# Time Series Analysis

## 1. Intuition

·问题：利用OLS对多元线性回归模型进行估计的估计值 何时无偏最优？

· MLR1.Linearity: Linear in parameter;

是线性模型

· MLR2. Random Sampling; ——**样本随机选择**

· MLR3. No Perfect Collinearity——**各个系数之间不存在完全共线性**

· MLR4. Zero Conditional Mean——

· MLR5. Homoskedasticity/Homoscedasticity—— **同方差性**

· MLR6. Normality of Error Terms——**残差项服从正态分布。若数据量足够大，**

·**MLR1 – MLR4:** Unbiased，满足条件1-4，利用OLS估计得出的参数是无偏的。

·**MLR1 – MLR5:** “BLUE”-Best Linear Unbiased Estimator 满足条件1-5，利用OLS估计得出的参数是无偏且最优的。

·**MLR1 – MLR6:**“Classical Linear Assumptions” 满足条件1-6，经典线性回归模型的基本假设。

·为什么要无偏最优？

## 2. Basic Regression Analysis with Time Series Data

### 2.1 The Nature of Time Series Data

·Obvious Characteristic：Temporal Ordering→Past can affect the future, but not vice versa.

·More subtle difference: Cross-Sectional→Samples are randomly draw from the appropriate population：来自于不同总体的样本自然会产生不同的变量。

·Randomness in Time Series Data: Being outcomes of random variables.

·虽然时间序列数据也满足随机变量直观上的定义（其结果是未知的），但是与Cross-Sectional数据相比，我们只能观察到该时间序列中的一种结果。

·A sequence of random variables indexed by time is called **Stochastic Process** or a **Time Series Process**. When we select a time series dataset, we obtain one possible realization of the stochastic process. **We can only see a single realization, because we cannot go back in time and start the process over again.**

**·**如果情况不同，会有不同的结果**→Outcome of random variables**

### 2.2 Examples of Time Series Regression Models

·Static Models：y与z同期:

Usually, a static model is postulated when a change in z at time t is believed to have an immediate effect on y:, when

·Finite Distributed Lag Models(FDL, 有限分布滞后模型):

one or more variables to affect y with a lag:

(FDL of Order 2)

**·Lag Distribution**→summarize the dynamic effect that a temporary increase in z has on y.

·**Impact Propensity**(即期倾向)：只有，前后都为。其参数被称之为**Impact Propensity。**

·Permanent Increase:当关注于某节点前后自变量永久的增加时，方程系数之和被称之为：**Long-Run Propensity (LRP)/ Long-Run Multiplier**

·**A Finite Distributed Lag Model(FDL) of order q is written as:**

·A primary purpose: Test whether z has a lagged effect on y.

·由于z的滞后项之间往往会存在多重共线性，因此精确估计得到每一个的值是不可能的。但是我们可以估计得到比较好的LRP值。

·**A Convention about the Time Index**→Initial observations：the first time period being represented by the regression equation.

### 2.3 Finite Sample Properties of OLS under Classical Assumptions

·How the assumptions must be altered from our cross-sectional analysis to cover Time Series Regressions.

·Unbiasedness of OLS

Assumption TS.1 Linear in parameter

Assumption TS.2 No Perfect Collinearity

**Assumption TS.3 Zero Conditional Mean:** the error at time t, is uncorrelated with each explanatory variable in every time period.

·Contemporaneously exogenous:

·Strictly Exogenous→OLS is Unbiased. 严格外生假设，即残差项不与自变量的任何滞后项相关。

·截面数据中，由于随机采样的假设，我们并不明确地假设残差项与每个个体的属性相关；由于时间序列数据并不是随机采样的结果，所以需要明确假设残差项与自变量的任何滞后项无关。

·静态模型中，遗漏某自变量的之后效应可能会违背上述假设

·FDL中，虽然控制了主要自变量与残差项是严格外生的（农业生产函数中的降雨量都不会受到产量的影响），残差项（一些未被考量的因素，如农业生产函数中的劳动力会输入受到过去的产量的影响，存在自选择与双向因果的变量）仍有可能对于未来的自变量有影响从而导致违背了严格外生的假设，换言之，严格外生的自变量无法对过去已经发生的效应作出反应。很多政策变量往往是动态调节的，受到过去结果的影响。

·Unbiasedness of OLS

·The Variance of the OLS Estimators and the Gauss-Markov Theorem

Assumption TS.4 Homoskedasticity: Conditional on X, the variance of is the same for all t:

Assumption TS.5 **No Serial Correlation**

Assumption TS.6 The errors are independent of X and are independently and identically distributed as Normal(0,)

### 2.4 Functional Form, Dummy Variables, and Index Numbers

·Short-Run Elasticity

·Long-Run Elasticity

·Binary or Dummy independent variables→A certain event has occurred.

·Event Study→Whether a particular event influences some outcome.

·The notion of an **index number** and the difference between nominal and real economic variables: Base Period and the best value

·太多的虚拟变量会损害变量的自由度(Degree of Freedom)

### 2.5 Trends and Seasonality

·Characterizing **Trending Time Series**:

·Time Trend: Two time series processes appear to be correlated only because they are both trending over time for reasons related to other unobserved factors.

·What kind of statistical models adequately capture trending behavior?

·A linear time trend:

·CLM Assumption: is linear in t；trending time series allows {} to be correlated over time.

·Exponential Trend: Series has the same average growth rate from period to period.

·If follows (1), , for all t.

·More complex form(e.g. Quadratic Form)

·Using Trending Variables in Regression Analysis:

·Trending Variables 并没有违背TS.1 –TS.6，但是要小心没有被观测到的具有时间趋势的影响因变量的因素会与解释变量相关，从而导致残差与解释变量相关而导致内生性问题。

·Spurious Regression Problem: 寻找多个具有时间趋势变量之间的关系。

·How omitting a time trend can result in Spurious Regression.

·In some cases, adding a time trend can make a key explanatory variable more significant(Different kinds of trends)

·重点是捕捉主效应的影响，所以不适合将太多的时间变量的多项式加入到模型之中。

·A Detrending Interpretation of Regressions with a Time Trend.

·Conclusions are much more general.

·Regress on ,,and t→Fitted Equation:

·趋势变量t何时被加入到模型之中？趋势变量在统计学学上显著。

·Computing R-Squared when the Dependent Variable Is Trending

·R-squareds in time series regressions are often very high, especially compared with typical R-squareds for cross-sectional data.

·Does this mean that we learn more about factors affecting y from time series data?

·On one hand, time series data often come in aggregate form，and aggregates are often easier to explain than outcomes on individuals, families, or firms, which is often the nature of cross-sectional data.

·the usual and adjusted R-squareds for time series regressions can be artificially high when the dependent variable is trending. Remember that R2 is a measure of **how large the error variance is relative to the variance of y.**

·Seasonality

·Seasonality: a time series is observed at monthly or quarterly intervals.

·**Seasonally adjusted:** series that do display seasonal patterns are often seasonally adjusted before they are reported for public use. A seasonally adjusted series is one that has had the **seasonal factors removed from it.**

**·seasonal dummy variables**

·Interaction

**·Deseasonalizing**

## Further Issues in Using OLS with Time Series Data

### 3.1 Stationary and Weakly Dependent Time Series

·Large Sample Approximation in regression analysis with time series data

·**Stationary and Nonstationary Time Series**: Probability distributions are stable over time: If we take any collection of random variables in the sequence and then shift that sequence ahead h time periods, the joint probability distribution must remain unchanged.

·Stationary:两相邻时间解释变量的联合分布概率不会随着时间的改变而改变。平稳性的定义要求邻接项的相关性都相似。换言之，其要求的是邻接项之间的关系，而对邻接项内部，两个时间节点的解释变量的关系并没有做特殊要求。

·Covariance Stationary Process: A stochastic process with a finite second moment is covariance stationary if (i) is constant; (ii) is constant; and (iii) for any , depends only on h and not on t.

·函数的形式反映了假设：边际效应不会随着时间发生改变，而根据TS假设，其方差与邻接项之间的相关性为0，也是常数。

·**Weakly Dependent Time Series：Covariance Stationary Process,**

.

·Why is weak dependence important for regression analysis

## Serial Correlation and Heteroskedasticity in Time Series Regressions

## Advanced Topics

·A unit root implies that a shock today has a long-lasting impact.

·**Spurious Regression:** Two time series process → Even if two unit root series are independent, it is quite likely that the regression of one on the other will yield a statistically significant t statistic.

·Cointegration

·How regression methods can be used to forecast future outcomes of a time series.

### 5.1 Infinite Distributed Lag Models

·Infinite Distributed Lag Models(IDL):

·We only observe a finite history of data, equation involves an infinite number of parameters.→We cannot estimated it without some restrictions.

·The Geometric (or Koyck) Distributed Lag Models(IDL):

·Rational Distributed Lag Models(IDL):