INTRODUCTION TO ARMA MODELS

JEN-WEN LIN, PhD, CFA

Date: 09 January, 2018

1

GRADUATE STUDENT PROJECTS

- **Past project topics:**
  - PM2.5 Time Series Analysis
  - Dynamic Asset Allocation for Pairs Trading
  - Macro economic factor (smart beta)
- **Evaluation**
  - Presentation
  - Research report (see next page for requirement)

1. Submit your project electronically. In the email, you have to include
   - Subject of the email: **STA2202 project**: **names** (**student IDs**)
   - Attachment in the email:
     1. write-up of your project (not including your codes)
     2. Computer codes (e.g. R, python, excel...etc) that can reproduce your analysis in your report
     3. Data used in your analysis
2. The content in the write-up (no more than 15 pages, including references and appendix):
   - Executive summary
   - Introduction to the problem of interest
   - Introduction the time series model that you use to solve the problem of interest
   - Explain why the chosen model is the right tool to solve the problem
   - Implement and describe your analysis
   - Carry out relevant statistical tests that show your implementation is good
   - Conclusion (good or bad)
   - References
   - Appendix (if any)

APPROACHES TO TIME SERIES ANALYSIS

Time domain analysis

- Focuses on modeling some future values of a time series as a parametric function of the current and past values

Frequency domain analysis

- Assumes that the primary characteristics of interest in time series analyses relate to periodic or systematic sinusoidal variations found naturally in most data

In many cases, the two approaches may produce similar answers for long series, but the comparative performance over short samples is better done in the time domain.

WHAT DID WE LEARN LAST WEEK

- Review what we learned in Stats 201
- Time series data exhibit trends, cyclical/seasonal patterns, as well as temporal dependence.
- Weak stationarity
- Sample autocorrelation functions (SACF) and their asymptotic distributions

LEFTOVER FROM LAST LECTURE

Be cautious about using conventional statistics on time series data

SAMPLE VARIANCE FOR DEPENDENT TIME SERIES

- For simplicity and for the demonstration purposes only, let's assume z_t follows an autoregressive process of order one

$$z_t = \phi z_{t-1} + \xi_t, \quad \xi_t \sim NID(0, \sigma_a^2)$$

- For an AR(1) process, the autocorrelation at lag k is simply ϕ^k. We have

$$\frac{E(\hat{s}^2)}{var(z_t)} = 1 - \frac{2}{n-1}\left[\sum_{k=1}^{n-1} \phi^k - \frac{1}{n}\sum_{k=1}^{n-1} k\phi^k\right].$$

That is, unlike what we learned in Stat 201, the sample variance is a biased estimator of $var(z_t)$.

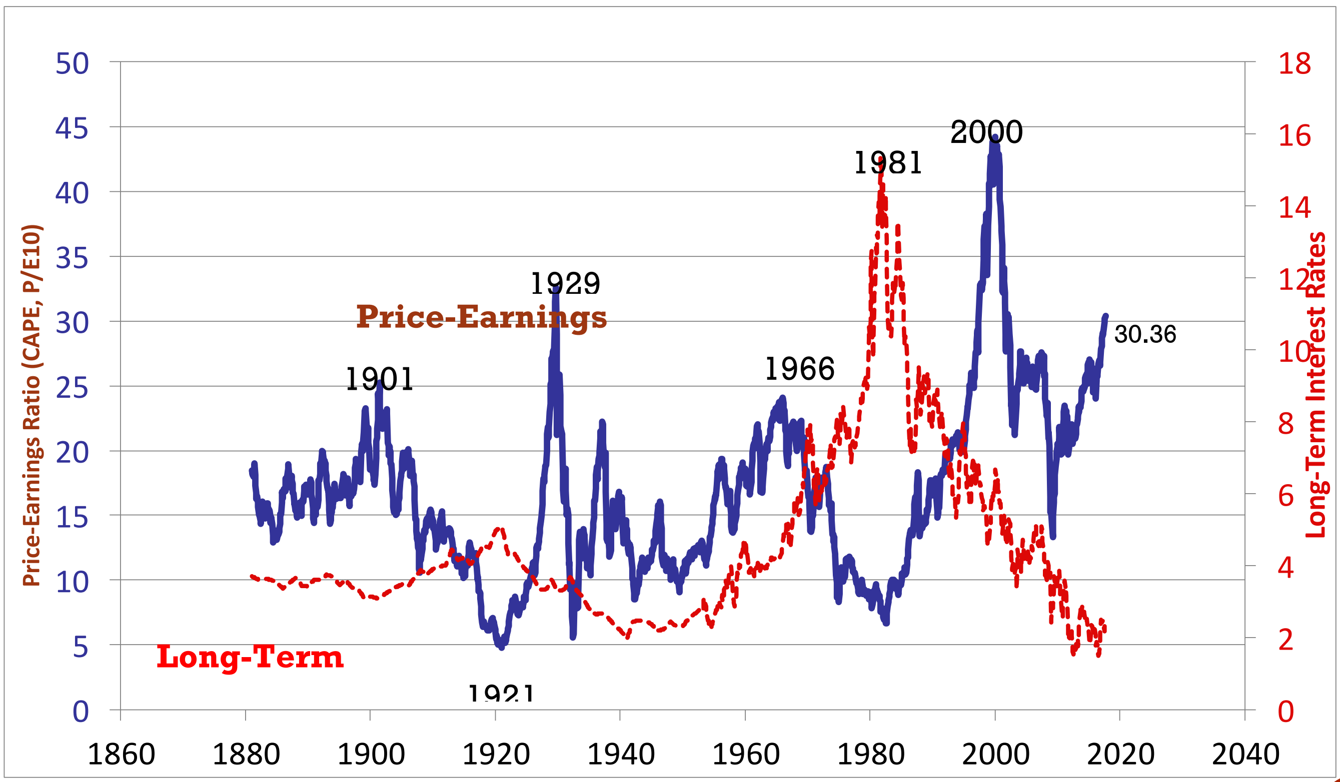
SAMPLE VARIANCE FOR DEPENDENT TIME SERIES

Table 1 summarizes the relationship between ϕ and the bias ratio, i.e. $E(\hat{s}^2)/var(z_t)$, when the series length is 240.

Table 1: The relationship of autocorrelation and bias ratio (n=240)

	variance	volatility
phi= 0	1.0000	1.0000
phi= 0.1	0.9991	0.9995
phi= 0.2	0.9979	0.9990
phi= 0.3	0.9964	0.9982
phi= 0.4	0.9945	0.9972
phi= 0.5	0.9917	0.9958
phi= 0.6	0.9876	0.9938
phi= 0.7	0.9807	0.9903
phi= 0.8	0.9672	0.9835
phi= 0.9	0.9278	0.9632
phi= 0.99	0.4858	0.6970

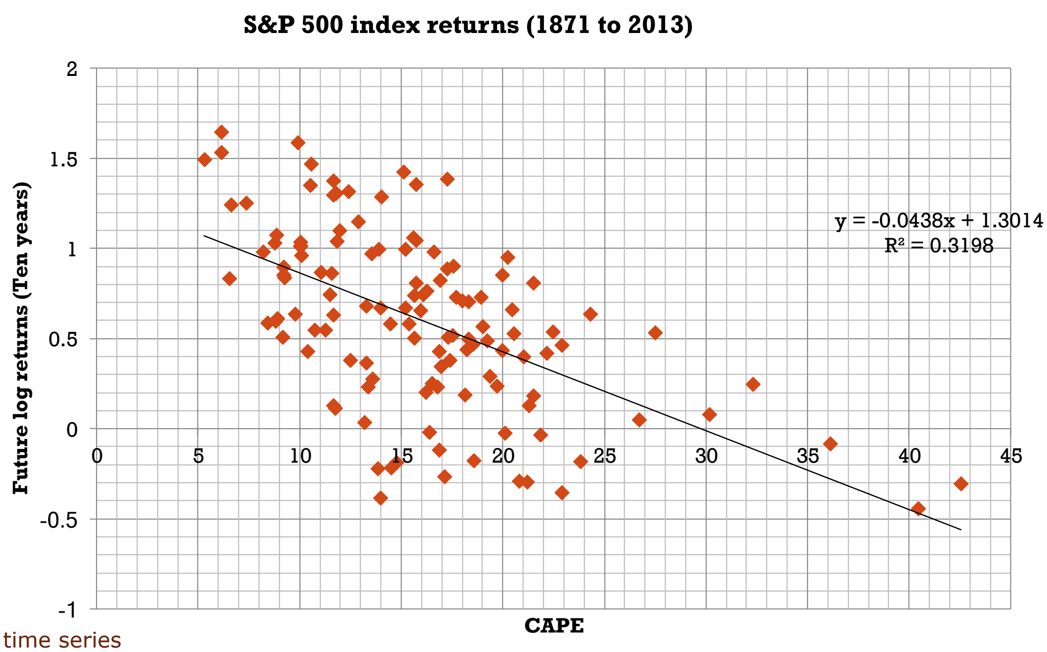
U.S. STOCK MARKETS 1871-PRESENT AND CAPE RATIO



Source: http://www.econ.yale.edu/~shiller/data.htm.

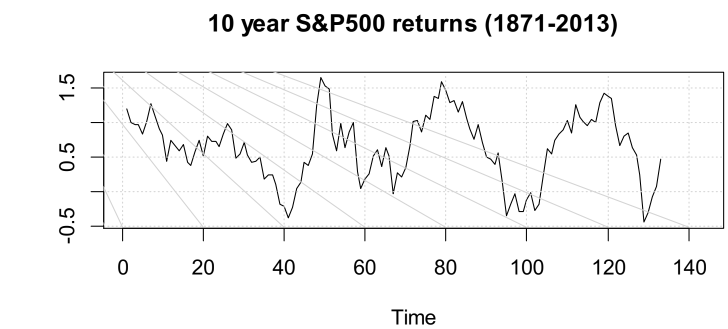
STOCK MARKET PREDICTABILITY (CAMPBELL AND SHILLER, 1998)

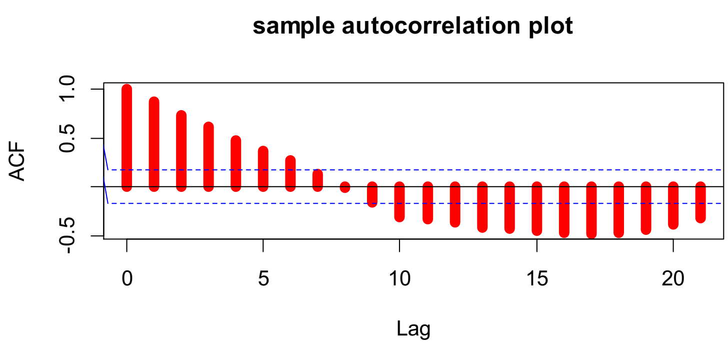
- Campbell and Shiller (1998) tested the hypothesis of stock market predictability by regressing forward ten year returns against Cyclically Adjusted PE ratio (CAPE).



Source: http://www.econ.yale.edu/~shiller/data.htm.

WHAT'S WRONG WITH THE CONVENTIONAL ANALYSIS





- The fitted model and the inference of the above hypothesis may be questionable for time series data.

- We are going to learn how to conduct the right analysis for time series data in this course

CORRELATION IN SIMPLE REGRESSION

$$y_t = \alpha + \beta x_t + e_t$$

$$cov(y_t, x_t) = cov(\alpha + \beta x_t + e_t, x_t) = \beta cov(x_t, x_t) = \beta \sigma_x^2$$

$$cov(y_t, x_t) = \rho_{xy}(0) \cdot \sigma_x \sigma_y$$

$$\beta = \frac{cov(y_t, x_t)}{\sigma_x^2} = \rho_{xy}(0) \cdot \frac{\sigma_y}{\sigma_x} \propto \rho_{xy}(0)$$

- Since standard deviation is nonnegative, we may test whether beta is zero via testing whether the cross correlation function (CCF) at lag zero is statistically different from zero.
- How to test whether CCF is statistical different from zero?

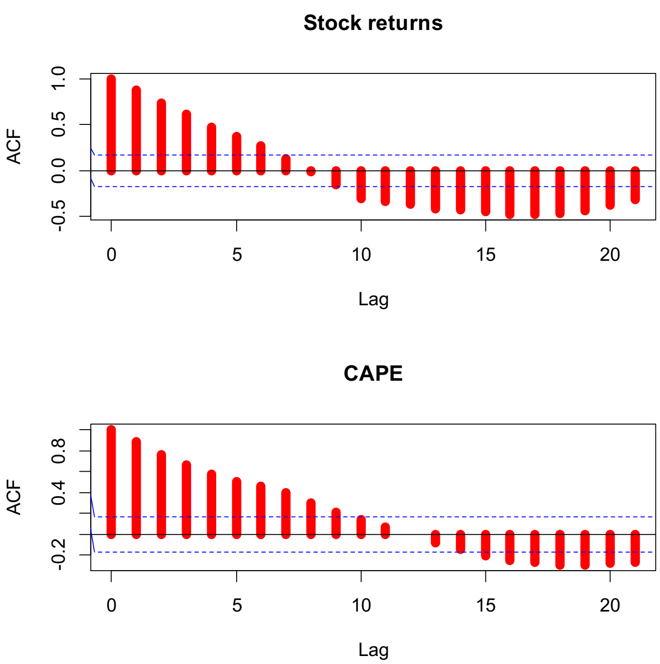
TEST ZERO CROSS-CORRELATION

- If both x_t and y_t contain no serial correlation and mutually independent, the sample CCF, $r_{xy}(k)$, is approximately normally distributed with zero mean and variance $1/n$, where n is the sample size (or the number of pairs of available).
  - The above statement is incorrect if x_t and y_t are serially correlated (even if the processes x_t and y_t are independent of each other).
- Under the assumption that both x_t and y_t are stationary and that they are independent of each other, $\sqrt{n}\, r_{xy}(k)$ is asymptotically normal with mean zero and variance

$$1 + 2\boxed{\sum_{j=1}^{\infty} \rho_X(j)\rho_Y(j),}$$

where $\rho(k)$ denotes the autocorrelation at lag k. For refinement of this asymptotic result, see Brockwell and Davis (1991, p.410).

REVISIT CAMPBELL AND SHILLER (1998, JPM)



- Both stock returns and CAPE exhibit severe serial correlation (persistent).
- Ignore the presence of serial correlation. The estimate of CCF at lag zero is -0.57 and the p-value for zero CCF is 3.6e-10.
- Take into account the serial correlation using the prewhitening technique (to discuss latter in the course). The prewhitened CCF at lag zero is -0.192 and the corresponding p-value for zero CCF is 0.033 (unable to reject H_0 at 99% significance level).
- In practice, we usually test the significance of regression coefficients based on HAC estimator, such as Newey and West (1987).

SPURIOUS CORRELATION IN TIME SERIES

Yule (1926), "Why do we Sometimes get Nonsense-Correlation between Time-Series?—A Study in Sampling and the Nature of Time-Series", *Journal of the Royal Statistical Society*, vol. 89, No. 1

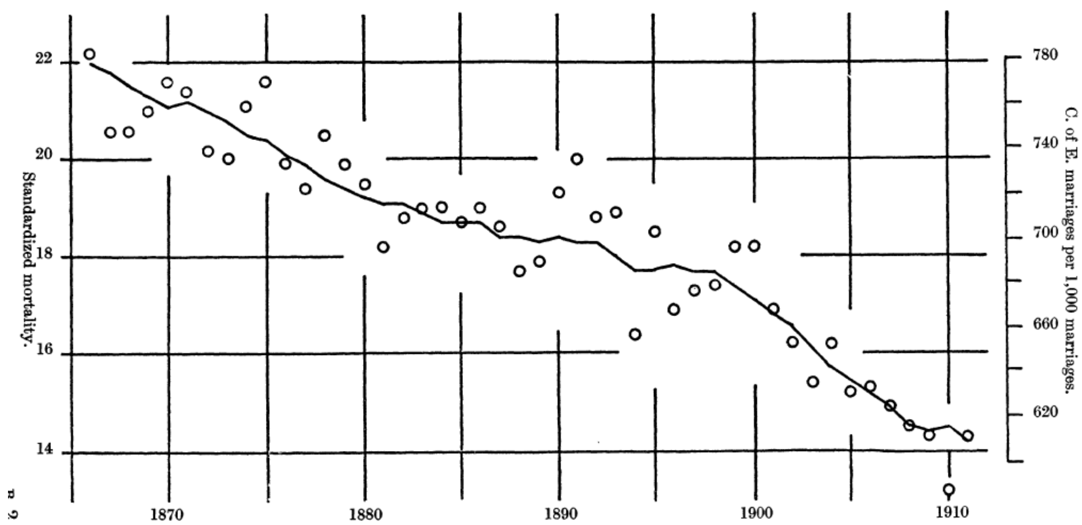


FIG. 1.—Correlation between standardized mortality per 1,000 persons in England and Wales (circles), and the proportion of Church of England marriages per 1,000 of all marriages (line), 1866–1911. $r = + 0{\cdot}9512$.

SPURIOUS CORRELATION IN TIME SERIES

- Characteristics of time series lead to various forms of spurious correlation/regression.
  - Theoretical justification, correct model specification, and tests are useful to avoid making decision based on spurious correlation.
- In this class, we will introduce some techniques to deal with spurious correlation in time series.
- **Remark:** Correlation does not imply causation.

CLASSICAL DECOMPOSITION OF TIME SERIES

DECOMPOSITION OF TIME SERIES

Statisticians usually decompose a time series into components representing

- trend
- seasonal variation
- other cyclic changes
- irregular fluctuations.

This decomposition is sometimes referred to as the classical decomposition model in time series analysis.

Seasonal variation

- Time series exhibit variation that is annual in period (or every 12 units of time).
- For example, the sales of electronic companies in the second quarter are typically the lowest.

Cyclic variation

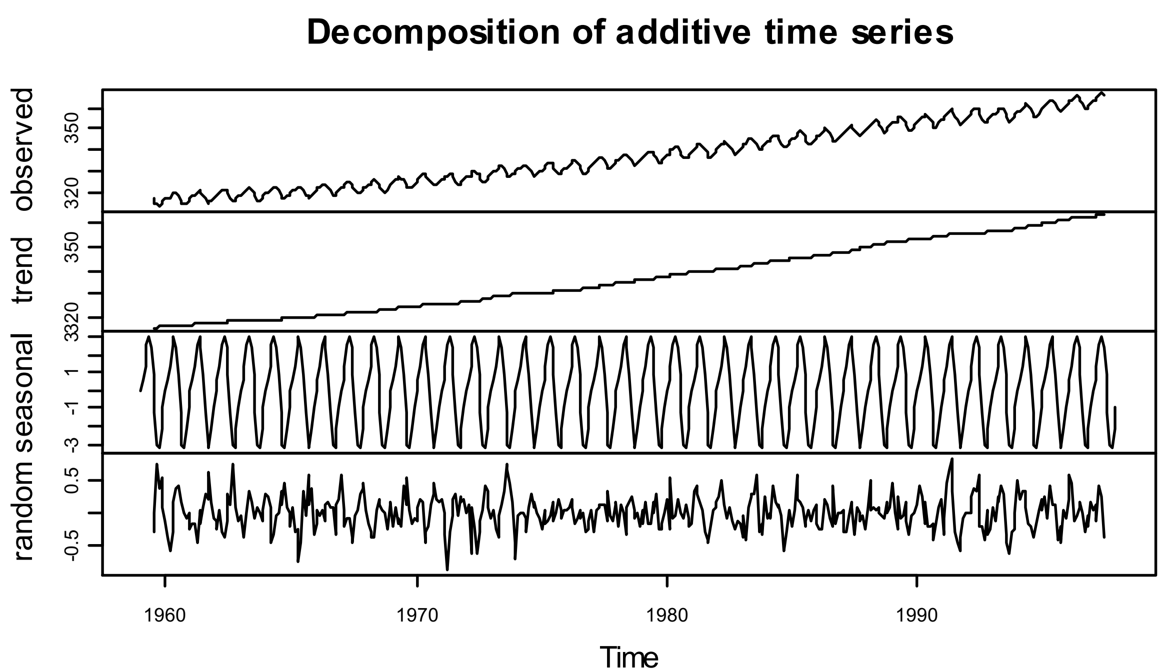
- Time series exhibit variation at a fixed period due to some other physical cause.
- Examples are daily variation in temperature and business cycles.

Trend

- This may be loosely defined as 'long-term change in the mean level'.

CLASSICAL DECOMPOSITION IN R

Atmospheric concentrations of $CO2$ are expressed in parts per million (ppm) and reported in the preliminary 1997 SIO manometric mole fraction scale.

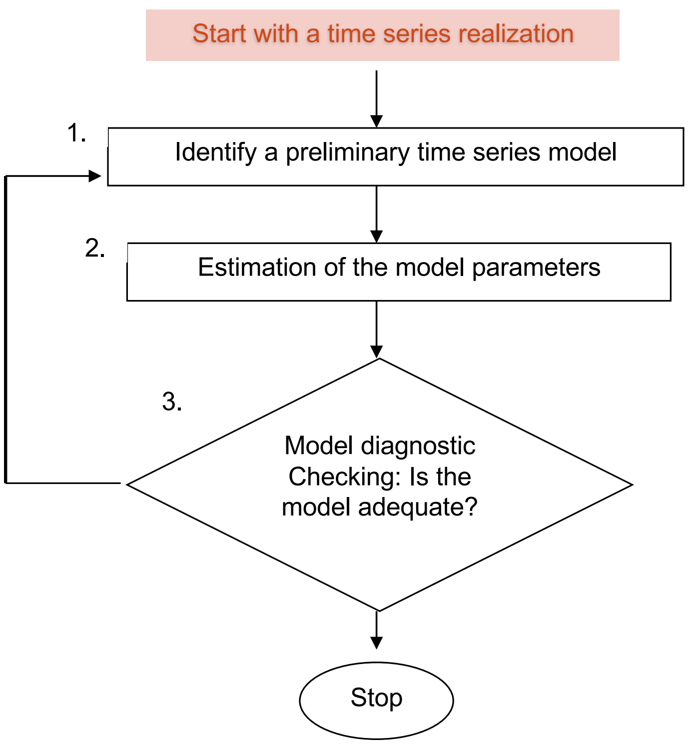


STEPS TO TIME SERIES MODELING

1. Plot the time series and check for
   - Trend, seasonal and other cyclic components, any apparent sharp changes in behavior, as well as any outlying observations
2. Remove trend and seasonal components to get residuals
3. Choose a model to fit the residuals
4. Forecasting can be carried out by forecasting residual and then inverting the transformation carried out in Step 2.

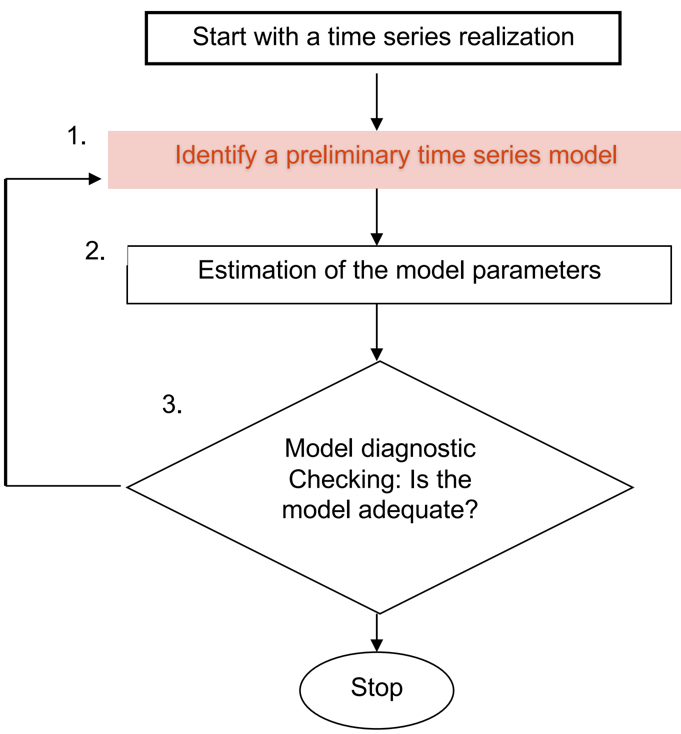
THREE STAGE OF BOX-JENKINS APPROACH

THREE STAGES OF BOX-JENKINS APPROACH



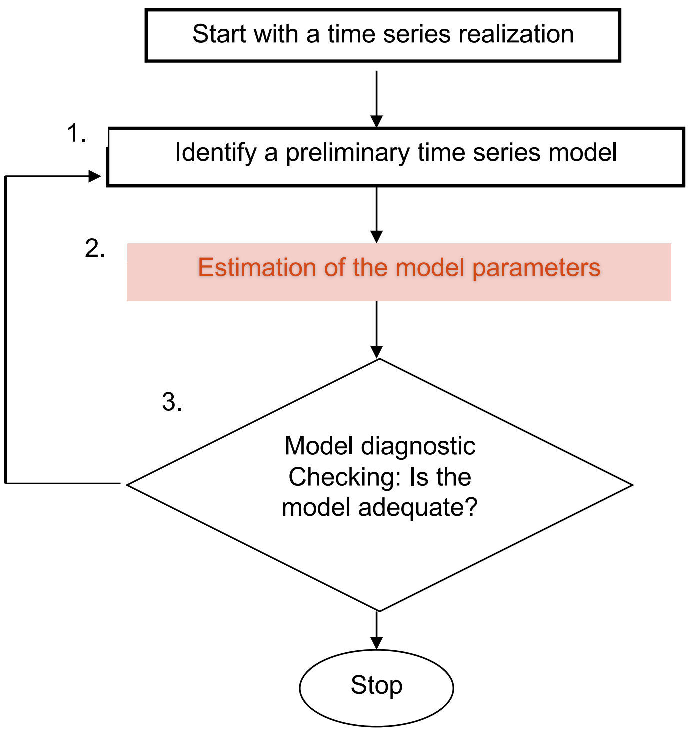
- Understand the problem of interest
- Collect data
- Plot time series data

FIRST STAGE—MODEL IDENTIFICATION



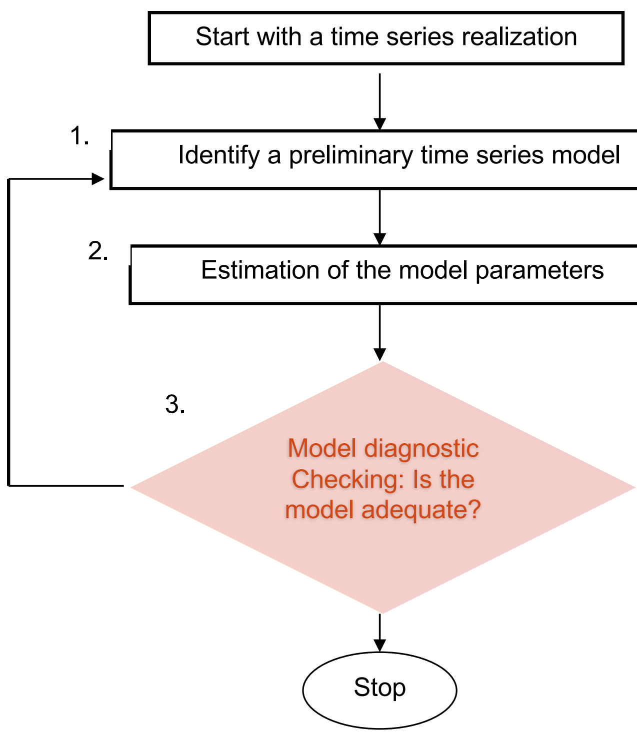
- Perform differencing and transformations to transform the data into stationarity.
- Identify preliminary ARMA(p,q) models using sample autocorrelations and sample partial autocorrelations

SECOND STAGE—MODEL ESTIMATION



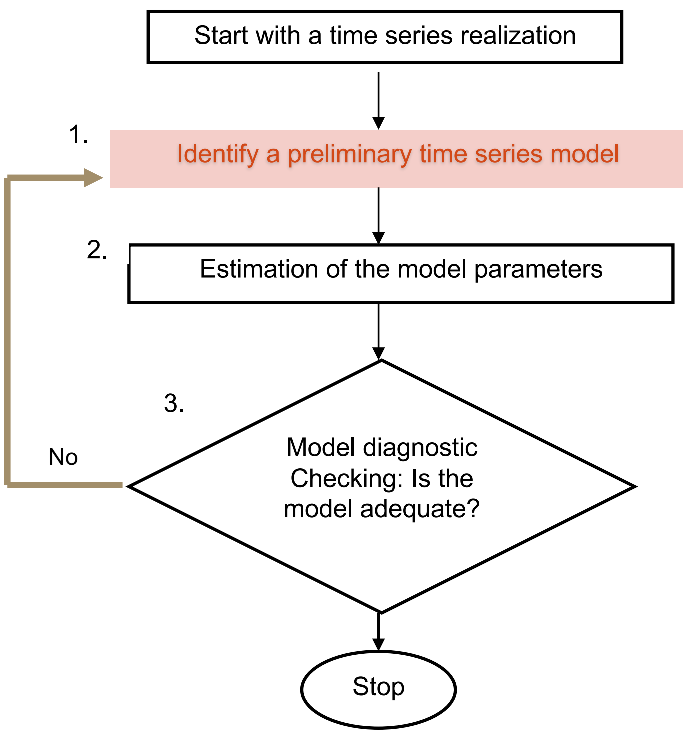
- Method of moments
- Maximum likelihood Estimation
- Kalman Filter
- Others
- To discuss latter in this course

THIRD STAGE—MODEL EVALUATION



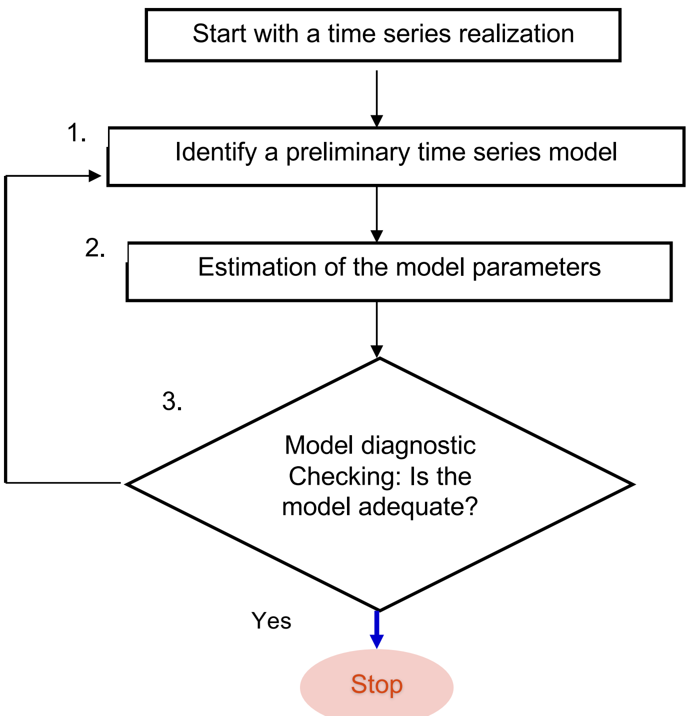
- Model adequacy is checked by examining if the residuals of the fitted model are approximately uncorrelated (after taking into account the effect of estimation)

MODEL IS NOT ADEQUATE



- The fitted model fails diagnostic checks.
- Return to the first stage and identify another time series model.

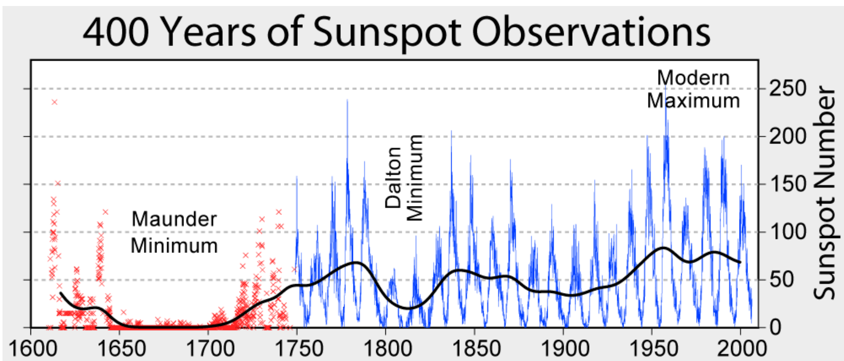
MODEL IS ADEQUATE



- If the fitted model passes diagnostic checks, we may the use the model for our analysis.

TIME SERIES MODEL FOR IRREGULAR COMPONENT

MODELING IRREGULAR COMPONENT



- The origin of ARMA models can be traced back to Yule's work on Wolfer's sunspot numbers.
- ARMA models were widely accepted by researchers and professionals after the work of Box and Jenkins (1970).

AUTOREGRESSIVE MOVING AVERAGE (ARMA) MODEL

- A process $\{X_t\}$ is said to be an ARMA(p, q) process if $\{X_t\}$ is stationary and if for every *t*

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q},$$

where $a_t \sim WN(0, \sigma^2)$.

  - $\{X_t\}$ is said to be an ARMA(*p*,*q*) process with mean μ if $\{X_t - \mu\}$ is an ARMA(*p*,*q*) process.

- Or, use compact notation $\Phi(B)(X_t - \mu) = \Theta(B)a_t$, where

$$\Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$$

and

$$\Theta(z) = 1 + \theta_1 z + \cdots + \theta_{\cancel{p}\,q} z^{\cancel{p}\,q}.$$

MA(∞) PROCESSES

- If $\{a_t\} \sim WN(0, \sigma^2)$ then we say that $\{X_t\}$ is a MA(∞) process of $\{a_t\}$ if there exists a sequence $\{\psi_j\}$ with $\sum_{j=0}^{\infty} |\psi_j| < \infty$ such that

$$X_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}, t = \cdots, -1,0,1,2, \dots$$

- We can calculate autocorrelation functions of a stochastic process $\{X_t\}$ using the aforementioned theorem as long as $\{X_t\}$ can be written in the form of a MA(∞) process.
- **Theorem** (Brockwell and Davis, 1992, p.91) The MA(∞) process is stationary with mean zero and autocovariance function

$$\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|}$$

MOVING AVERAGE PROCESS OF ORDER q

- The moving average process of order q, denoted as $MA(q)$, is given by

$$X_t = a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q} = \Theta(B) a_t$$

where B is the backward shift operator, $B^h X_t = X_{t-h}$,

$$\Theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$$

and $a_t \sim WN(0, \sigma^2)$

- **Question:** What conditions do we need for $MA\ (p)$ processes to be weakly stationary?

AUTOCOVARIANCE OF MA(2) PROCESS

$$\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|}$$

$$X_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2}, a_t \sim WN(0, \sigma^2)$$

$$X_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}, t = \cdots, -1,0,1,2, \ldots$$

$$\psi_j = \theta_j, \, j = 0,1,2 \text{ and } \psi_j = 0, \, j > 2$$

$$\gamma(1) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+1} = \sigma^2 (1 \cdot \theta_1 + \theta_1 \cdot \theta_2 + \theta_2 \cdot 0) = \sigma^2 (\theta_1 + \theta_1 \theta_2)$$

$$\gamma(2) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+2} = \sigma^2 (1 \cdot \theta_2 + \theta_1 \cdot 0 + \theta_2 \cdot 0) = \sigma^2 \theta_2$$

$$\gamma(3) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+3} = \sigma^2 (1 \cdot 0 + \theta_1 \cdot 0 + \theta_2 \cdot 0) = 0$$

AUTOCOVARIANCE OF MA(q) PROCESSES

$$\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|}$$

$$X_t = a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}, a_t \sim WN(0, \sigma^2)$$

$$\psi_j = \theta_j,\ j = 0, 1, \dots, q \text{ and } \psi_j = 0\ \forall j > q$$

$$\gamma(k) = cov(X_t, X_{t+k})$$

$$= cov\left(\sum_{j=0}^{q} \theta_j a_{t-j}, \sum_{j=0}^{q} \theta_j a_{t+k-j}\right)$$

$$= \begin{cases} \sigma^2 \sum_{i=0}^{q-k} \theta_i \theta_{i+k}, & k = 0, \pm 1, \dots, \pm q \\ 0, & otherwise \end{cases}$$

SUMMARY

- The maximum lag of the non-zero sample autocorrelation is a good indicator of the MA(q) processes.

- The ACF of MA(q) processes

$$X_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q},$$

cut off after lag q, i.e.

$$\rho_k = \begin{cases} \dfrac{\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2}, & k = 1, \ldots, q \\ 0 & , k > q \end{cases}$$

AUTOREGRESSIVE MODEL OF ORDER p

- The autoregressive process (model) of order p, denoted as AR(p), is given by

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \Phi(B) X_t = a_t,$$

where $a_t \sim WN(0, \sigma^2)$, $B^h X_t = X_{t-h}$, $h \in Z$, and

$$\Phi(B) = \left(1 - \phi_1 B - \cdots - \phi_p B^p\right)$$

Questions to think about

1. How to check the weak stationarity of $\{X_t\}$?
2. Derive the autocovariance function of $\{X_t\}$?
3. Use ACF for model identification?

AR(1) PROCESSES

- The autoregressive process of order one is given by

$$X_t - \phi X_{t-1} = a_t, \{a_t\} \sim WN(0, \sigma^2),$$

where a_t is uncorrelated with X_s for all $s < t$.

AR(1) PROCESSES

An AR(1) process with $|\phi| <$ 1 is called causal or a future-independent autoregressive process. In this course, we consider only causal processes.

$$
\begin{aligned}
\gamma(k) &= \operatorname{cov}(X_t, X_{t+k}) \\
&= \operatorname{cov}\Big(\sum_{l=0}^{\infty} \phi^l a_{t-l}, \sum_{j=0}^{\infty} \phi^j a_{t+k-j}\Big) \\
&= \operatorname{cov}\Big(\boxed{\sum_{l=0}^{\infty} \phi^l a_{t-l}}, \sum_{s=0}^{k-1} \phi^s a_{t+k-s} + \boxed{\sum_{l=0}^{\infty} \phi^{l+k} a_{t-l}}\Big) \\
&= \phi^k \gamma(0)
\end{aligned}
$$

$$
\gamma(0) = Var(X_t) = \sigma^2 / (1 - \phi^2)
$$

finite if $|\phi| < 1$, why?

REMARK AND QUESTIONS

- We have shown answers for an AR(1) process using definition and its MA(∞) representation.
- Can we apply the same technique to higher order AR process and answer the questions asked earlier?
  1. Check weak stationarity of an AR(p) model
  2. Derive autocorrelation functions of an AR(p) model
  3. Can we use the pattern of ACFs of an AR(p) process for model identification

General approach to checking stationarity of an AR(p) process

- A general way of checking the (weak) stationarity condition of autoregressive processes is that the roots of the following equation

$$\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p = 0$$

must lie outside the unit circle.

- *Example:* Check the stationarity of an AR(1) model

$$(1 - \phi B)X_t = a_t$$

  - $\Phi(B) = 1 - \phi B = 0$ so $B = 1/\phi$
  - $|B| = |1/\phi| > 1$ so $|\phi| < 1$

DERIVE AUTOCOVARIANCE FUNCTIONS OF AUTOREGRESSIVE PROCESSES

For simplicity, consider a stationary AR(1) process

$$X_t = \phi X_{t-1} + a_t,\ \ a_t \sim NID(0, \sigma^2), \qquad (\star)$$

For $k = 1$: multiply X_{t-1} on both sides of ($\star$) and take expectation on both sides of the equation

$$X_t X_{t-1} = \phi X_{t-1}^2 + X_{t-1} a_t$$

Take Expecation:

$$E(X_t X_{t-1}) = \phi \cdot Var(X_t)$$

$$\Rightarrow \gamma(1) = \phi \cdot \gamma(0)$$

$$X_t X_{t-1} = \phi X_{t-1}^2 + X_{t-1} a_t$$

Take Expecation:

$$\boxed{E(X_t X_{t-1})} = \boxed{\phi \cdot Var(X_t)}$$

$$\Rightarrow \gamma(1) = \phi \cdot \gamma(0)$$

$$X_{t-1} = \sum_{j=0}^{\infty} \phi^j \cdot a_{t-1-j}$$

$$\mathrm{cov}(a_t, X_{t-1}) = \mathrm{cov}(a_t, \sum_{j=0} \phi^j a_{t-1-j}) = 0$$

For $k = 2$: multiply X_{t-2} on both sides of (*) and take expectation on both sides of the equation

$$X_t X_{t-2} = \phi X_{t-1} X_{t-2} + X_{t-2} a_t$$

Take Expecation:

$$E(X_t X_{t-2}) = \phi \cdot E(X_{t-1} X_{t-2})$$

$$\Rightarrow \gamma(2) = \phi \cdot \gamma(1)$$

Using the result that $\gamma(1) = \phi \cdot \gamma(0)$

$$\Rightarrow \gamma(2) = \phi \cdot \gamma(1) = \phi^2 \cdot \gamma(0)$$

For k $\geq$3, similarly we have

$$X_t X_{t-k} = \phi X_{t-1} X_{t-k} + X_{t-k} a_t$$

Take Expecation:

$$E(X_t X_{t-k}) = \phi \cdot E(X_{t-1} X_{t-k})$$

$$\Rightarrow \gamma(k) = \phi \cdot \gamma(k-1)$$

$$\Rightarrow \qquad \cdots\cdots$$

$$\Rightarrow \boxed{\gamma(k) = \phi^k \gamma(0)}$$

Question: Can we apply the above technique to a general AR(p) process?

ACF OF AR(p) PROCESSES

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots\ldots\phi_p X_{t-p} + a_t$$

Stationarity condition of an AR(p) models are that all *p* roots of the characteristic equation outside of the unit circle

ACF

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots\ldots\phi_p\rho_{k-p}$$

$$\left.\begin{array}{l} \rho_1 = \phi_1\rho_0 + \phi_2\rho_1 + \ldots\ldots\phi_p\rho_{p-1} \\ \rho_2 = \phi_1\rho_{11} + \phi_2\rho_0 + \ldots\ldots\phi_p\rho_{p-2} \\ \vdots \\ \rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \ldots\ldots\phi_p\rho_0 \end{array}\right\}$$

System (Yule-Walker equations) to solve for the first *p* autocorrelations: *p* unknowns and *p* equations

In general, ACFs of an AR(p) model decay as mixture of exponentials and/or damped sine waves--depending on real/complex roots

ALTERNATIVE MEASURE OF TEMPOROAL DEPENDENCE
PARTIAL AUTOCORRELATION FUNCTION

- The correlation between X_t and X_{t+k} after mutual linear dependency on the intervening variables, $X_{t+1}, X_{t+2}, \ldots,$ and X_{t+k-1} has been removed.
- The conditional correlation

$$\phi_{kk} = corr(X_t, X_{t+k} | X_{t+1}, \ldots, X_{t+k-1})$$

  is usually referred to as the partial autocorrelation functions (PACF) in time series analysis.
  - PACF between X_t and X_{t+k} can be obtained as the regression coefficient associated with X_t when regressing X_{t+k} on its k lagged variables $X_{t+k-1}, X_{t+k-2}, \ldots,$ and X_t.

CALCULATE PACF USING YULE-WALKER EQUATIONS

- A general method for finding the partial autocorrelation function for any stationary process with autocorrelation function ρ_k is as follows.
- For a given lag k, it can be shown that the ϕ_{kk} satisfy the Yule-Walker equations.

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \phi_{k3}\rho_{j-3} + \cdots + \phi_{kk}\rho_{j-k},$$

for $j = 1,2, \dots, k$. That is, we regard $\rho_1, \dots, \rho_k$ as given and wish to solve for ϕ_{kk}.

SOLVING YW EQUATIONS USING CRAMER'S RULE

- For the AR(2) process, the Yule-Walker equations may be written as

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-1}, \text{ for } k \geq 1.$$

- Given a set of ACFs, we can solve $\phi_{11}, \phi_{22}, \phi_{33}$ based on Yule-Walker equations:

$$\phi_{11} = \rho_1 = \frac{\phi_1}{1-\phi_2} \qquad \phi_{22} = \frac{\det\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{bmatrix}}{\det\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}} = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2} = \phi_2$$

CRAMER'S RULE

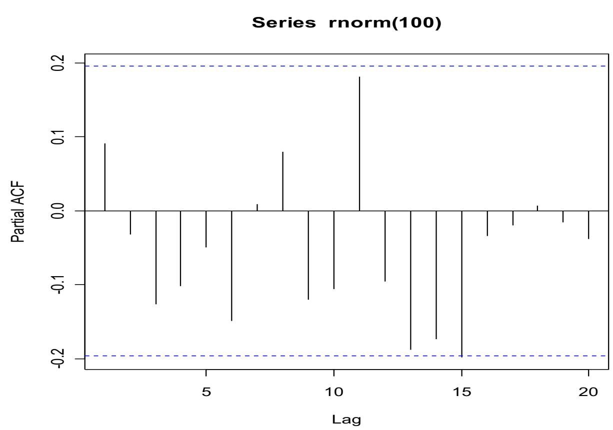
$$\phi_{33} = \frac{\det\begin{bmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{bmatrix}}{\det\begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{bmatrix}} = \frac{\det\begin{bmatrix} 1 & \rho_1 & \phi_1 + \phi_2\rho_1 \\ \rho_1 & 1 & \phi_1\rho_1 + \phi_2 \\ \rho_2 & \rho_1 & \phi_1\rho_2 + \phi_2\rho_1 \end{bmatrix}}{\det\begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{bmatrix}} = 0 \qquad \phi_{kk} = 0,\ k \geq 3$$

This example show that for an AR(p) model, PACF at lag k equals zero if k is greater than p, where k and p are integers.

DISTRIBUTION OF SAMPLE PACF

- Under the hypothesis that the underlying process is white noise sequence, sample PACF are normally distributed with $var(\widehat{\phi_{kk}}) = 1/n$ asymptotically.
- Hence, $\pm {}^{2}/_{\sqrt{n}}$ can be used as critical limits (95% confidence level) to test for the hypothesis of a white noise process.

Sample autocorrelation function of a simulated NID sequences



CAUSAL AND INVERTIBLE PROCESSES (BROCKWELL AND DAVIS, 1992)

Causal process:	$\{X_t\}$ can be expressed in terms of $\{a_s, s \leq t\}$. Such processes are called causal or future-independent autoregressive process.
Invertible process:	No restrictions on the $\{\theta_i\}$ are required for a (finite-order) MA process to be stationary. The imposition of the invertible condition ensures that there is a unique MA process for a given set of ACF.

DUALITY BETWEEN AR AND MA PROCESSES

- A finite-order stationary AR(p) process corresponds to a MA(∞) process, and a finite-order invertible MA(q) corresponds to an AR process of infinite-order .

Casual/stationary AR(p) ➔ MA(∞)

Invertible MA(q)➔ AR(∞)

- This dual relationship also exists in autocorrelation and partial autocorrelation functions.

REMARKS ON INVERTIBILITY

- Consider the following first-order MA processes:

A: $X_t = a_t + \theta a_{t-1}$

B: $X_t = a_t + \dfrac{1}{\theta} a_{t-1}$

A: $a_t = X_t - \theta X_{t-1} - \theta^2 X_{t-2} - \cdots$, if $|\theta| < 1$

B: $a_t = X_t - \dfrac{1}{\theta} X_{t-1} - \dfrac{1}{\theta^2} X_{t-2} - \cdots$, if $|\theta| > 1$

$$\gamma_A(0) = \operatorname{var}(a_t + \theta a_{t-1}) = (1+\theta^2)\sigma^2$$
$$\gamma_A(1) = \operatorname{cov}(a_t + \theta a_{t-1}, a_{t+1} + \theta a_t) = \theta\sigma^2$$
$$\rho_A(1) = \gamma_A(1)/\gamma_A(0) = \theta/(1+\theta^2)$$

$$\gamma_B(0) = \operatorname{var}(a_t + \frac{1}{\theta} a_{t-1}) = (1+\frac{1}{\theta^2})\sigma^2$$
$$\gamma_B(1) = \operatorname{cov}(a_t + \frac{1}{\theta} a_{t-1}, a_{t+1} + \frac{1}{\theta} a_t) = \frac{1}{\theta}\sigma^2$$
$$\rho_B(1) = \frac{\frac{1}{\theta}\sigma^2}{(1+\frac{1}{\theta^2})\sigma^2} = \theta/(1+\theta^2)$$

Two different models possess the same autocorrelation functions

CAUSAL ARMA PROCESSES

- An ARMA(p,q) process, defined by the equation $\Phi(B)X_t = \Theta(B)a_t$, is said to be causal if there exists a sequence of constants $\{\psi_j\}$ such that $\sum_{j=0}^{\infty}|\psi_j| < \infty$ and $X_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$, $t = 0, \pm 1, \pm 2, \ldots$
- In compact notation, we have

$$\Psi(z) = \sum_{j=0}^{\infty} \psi_j \cdot z^j = \frac{\Theta(z)}{\Phi(z)}, \quad |z| \leq 1$$

Show how to calculate $\{\psi_j\}$ in class

INVERTIBLE ARMA MODELS

- An $ARMA(p, q)$ process is said to be invertible if there exists a sequence of constants $\{\pi_j\}$ such that $\sum_{j=0}^{\infty} |\pi_j| < \infty$ and

$$a_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, t = 0, \pm 1, \pm 2, \ldots$$

$$\Pi(z) = \sum_{j=0}^{\infty} \pi_j \cdot z^j = \frac{\Phi(z)}{\Theta(z)}, \quad |z| \leq 1$$

Show how to calculate $\{\pi_j\}$ in class

WOLD DECOMPOSITION

- Any zero-mean process $\{X_t\}$ which is not *deterministic* can be expressed as a sum of $X_t = U_t + V_t$, where $\{U_t\}$ denotes an $MA(\infty)$ process and $\{V_t\}$ is a deterministic process which is uncorrelated with $\{U_t\}$.
  - If the values $X_{n+j}, j \geq 1$ of the process $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ were perfectly predictable in terms of $\mu_n = sp\{X_t, -\infty < t \leq n\}$. Such processes are called *deterministic*.
  - If X_n comes from a deterministic process, it can be predicted (or determined) by its past observations of the process, i.e., $X_t, t < n$.

MODEL IDENTIFICATION

The first stage of Box-Jenkins analysis

Review theoretical ACFs and PACFs of ARMA models

Model identification using ACF and PACF with R examples

MA(1) PROCESSES

- Autocorrelation functions

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta\sigma^2}{(1+\theta^2)\sigma^2} = \frac{\theta}{1+\theta^2}$$

$$\rho_j = 0 \quad j > 1$$

- Partial autocorrelation functions

$$\phi_{11} = \rho_1 = \frac{\theta}{1+\theta^2} = \frac{\theta(1-\theta^2)}{1-\theta^4}$$

$$\phi_{22} = -\frac{\rho_1^2}{1-\rho_1^2} = -\frac{\theta^2}{1+\theta^2+\theta^4} = -\frac{\theta^2(1-\theta^2)}{1-\theta^6}$$

$$\phi_{kk} = -\frac{\theta^k(1-\theta^2)}{1-\theta^{2(k+1)}}, \qquad k \geq 1$$

MA(Q) PROCESSES

- The ACF of MA(q) processes,

$$X_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q},$$

cut off after lag q, i.e.

$$\rho_k = \begin{cases} \dfrac{\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2}, & k = 1, \ldots, q \\ 0 & , k > q \end{cases}$$

- The maximum lag of the non-zero sample autocorrelation is a good indicator of the MA(q) processes.

AR(1) PROCESSES

- Autocovariance and Autocorrelation:

$$\gamma_j = \phi\gamma_{j-1} \quad j \geq 1$$

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \frac{\phi\gamma_{j-1}}{\gamma_0} = \phi\rho_{j-1}, \qquad j \geq 1$$

- Partial autocorrelation functions:

$$\phi_{11} = \rho_1 = \phi$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2} = \frac{\phi^2 - \phi^2}{1-\rho_1^2} = 0$$

$$\boxed{\phi_{kk} = 0 \quad k \geq 2}$$

AR(p) PROCESSES

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots\ldots\, \phi_p X_{t-p} + a_t$$

Stationarity condition is that all p roots of the characteristic equation outside of the unit circle

ACF

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \ldots\ldots \phi_p \rho_{k-p}$$

$$\left.\begin{aligned} \rho_1 &= \phi_1 \rho_0 + \phi_2 \rho_1 + \ldots\ldots \phi_p \rho_{p-1} \\ \rho_2 &= \phi_1 \rho_{11} + \phi_2 \rho_0 + \ldots\ldots \phi_p \rho_{p-2} \\ &\vdots \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \ldots\ldots \phi_p \rho_0 \end{aligned}\right\}$$

System to solve for the first p autocorrelations: p unknowns and p equations

ACFs decay as mixture of exponentials and/or damped sine waves--depending on real/complex roots

PACF

$$\phi_{kk} = 0 \text{ for } k > p$$

TABLE 6.1(WEI, 2006)

CHARACTERISTICS OF THEORETICAL ACF AND PACF FOR STATIONARY PROCESSES

Process	ACF	PACF
AR(p)	Tails off as exponential decay or damped sine wave	Cuts off after lag p
MA(q)	Cuts off after lag q	Tails off as exponential decay or damped sine wave
ARMA(p,q)	Tails off after lag (q-p)	Tails off after lag (p-q)

MODEL ADEQUACY

The Final stage of Box-Jenkins approach

This stage is also called model diagnostic checking which involves techniques like over-fitting, residual plots and, more importantly, checking if residuals are approximately uncorrelated.

WHY RESIDUALS ARE UNCORRELATED?

- The residuals of a fitted ARMA(p,q) mode is

$$\hat{a}_t = X_t - \hat{\phi}_1 X_{t-1} - \cdots - \hat{\phi}_p X_{t-p} - \hat{\theta}_1 a_{t-1} - \cdots - \hat{\theta}_q a_{t-q},$$

where $\hat{\phi}_k$, $\hat{\theta}_k$ $\forall k$ are the parameter estimates obtained from the second stage, and $\{\hat{a}_t\}$ are the residuals of the fitted model

  - Residuals can be seen as the sample estimates of $\{\hat{a}_t\}$ and therefore are approximately uncorrelated (white noise) because of the estimation process.
  - Remark: residuals are not independent in the classical regression model

AUTOCORRELATION AMONG RESIDUALS

- Residual autocorrelation functions at lag k

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} \hat{a}_t \hat{a}_{t+k}}{\sum_{t=1}^{n} \hat{a}_t^2}$$

- How many lags are enough??
- The overall tests that check an entire group of residual autocorrelation functions (assuming that the model is adequate) are called portmanteau tests.
- In spirits, portmanteau tests may be seen as a variant of the goodness of fit tests.

POPULAR PORTMANTEAU TESTS

Box and Pierce (1970)

$$Q_{BP} = n \cdot \sum_{1}^{m} \hat{\rho}_k^2 \sim \chi^2_{m-(p+q)}$$

Ljung and Box (1978)

$$Q_{LB} = \sum_{1}^{m} \frac{n \cdot (n+2)}{(n-k)} \hat{\rho}_k^2 \sim \chi^2_{m-(p+q)}$$

EXAMPLE: LI (2004), P11

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.4	.15	.07	.06	.09	.03	.05	.06	.5	.01

- $X_t = (1-0.4B)a_t$ was fitted to a series of 80 observations.

$$Q_{BP} = 80(.4^2 + .15^2 + \ldots + .01^2) = 16.696$$

$$Q_{LB} = 80(82)(.4^2/79 + .15^2/78 + \ldots + .01^2/70) = 17.488$$

- The upper *5%* critical value from the chi-squared distribution with *9*(=*10*-*1*) degrees of freedom is 16.92.

MORE ABOUT PORTMANTEAU TESTS

- Pros:
  - Practical purposes
  - Minimal requirement for using the fitted model
- Cons:
  - Lack power if comparing with traditional statistical tests, such as likelihood ratio tests
- Possible improvements and other applications:
  - 1. Finite sample adjustments
  - 2. Complicated functional of residual autocorrelations
  - 3. Monte Carlo test: See Lin & McLeod(2006)
  - 4. Other applications: portmanteau tests for randomness and ARMA models with infinite variance innovations

MODEL SELECTION

In time series analysis, several models may adequately represent a given data set.

How to select the best model among these candidates is called model selection or order selection.

METHODS FOR MODEL SELECTION

- Two of the most popular methods are Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC):

$$AIC=-2 \log ML+2 k, \qquad (a)$$

$$BIC=-2 \log ML+k \log(n), \qquad (b)$$

where ML denotes maximum likelihood, log ML is the value of maximized log-likelihood function for a model fitted to a give data set, and k is the number of independently adjusted parameters within the model.

- *Remark*: BIC puts more penalties on the number of parameters used by fitted models, and some empirical studies indicate that the model selected by BIC performs better in the post-sample analysis, such as forecasting.