Time Series

Junhui Qian

December 2, 2012

Outline

- Introduction
- Conditional Mean
  - Serial Correlation
  - AR, MA, and ARMA Models
- Conditional Variance
  - Volatility Clustering
  - ARCH/GARCH Models
- Time Series Regression

What is time series

- Time series is a sequential collection of random variables.
- Example: GDP, CPI, unemployment, interest rate, foreign exchange rate, electricity usage, stock market index, EPS (earning per share) of a stock, yearly income of a household, etc.
- While cross-section data is a "sample" of population, time series is a "tracking" of an chosen individual (household, province, country, etc.).
- Time series is also called "stochastic process". A time series data is a "realization" of a stochastic process.

Stationary Time Series

- A stationary time series has a stable correlative structure.
- Weak stationarity.

$$\mathbb{E}x_t = \mu, \quad \mathrm{cov}(x_t, x_{t-\tau}) = \gamma_\tau$$

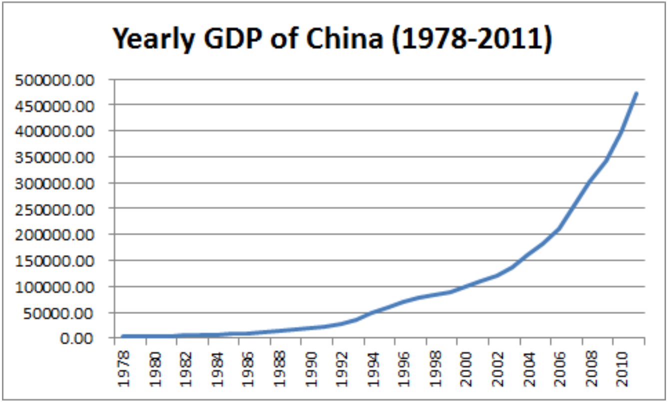
- Strict stationarity.

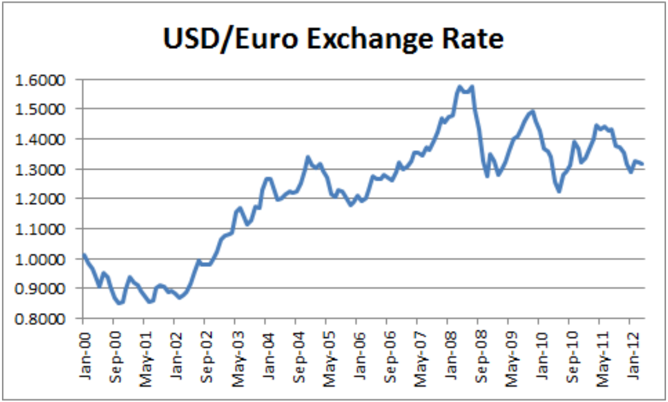
$$F(X_t, ..., X_T) = F(X_{t+\tau}, ..., X_{T+\tau}),$$

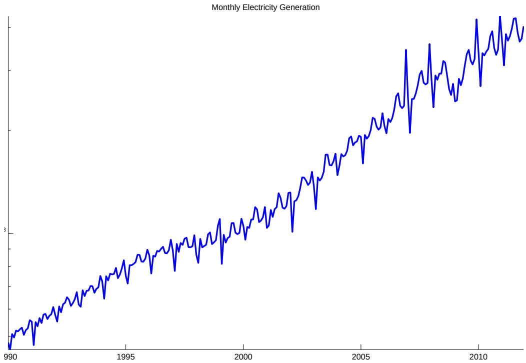
where F is the joint distribution function.

Nonstationary Time Series

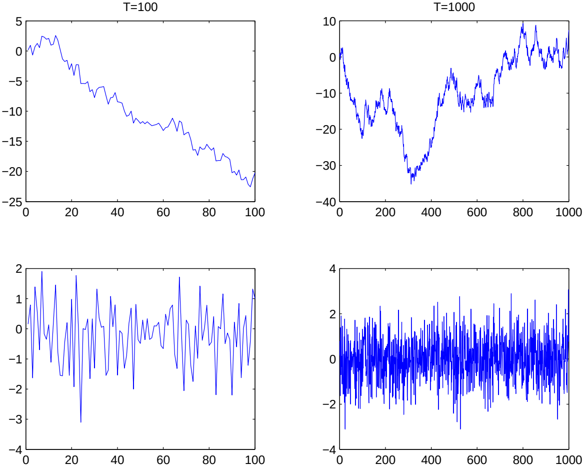
- Trend process (Yearly GDP, population, etc.)
- Unit root process (Interest rate, foreign exchange rate, stock prices, etc.)
- Seasonal process (Quarterly agricultural output, monthly monetary base, quarterly GDP, quarterly EPS of a stock, etc.)


Yearly GDP of China (1978-2011)
500000.00
450000.00
400000.00
350000.00
300000.00
250000.00
200000.00
150000.00
100000.00
50000.00
0.00
1978
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010


USD/Euro Exchange Rate
1.6000
1.5000
1.4000
1.3000
1.2000
1.1000
1.0000
0.9000
0.8000
Jan-00
Sep-00
May-01
Jan-02
Sep-02
May-03
Jan-04
Sep-04
May-05
Jan-06
Sep-06
May-07
Jan-08
Sep-08
May-09
Jan-10
Sep-10
May-11
Jan-12


Monthly Electricity Generation
990
1995
2000
2005
2010

The Illusion of Pattern



White Noise

- Weak white noise (ε_t).

$$\mathbb{E}\varepsilon_t = 0, \quad \mathrm{var}(\varepsilon_t) = \sigma^2, \quad \mathrm{cov}(\varepsilon_t, \varepsilon_{t-\tau}) = 0 \text{ for } \tau \neq 0.$$

- Strong white noise.
- Gaussian (strong) white noise.

$$(\varepsilon_t) \sim i.i.d. \;\; N(0, \sigma^2)$$

- White noise is often used to characterize shocks.

Describing Serial Correlation

- Time series differ from cross-section data in that the former usually exhibit serial correlation, ie, the conditional mean of future variable depends on past observations.
- To examine serial correlation graphically, we use
  - Scatter plots
  - Correlogram (Autocorrelation function, partial Correlation)

ACF (AutoCorrelation Function)

The k-th order ACF is defined as

$$\rho_k = \frac{\mathrm{cov}(x_t, x_{t-k})}{\mathrm{var}(x_t)} = \frac{\gamma_k}{\gamma_0}.$$

PACF (Partial AutoCorrelation Function)

The k-th order PACF is defined as the coefficient δ_k in the following regression model,

$$x_t = \delta_0 + \delta_1 x_{t-1} + \delta_2 x_{t-2} + \cdots + \delta_k x_{t-k} + \varepsilon_t.$$

- PACF(k) measures the "direct" impact of x_{t-k} on x_t.
- Obviously, x_{t-k} also intermediately affect x_t through $x_{t-k+1}, \ldots, x_{t-1}$.

Modeling Conditional Mean

- For stationary time series, typical models for conditional mean are the following,
  - MA (Moving Average)
  - AR (Auto-Regression)
  - ARMA (AutoRegressive Moving Average)
- Stochastic processes characterized by these models are called linear processes.

AR(1) Model

A zero-mean first order AR process, $x_t \sim AR(1)$, if

$$x_t = ax_{t-1} + \varepsilon_t,$$

where (ε_t) is a white noise and $a \in (-1, 1)$.

A Different Form

If $\mathbb{E}x_t \neq 0$, an AR(1) process can be represented as

$$x_t = c + a x_{t-1} + \varepsilon_t,$$

where (ε_t) is a white noise and $a \in (-1, 1)$.

Mean and Variance

- Let $\mu = \mathbb{E}x_t$, we have

$$\mu = \frac{c}{1-a}$$

- If var(ε_t)=σ^2,

$$\mathrm{var}(x_t) = \frac{\sigma^2}{1-a^2}.$$

- What if $|a| \geq 1$?

ACF and PACF

For an AR(1) process, it is easy to obtain

$$\mathrm{ACF}(k) = a^k,$$

and

$$\mathrm{PACF}(k) = \begin{cases} a, & k = 1 \\ 0, & k > 1 \end{cases}$$

AR(p) Model

A zero-mean p-th order AR process, $x_t \sim AR(p)$, if

$$x_t = a_1 x_{t-1} + \cdots + a_p x_{t-p} + \varepsilon_t,$$

where (ε_t) is a white noise and the roots of $1 - a_1 z - \cdots - a_p z^p = 0$ are outside the unit circle.

ACF and PACF of AR(p)

For a general AR(p),

- ACF exponentially decays to zero but never reaches zero.
- And there are finite number of nonzero PACF's.
- Question: $\mathrm{PACF}(1) = a_1$?
- AR models are suitable for those processes with "long" ACF and "short" PACF.

Estimation of AR Models

- OLS. Consistent, robust to the distribution of (ε_t), but may not be efficient.
- MLE. Most efficient under correct specification.
- Yule-Walker Equation

Yule-Walker Method

- The Yule-Walker Equation

$$\mathbb{E}[x_{t-k}(x_t - a_1 x_{t-1} - \cdots - a_p x_{t-p})] = 0, k = 1, ..., p$$

- AR(1)

$$\gamma_1 - a\gamma_0 = 0$$

- AR(2)

$$\begin{aligned} \gamma_1 - a_1\gamma_0 - a_2\gamma_1 &= 0 \\ \gamma_2 - a_1\gamma_1 - a_2\gamma_0 &= 0 \end{aligned}$$

Estimation of Covariance

- We estimate γ_k by

$$\hat{\gamma}_k = \frac{1}{n-k} \sum_{t=k+1}^{n} x_t x_{t-k}.$$

- As k increases, $\hat{\gamma}_k$ becomes less accurate.

Forecasting Based on AR(p)

- 1-step ahead forecast

$$\hat{x}_{n+1} = \hat{a}_1 x_n + \hat{a}_2 x_{n-1} + \cdots + \hat{a}_p x_{n-p}.$$

- 2-step ahead forecast

$$\hat{x}_{n+2} = \hat{a}_1 \hat{x}_{n+1} + \hat{a}_2 x_n + \cdots + \hat{a}_p x_{n-p+1}.$$

- $\vdots$

Forecast Error for 1-step Ahead Forecast

- If n is large, $\hat{a}_i \approx a_i$.
- Then

$$\hat{x}_{n+1} - x_{n+1} \approx \varepsilon_{n+1}$$

- Then

$$\mathrm{var}(\hat{x}_{n+1} - x_{n+1}) \approx \mathrm{var}(\varepsilon_{n+1}) = \sigma^2.$$

Interval Forecast

If, furthermore, we assume $\varepsilon_t \sim N(0, \sigma^2)$, then we obtain the 95% 1-step ahead interval forecast,

$$[\hat{x}_{n+1} - 1.96\hat{\sigma}, \hat{x}_{n+1} + 1.96\hat{\sigma}],$$

where $\hat{\sigma}$ is the estimate of σ,

$$\hat{\sigma}^2 = \frac{1}{n - 2p - 1} \sum_{t=p+1}^{n} \hat{\varepsilon}_t^2.$$

Forecast Error for 2-step Ahead Forecast

We have

$$\hat{x}_{n+2} - x_{n+2} \approx a_1 \varepsilon_{n+1} + \varepsilon_{n+2}.$$

Hence

$$\mathrm{var}(\hat{x}_{n+2} - x_{n+2}) = (1 + a_1^2)\sigma^2.$$

MA (Moving Average) Models

A q-th order Moving Average model, $x \sim MA(p)$, is defined as

$$x_t = \varepsilon_t + b_1 \varepsilon_{t-1} + \cdots + b_q \varepsilon_{t-q}.$$

ACF and PACF

- It is obvious that an MA(q) process has finite number of nonzero ACF's.
- It can be shown that an MA(q) process has infinite number of nonzero PACF's.
- Hence MA models are suitable for those processes with long PACF and short ACF.

The Estimation of MA Model

Take MA(1) as an example, $x_t = \varepsilon_t + b\varepsilon_{t-1}$, we have

$$\varepsilon_t = x_t - b\varepsilon_{t-1}.$$

- Nonlinear least square.
- MLE, assuming $\varepsilon_t \sim iid\ N(0, 1)$.

One-Step Ahead Forecasting Based on MA(1)

- Forecasting based on MA(1), $x_t = \varepsilon_t + b\varepsilon_{t-1}$,

$$\hat{x}_{n+1} = b\varepsilon_n,$$

  where ε_n must be estimated.
- $\mathrm{var}(\hat{x}_{n+1} - x_{n+1}) = \sigma^2$.

Estimation of ε_n

One procedure for estimating ε_n is as follows,

1. Assume $\varepsilon_0 = 0$,
2. $\varepsilon_1 = x_1 - \hat{b}\varepsilon_0$
3. $\varepsilon_2 = x_2 - \hat{b}\varepsilon_1$
4. $\vdots$
5. $\varepsilon_n = x_n - \hat{b}\varepsilon_{n-1}$

Two-Step Ahead Forecasting Based on MA(1)

We have

$$x_{n+2} = \varepsilon_{n+2} + b\varepsilon_{n+1}.$$

It is obvious that

$$\mathbb{E}(x_{n+2}|x_{n+1}, x_n, ...) = 0.$$

Hence

$$\hat{x}_{n+2} = 0.$$

ARMA Model

- ARMA model combines AR and MA together, the following is an ARMA(1,1) model,

$$x_t - a x_{t-1} = \varepsilon_t + b\varepsilon_{t-1}.$$

- When $|a| < 1$, the ARMA(1,1) process is stationary.
- ARMA(p,q) is similarly defined,

$$x_t - a_1 x_{t-1} - \cdots - a_p x_{t-p} = \varepsilon_t + b_1\varepsilon_{t-1} + \cdots + b_q\varepsilon_{t-q}.$$

- When the roots of $1 - a_1 z - \cdots - a_p z^p = 0$ are outside the unit circle, ARMA(p,q) is stationary.
- ARMA has long but rapidly declining ACF and PACF's.

The Estimation of ARMA Model

- Yule-Walker for the AR part.
- Nonlinear least square.
- MLE

Forecasts Based on ARMA

Suppose $x_t \sim ARMA(1,1)$,

$$x_t = a x_{t-1} + \varepsilon_t + b\varepsilon_{t-1}.$$

- The one-step-ahead forecast is

$$\hat{x}_{n+1} = \hat{a} x_n + \hat{b}\hat{\varepsilon}_n,$$

- The variance of $\hat{x}_{n+1}$ is given by

$$\mathrm{var}(\hat{x}_{n+1} - x_{n+1}) = \sigma^2$$

- Homework: write the two-step-ahead forecast based on ARMA(1,1) and calculate its variance.

The Selection of Order

- Use significant tests of parameters.
- Information criteria.
  - BIC (Bayes Information Criterion)

$$\mathrm{BIC}(p, q) = \log\left(\frac{SSR(p, q)}{n}\right) + (p + q + 1)\frac{\log n}{n},$$

    where $SSR(p, q)$ is the SSR of the estimated model.
  - AIC (Akaike Information Criterion)

$$\mathrm{BIC}(p, q) = \log\left(\frac{SSR(p, q)}{n}\right) + (p + q + 1)\frac{2}{n}.$$

Diagnostics of ARMA Models

- After estimating an ARMA model, we should check whether the assumptions we made are valid. In particular, we should check whether the residuals are indeed a white noise.
- Correlogram
- Ljung-Box test, also called Q test,

$$Q = n(n+2)\sum_{j=1}^{m}\frac{\hat{\rho}_j^2}{n-j},$$

where n is the sample size, $\hat{\rho}_j$ is the jth-order sample autocorrelation, and m is the number of lags being tested. Under the null hypothesis (white noise), $Q \sim \chi^2_m$.

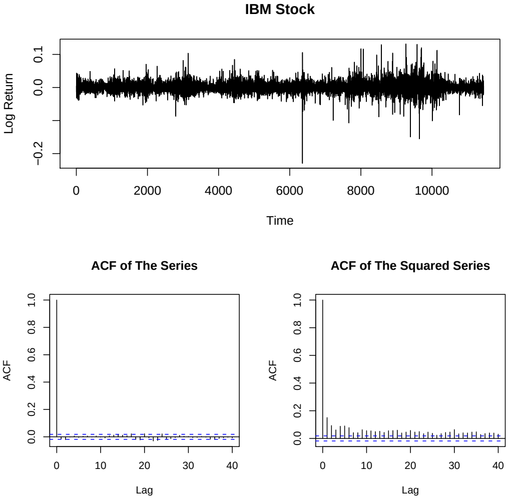
Why do we care about conditional variance?

- For many types of time series (stock return, interest rate variation, foreign exchange rate variation, etc.), conditional variance is a measure of risk facing investors.
- Conditional variance is essential for interval forecast.
- Conditional variance is essential for pricing derivatives.

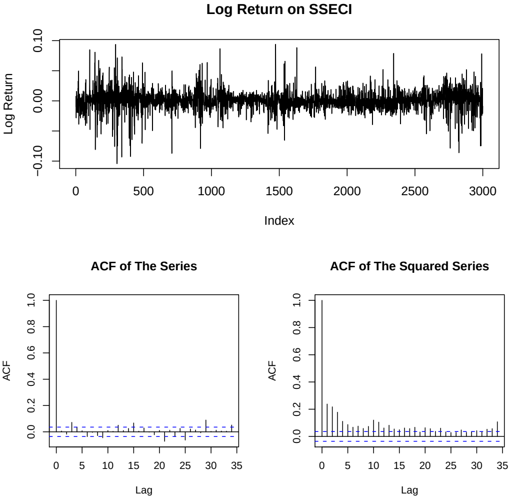
Some Stylized Facts of Financial Returns

- There exists little or weak linear serial correlation in asset returns.
- The marginal distribution of asset returns is usually skewed and peaked (heavy-tailed).
- There exists volatility clustering, ie, big changes follow big changes.

The Log Return of IBM Stock



The Log Return on SSECI



The ARCH Model

- ARCH models are introduced to describe volatility clustering.
- An ARCH(p) white noise is defined as

$$u_t = \sigma_t \varepsilon_t,$$

where ε_t is an iid white noise and h_t satisfies

$$\sigma_t^2 = c + a_1 u_{t-1}^2 + \cdots + a_p u_{t-p}^2, \tag{1}$$

where $c > 0$, $a_i \geq 0$ for all i, and $\sum_i a_i < 1$.

Volatility Clustering

- $\sigma_t^2 = \mathrm{var}(u_t | u_{t-1}, u_{t-2}, ...)$ is the conditional variance of u_t, given information available at time $t-1$.
- For example, in the following ARCH(1) model, σ_t^2 evolves in a "Markovian" manner,

$$\sigma_t^2 = c + a u_{t-1}^2,\ c > 0,\ 0 < a < 1.$$

- From the dynamic equation of σ_t^2, we can see that big shocks produce big volatilities ahead. Or, big changes tend to follow big shocks.
- Conditional variance is also called conditional heteroscedasticity, ARCH is in fact "AutoRegressive Conditional Heteroscedasticity".

Simulated ARCH(1) Processes

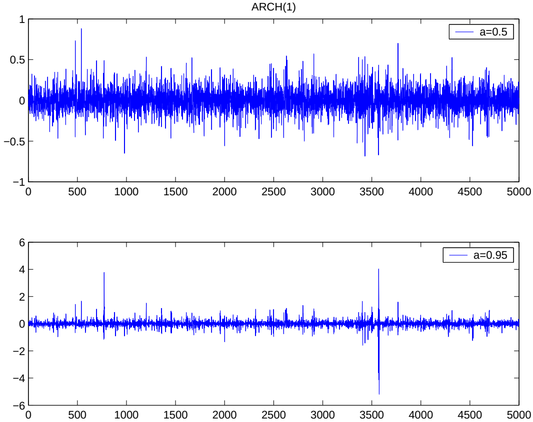


Figure : DGP: $u_t = \sigma_t \varepsilon_t$, where $\varepsilon_t \sim iidN(0,1)$ and $\sigma_t^2 = 0.01 + a u_{t-1}^2$. The upper panel sets $a = 0.5$ and the lower panel sets $a = 0.95$.

Properties of ARCH Processes

(a) Let $\eta_t = u_t^2 - \sigma_t^2$. (η_t) is a martingale difference sequence, ie, $\mathbb{E}(\eta_t | u_{t-1}, u_{t-2}, ...) = 0$.

(b) (u_t^2) is an AR(p) process.

(c) (u_t) is a white noise with a (unconditional) variance of $\mathrm{var}(u_t) = c/(1 - a_1 - \cdots - a_p)$.

(d) The (unconditional) distribution of u_t is of heavy tails, ie, the kurtosis of u_t is greater than 3.

Heavy Tail

- We can show that

$$\text{Kurtosis}(u_t) = \frac{\mathbb{E}u_t^4}{(\mathbb{E}u_t^2)^2} = 3\frac{1-a^2}{1-3a^2} > 3.$$

- For kurtosis to be positive, a must satisfy $a \in [0, \sqrt{3}/3)$.
- As a increases, the tail becomes heavier.
- When $a \geq \sqrt{3}/3$, the kurtosis ceases to exist.

GARCH Model

An order-(p, q) GARCH model, or GARCH$(p,\ q)$, is defined as,

$$u_t = \sigma_t \varepsilon_t,$$

where

$$\sigma_t^2 = c + a_1 u_{t-1}^2 + \cdots + a_p u_{t-p}^2 + b_1 \sigma_{t-1}^2 + \cdots + b_q \sigma_{t-q}^2.$$

- $c > 0$, $a_i \geq 0$, $b_i \geq 0$ for all i.
- $\sum_{i=1}^{\max(p,q)} (a_i + b_i) < 1$.
- GARCH generalizes ARCH, hence its name, Generalized ARCH.

Properties

(a) Let $\eta_t = u_t^2 - \sigma_t^2$. (η_t) is a martingale difference sequence, ie, $\mathbb{E}(\eta_t | u_{t-1}, u_{t-2}, ...) = 0$.

(b) (u_t^2) is an ARMA(p, q) process,

$$u_t^2 = c + \sum_{i=1}^{\max(p,q)} (a_i + b_i) u_{t-i}^2 + \eta_t - \sum_{i=1}^{q} b_i \eta_{t-i}.$$

(c) u_t is a white noise, with an (unconditional) variance of $\mathrm{var}(\omega_t) = c/(1 - \sum_{i=1}^{\max(p,q)} (a_i + b_i))$.

(d) The (unconditional) distribution of ω_t is of heavy tails.

Estimation of ARCH and GARCH Models

- MLE
- Methods applicable to the estimation of ARMA models.

Forecasting Variance: One Step Ahead

For one-step-ahead forecast, we have

$$\sigma_{n+1}^2 = c + a_1 u_n^2 + \cdots + a_p u_{n-p+1}^2 + b_1 \sigma_n^2 + \cdots + b_q \sigma_{n-q+1}^2,$$

where $(u_n^2, ..., \omega_{n-p+1}^2)$ and $(\sigma_n^2, ..., \sigma_{n-q+1}^2)$ are known at time n. Note that the one-step-ahead forecast is deterministic.

Forecasting Variance: Two Steps Ahead

For two-step-ahead forecasting, we have

$$\begin{aligned}\hat{\sigma}^2_{n+2} &= c + a_1\mathbb{E}(u^2_{n+1}|\mathcal{F}_n) + a_2 u^2_n + \cdots + a_p u^2_{n-p+2} + b_1\sigma^2_{n+1} + b_2\sigma^2_n + \\ &= c + a_2 u^2_n + \cdots + a_p u^2_{n-p+2} + (a_1 + b_1)\sigma^2_{n+1} + b_2\sigma^2_n + \cdots + b_q\end{aligned}$$

For GARCH(1, 1) model, the n-step-ahead forecast can be written as

$$\hat{\sigma}^2_{T+n} = \frac{c(1-(a+b)^{n-1})}{1-a-b} + (a+b)^{n-1}\sigma^2_{T+1} \to \frac{c}{1-a-b}.$$

Estimating the Volatility

Take GARCH(1,1) as example,

1. Set $\sigma_0^2 = \frac{\hat{c}}{1-\hat{a}-\hat{b}}$.
2. $\hat{\sigma}_1^2 = \hat{c} + \hat{a}u_0^2 + \hat{b}\sigma_0^2$
3. $\hat{\sigma}_2^2 = \hat{c} + \hat{a}u_1^2 + \hat{b}\hat{\sigma}_1^2$

$\vdots$

Diagnostics of GARCH Models

1. We should check whether $\varepsilon_t = \frac{u_t}{\sigma_t}$ is iid.
2. We should first estimate ε_t,

$$\hat{\varepsilon}_t = \frac{u_t}{\hat{\sigma}_t}.$$

3. Test for iid
   - Non-correlation: correlogram, Ljung-Box test, etc.
   - If normality is assumed, test for normality
   - Test on independence.

Time Series Regression Assumptions

(1) Linearity

$$y_t = \beta_0 + \beta_1 x_{1t} + \cdots \beta_k x_{kt} + u_t.$$

(2) (x_t, y_t) are jointly stationary and ergodic.

(3) No perfect collinearity.

(4) Past and contemporary exogeneity $\Leftrightarrow$

$$\mathbb{E}(u_t | x_t, x_{t-1}, ...) = 0.$$

Ergodicity

- An ergodic time series (x_t) has the property that x_t and x_{t-k} are independent if k is large.
- If (x_t) is stationary and ergodic, then a law of large number holds,

$$\frac{1}{n}\sum_{t=1}^{n} x_t \to \mathbb{E}x \quad a.s. \; .$$

Exogeneity in Time Series Context

- Strict exogeneity.

$$\mathbb{E}(u_t|X) = \mathbb{E}(u_t|..., x_{t+1}, x_t, x_{t-1}, ...) = 0.$$

- Past and Contemporary exogeneity.

$$\mathbb{E}(u_t|x_t, x_{t-1}, ...) = 0.$$

The Puzzling Exogeneity

- US GDP and East Asian Export
- Oil prices and inflation
- Monetary policy and inflation

Consistency of OLS

Under TSR Assumptions (1)-(4), the OLS estimator of the time series regression is consistent.

Special Cases

- Autoregressive models (AR),

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + u_t.$$

- Autoregressive distributed lag models (ARDL)

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \gamma_1 x_{t-1} + \cdots + \gamma_q x_{t-q} + u_t.$$

- Autoregressive models with exogenous variable (ARX)

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \gamma_1 x_t + \cdots + \gamma_q x_{t-q+1} + u_t,$$

where (x_t) is past and contemporary exogenous.

Beat OLS in Efficiency

- OLS is consistent, but is not efficient in general.
- u_t may be serially correlated and/or heteroscedastic. In such cases, GLS would be a better alternative.
- A simple way to account for serial correlation is to explicitly model u_t as an ARMA process:

$$y_t = x'\beta + u_t,$$

  where $u_t \sim ARMA(p, q)$. But OLS is no longer able to estimate this model. Instead, nonlinear least square or MLE should be used.

Granger Causality

- Granger causality means that if x causes y, the x is a useful predictor of y_t.
- Granger Causality Test. In the model

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \gamma_1 x_{t-1} + \cdots + \gamma_q x_{t-q} + u_t.$$

  We test:

$$\mathrm{H}_0 : \gamma_1 = \cdots = \gamma_q = 0.$$

- The above test should be more appropriately called a non-causality test. Or even more precisely, a non-predicting test.
- Example: Monetary cause of inflation.

$$\pi_t = \beta_0 + \beta_1 \pi_{t-1} + \cdots + \beta_p \pi_{t-p} + \gamma_1 M1_{t-1} + \cdots + \gamma_q M1_{t-q} + u_t.$$

Two Applications

- Inflation forecast and causality test.
- Estimating CAPM model.