation may **invalidate R square** or adjusted R square measures. If it is a result of **spurious regression** because y and (some of the) x have **unit roots**.
- But **R square are reliable in serial correlation case** if the data are **weakly dependent**. I.e. serial correlation itself does not invalidate R-squared. In fact, if the data are **stationary and weakly dependent**/
- If the explanatory variables are *contemporaneously exogenous and weakly dependent*, then OLS is still consistent (not necessarily unbiased).

6.1 Testing for Serial Correlation

We specify simple alternative models that allow the errors to be serially correlated and use the model to test the null that the errors are not serially correlated. From this, we can derive the **first-order autocorrelation / AR(1) serial correlation model**:

$$\epsilon_t = \rho\epsilon_{t-1} + v_t, v_t \text{ iid}(0, \sigma^2) \quad (40)$$

So that if $\rho = 0$, then that means the error term is just iid. The error term is a function of its own lag. We dont include an intercept due to zero conditional mean.

We can think of this as seeing whether does the previous error term have a relationship/effect on the current error term. If it does ($\rho \neq 0$), then there is autocorrelation as the error terms are related (correlated). The null hypothesis of this is that there is no serial correlation such that

H0: $\rho = 0$, errors are not serially correlated

Often though, $\rho > 0$, when there is serial correlation but we still use a two-sided alternative.

However, in practice, we **can not t actually observe error terms so we instead use the OLS residual $\hat{\epsilon}$** .

6.1.1 T-Test for AR(1) serial correlation under strict exogeneity:

If the **explanatory variables are strictly exogenous**, we can use a simple **t-test** for $H_0: \rho = 0$.

1. Set up a time series model and run the regression:

$$y_t = \beta' x_t + \epsilon_t$$

where $t = 1, 2, 3 \dots T$, and β is vector and x is also vector in multi-variate case.

2. **First-order correlation:** Using the residuals from step 1, run the regression

$$\hat{\epsilon}_t = \rho \hat{\epsilon}_{t-1} + v_t$$

where $t = 2, 3, 4 \dots$ (lost the first observation), and v_t is white noise iid. If we do not reject $H_0: \rho = 0$, and then $v_t = \epsilon_t$

3. Compute the **t-statistic** for $\hat{\rho}$ and test whether $H_0: \rho = 0$. Therefore if we reject, there is autocorrelation with the error terms and we cannot do inferences.

Remark:

- The test mentioned has **large-sample justifications** and tends to work well. Standard errors are wider if small time period T , and therefore we might not reject even if $\hat{\rho}$ is large since there is a lot of room for error. And Standard errors are smaller if large time period T , so much more likely to reject $\hat{\rho}$ even if its small. This is due to the fact that the **practical significance is nothing to do with the statistical significance**. It may be the case that for a large sample size, we just happened to have found some correlation (and therefore since sample size is large, its easy to reject things).
- Note that we must assume **homoskedasticity** for this test, if heteroskedasticity, then we use **robust t-statistics** (these are smaller standard errors so it is harder to reject).
- we can actually use the usual t-test as long as $E(\epsilon_t | x_t, x_{t+1}) = 0$ in AR(1) serial correlation case. (i.e. the error in a given time period is uncorrelated with regressors contemporaneously and in the next time period.) If we have higher order serial correlation, still need strict exogenous assumption.

6.1.2 F-Test for AR(p) serial correlation under strict exogeneity:

If we find v_t is not white noise in AR(1) process (i.e. higher order correlation exists), we shall consider **AR(p) for serial correlation**. In step two,

$$\hat{\epsilon}_t = \rho_1 \hat{\epsilon}_{t-1} + \rho_2 \hat{\epsilon}_{t-2} + \dots + \rho_q \hat{\epsilon}_{t-q} + v_t$$

where $t = q+1, q+2, \dots$ (we had to loss q observations!)

Then, in step three, we compute the F test for joint significance of $\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-q}$, use **F statistic** to test the joint hypothesis that

$H_0: \rho_1 = \rho_2 = \dots = 0$, reject H_0 means **jointly statistically significant**.

The test require homoskedasticity assumption, otherwise, a heteroskedasticity-robust version can be applied.

Note that if we are using just a single lagged residual, we lose one observations since if we had 100 observations, we can only take 99 lag variables and therefore we cant test it on one observation so we drop that observation.

6.1.3 Breusche-Godfrey Test under weak exogeneity:

When **weak exogeneity** happen the $\hat{\rho}$ in previous formula is biased, and we need to correct it. A simple adjustment is to consider these *endogenous regressors* when testing for serial correlation.

What we can do is to include all of the explanatory variables from step 1 to the residual autocorrelation regression in step 2 and then run the **F-statistic** again. Note we **include the intercept** γ_0 :

$$\hat{\epsilon}_t = \rho_1 \hat{\epsilon}_{t-1} + \rho_2 \hat{\epsilon}_{t-2} + \dots + \rho_q \hat{\epsilon}_{t-q} + \gamma_0 + \gamma_1 z_{1,t} + \gamma_2 z_{2,t} + \dots + \gamma_j z_{j,t} \quad (41)$$

Then, in step three, we use **F statistic** that

$H_0: \rho_1 = \rho_2 = \dots = 0$

F test by Lagrange multiplier form for statistic. **LM statistic**:

$$LM = (n - q)R^2$$

where R^2 is just usual R-squared from regression. Under H_0 , LM follow Chi-square dist with q d.f. It also can be made robust to heteroskedasticity **Note that we definitely incorporate an intercept here**, this test called **Breusche-Godfrey Test for q^{th} order residuals**

6.2 Correcting Serial Correlation

Let us suppose we know model suffers from serial correlation, although the estimate is still unbiased and consistent, but **inefficiency/ not BLUE**, as a result, our standard errors will be off meaning that we can no longer carry out inferences. From that, we should correct the serial correlation.

6.2.1 AR(1) serial correlation with static model:

Assume we have a linear regression with serially correlated errors and strict exogeneity assumption holds:

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t, \quad t = 1, 2, \dots, T$$

$$\epsilon_t = \rho \epsilon_{t-1} + v_t, \quad v_t \sim iid(0, \sigma^2)$$

and ϵ_t and v_t are assumed to be **independent error process**. Assume error follow AR(1) process.

Then,

$$\tilde{y}_t = \tilde{\beta}_0 + \beta_1 \tilde{x}_t + \tilde{v}_t, \quad v_t \sim iid(0, \sigma_v^2), \quad t = 2, 3, \dots, T \quad (42)$$

This equation **satisfied all Causs-Markov assumptions**. Note that we lost one obs here. and \tilde{y}_t, \tilde{x}_t are called **quasi-difference variables**

Proof 9

$$\tilde{y}_t = y_t - \rho y_{t-1}$$

$$\tilde{x}_t = x_t - \rho x_{t-1}, \quad (t = 1, 2, \dots, T)$$

$$\tilde{y}_t = \tilde{\beta}_0 + \beta_1 \tilde{x}_t + \tilde{\epsilon}_t, \quad (t = 2, 3, \dots, T)$$

$$\tilde{\beta}_0 = (1 - \rho)\beta_0$$

$$v_t \sim iid(0, \sigma_v^2)$$

Note that if $\rho = 1$, we have then just differenced it (not quasi), but we assumed that its absolute vakue less than 1 . Also, note that this only works for the observations from 2,...,t, since we cannot difference the first period. Therefore, for estimators to be **BLUE**, we need to specify that the first time period is simply:

$$y_1 = \beta_0 + \beta_1 x_1 + \epsilon_1$$

and all the $corr(v_t, \epsilon_1) = 0, \forall t$.

6.2.2 Cochrane-Orcutt Correcting Serial Correlation (Feasible GLS)

1. Regress y_t on x_t and obtain the residuals $\hat{\epsilon}$

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t, \quad t = 1, 2, 3, \dots, T$$

2. Regress $\hat{\epsilon}_t$ on $\hat{\epsilon}_{t-1}$ and obtain $\hat{\rho}$

$$\hat{\epsilon}_t = \rho \hat{\epsilon}_{t-1} + v_t$$

3. From **quasi-differenced variables**, we regress on to obtain parameter estimates

$$\tilde{y}_t = \tilde{\beta}_0 + \beta_1 \tilde{x}_t + \tilde{\epsilon}_t, \quad (t = 2, 3, \dots, T)$$

Remark:

- The **usual standard errors, t statistics, F statistics** are now **asymptotically valid**.

- The cost of using $\hat{\rho}$ in place of ρ is that the estimator is **biased**, however, it is **asymptotically more efficient** (assuming the time series are weakly dependent) and consistent. $\hat{\rho}$ is **consistent** estimator of ρ , since the whole procedure results in a **feasible GLS**. The FGLS is biased, but consistent. In other words, we give up the unbiased property, to get the estimator is asymptotically more efficient than OLS.
- By FGLS, our **standard error will be corrected and larger**, make inference valid.
- Note that, in Cochrane-Orcutt procedure, we need **strictly exogenous**, while OLS only need weakly exogenous to be consistent.
- In practice the Cochrane-Orcutt procedure is an iterative procedure, and steps (2) and (3) are executed repeatedly to derive new parameter estimates until there is no change in the estimate of ρ obtained from successive iterations.
- More importantly, we may notice that these the Cochrane-Orcutt method may result in coefficients that do not differ too much from OLS, BUT the standard errors can be significantly higher to account for the serial correlation (That is the good news and what we want from CO procedure).
- Note that we **cannot compare the R^2** of CO result with the OLS models since the dependent and independent variables are now different.

6.2.3 AR(1) serial correlation in FDL(1) model:

$$y_t = \alpha + \beta z_{t-1} + \epsilon_t$$

$$\epsilon_t = \rho \epsilon_{t-1} + \nu_t$$

Combine them, we get

$$y_t = (1 - \rho) \alpha + \rho y_{t-1} + \beta z_{t-1} - \beta \rho z_{t-2} + \nu_t$$

The combined model with a lag of z and AR(1) serial correction is a **special case of a general model**

$$y_t = \delta + \rho y_{t-1} + \gamma_1 z_{t-1} + \gamma_2 z_{t-2} + \nu_t$$

where $\delta = (1 - \rho) \alpha$, $\gamma_1 = \beta$, $\gamma_2 = -\beta \rho$, and the coefficient of y_{t-1} is same (we have said that the serial correlation does not cause bias of inconsistency here, but less efficient, OLS estimator not BLUE). Also find that γ_2 is entirely determined by β and ρ .

For forecasting, the AR(1) serial correlation model may be too **restrictive**. As it may impose restrictions on the parameters that are not met. On the other hand, if the AR(1) serial correlation model holds, it captures the conditional mean $E(y_t | I_{t-1})$ with one fewer parameter than the general model; i.e. AR(1) is more **parsimonious** than the general model (AR(1) is suggested). AR(1) is more parsimonious than general model here means: given the same explanatory power (they have same regressors), It requires less parameters to estimate (β and ρ only), compared with general model (ρ, γ_1, γ_2), note that we ignore the intercept term here, as it is just a drift. Intuitively, less estimated parameter is

always better, as less estimation, less error we may occur (also less assumptions needed). And can help to avoid overfitting issue. It is related to **Occam's razor** (the less/least speculation is usually better. Another way of saying it is that the more assumptions you have to make, the more unlikely an explanation.)

7 Cointegration and Error Correction Model

7.1 Cointegration

Cointegration: when two (or more) **I(1)** series, but linear combination is **I(0)**. In Econometric, Cointegration means **Long term equilibrium/co-move in the long run** exists! In such case, we say y and z are *integrated*, i.e they do have a specific relationship, related to regularity in the long run.

Thus, the idea of test is: if we can find this β , s.t. η_t is **I(0)**? Since if y and z cannot be cointegrated, then it is impossible to find a stationary process for the error term.

Do not confuse it with Spurious case. Recall, the spurious series is when two unit root series are **independent**, but the regression of one on another yield statistical significant result (the standard inference fail in **I(1)** case). However, in Cointegration case, the regression of one on another is not spurious, but instead tells us something about the **Long-run relationship**.

Also Note that we get inconsistent estimator if both y_t and x_t are IID random walk process. In fact, we cannot even know the distribution of estimator. However, we now can obtain a **consistent** estimator if those random walk processes time series are cointegrated.

Suppose y_t and z_t are two **I(1)** time series, given by:

$$y_t = y_{t-1} + \epsilon_t, \epsilon_t \sim iid(0, \sigma_\epsilon^2) \quad (43)$$

$$z_t = z_{t-1} + \nu_t, \nu_t \sim iid(0, \sigma_\nu^2) \quad (44)$$

ϵ_t, ν_t are white noise process, generally, if no linear relationship, any **linear combination** bet them are still **I(1)**. Note that it just say '**linear combination**', **does not imply independent**, as they also can have non linear relationship.

But, in some cases, we can find a slope β , s.t. $y_t - \beta z_t$ can come to **I(0)**, white noise process.

For example, when $\beta = 1$, two RW process are cointegrated if y - z come to **I(0)**, that is, a linear combination of the two RW processes has a **tendency of returning to zero** with some regularity.

7.2 EngleSX Granger test for Cointegration:

Recall that if $\{y_t\}$ and $\{z_t\}$ are two **I(1)** process, then in static model without cointegration, error term still **I(1)**. Thus, Cointegration thus similar with Dickey-Fuller test on residuals from static model, but these two tests do has different critical value. as the Granger test account for estimation of the **cointegration parameter** and we use the **Asymptotic critical value for cointegration test**.

H0: Z and Y has are no co-integrated. i.e. Regression residual is **I(1)**.

H1: Z and Y has are co-integrated. i.e. residual can be **I(0)**, there is long run relationship between them.

$$y_t = \alpha + \beta z_t + \eta_t \quad (45)$$

1. Regress y on z ($\hat{\beta}$ here is consistent with β if cointegrated). Otherwise, we need to be careful with the standard error of β because we haven't met the condition of strict exogeneity. So the β would be consistent but t -statistic may not be correct.
2. Get the residuals of this estimated static model.
3. Check the unit root in the residuals and use the asymptotic critical value for cointegration test. (Note that don't use Dickey-Fuller distribution, shall use cointegration table instead)

7.3 Error Correction Model for Cointegration

The Cointegration between two series implies *error correction model*.

Let γ be the **adjustment parameter** and $\gamma < 0$, the larger of the absolute value of γ , the faster to the adjustment long run equilibrium.

Let the cointegrating relationship bet the **two I(1) series** is

$$\eta_t = y_t - \alpha - \beta z_t \quad (46)$$

Note that it just the formula in Granger test, where η_t is $I(0)$, stationary. In short run, the error correction model is:

$$\Delta y_t = \delta + \theta \Delta z_t + \gamma \eta_{t-1} + \epsilon_t \quad (47)$$

We can also incorporate the dynamic properties of relationship:

$$\Delta y_t = \delta + \theta_1 \Delta z_t + \theta_2 \Delta z_{t-1} + \gamma \eta_{t-1} + \epsilon_t$$

where $\gamma \eta_{t-1}$ is the **error correction term**, which is **stationary** (by construction, unless it is not, in which case we don't have cointegration) and lead the result show long run relationship, the difference term of y and z also turn into $I(0)$. We can think γ as an omitted variable, and add information to adjust $I(1)$ problem. the **derivation** from the equilibrium in the previous period is measured by η_{t-1} . And if η_{t-1} close to zero, it means the current point is close to the long - run equilibrium. If we impose $\gamma = 0$, then the error correction model reduces to a static model in first difference!

The interpretation of error correction model **proof 6** is regress Δy_t on Δz_t and η_{t-1} , and $\eta_{t-1} = y_{t-1} - \alpha - \beta z_{t-1}$, which is the static model on two $I(1)$ series. and get the η_{t-1} be the white noise. The issue is we do not know β , so we need to find $\hat{\beta}$ by Enger Granger method.

If $\eta_{t-1} > 0$, positive derivation, then $\gamma \eta_{t-1} < 0$. In this case, the current position is above the equilibrium, so next period t , we subtract the $\gamma \eta_{t-1}$ in error correction model, let y_t decrease.

If $\eta_{t-1} = 0$, the model reduces to static model in first difference. i.e. if we already in equilibrium, we can trust static model with first difference. But generally we consider

estimate η , as we do not know if we come to the equilibrium already.

The error correction is the **reparametrization of General dynamic models**:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 z_t + \beta_1 z_{t-1} + \epsilon_t \quad (48)$$

The lagged first difference is dynamic properties of relationship. but note that this equation contains **nonstationary variables**, while the error correction model with delta does not.

Proof 11: convert error correction model to General dynamic model

P 18 - 19 Vector Error Correction model

8 Forecast

9 Appendix: Basic Knowledge

- **consistent:** Intuitively, it means that we can get a constant estimate of an estimator when n goes to infinity, as the standard error goes to zero. The constant is a number not very (variance is zero).
- **unbiased:** On average, it's true, but does not mean unbiased estimator is better than other estimator, for example, if the variance of unbiased estimator is relative large.
- **Random Variable:** RV is just a *function!* that map the sample space/events to the real line (\mathbb{R}). We need define RV as we need convert 'event' to data, and then can work with it. For example, dealing with unstructured data.
Sample space is a set of all possible outcomes.
Outcome is only one, but *event* can be a collection of several outcomes.
- **Parameter and Estimator:** In classical statistics, parameter is the true and unknown constant, based on the population. As it is unknown, we need to estimate it. The estimator is a RV $\hat{\theta}(X)$, and estimate is an estimated value based on the data/sample we have. And from this., we have distribution of estimator and want n goes to infinity.
- **Deterministic Model VS Stochastic Model:** Deterministic model is **fully** determined by the parameter on values and initial conditions. But statistic model has **some inherent randomness**, the same parameter and initial condition can give different result.