DATA ANALYTICS

Unit 3: Box Jenkins and ARIMAX

Jyothi R.
Department of Computer Science and
Engineering

Contents

- Overview

- Identifying d

- Seasonality

- Identifying p and q

- Estimation and Information Criteria

- Diagnostic Checking: Residual Diagnostics

INTRODUCTION

Present the practical and pragmatic approach of Box and Jenkins in order to build ARIMA models

Step 1 : Identification

Step 2 : Estimation (and selection)

Step 3 : Diagnostic checking

Step 4 : Model's use

INTRODUCTION

- **Step 1** (identification) involves determining the order of the model required (p, d, and q) in order to capture the salient dynamic features of the data. This mainly leads to use graphical procedures (plotting the series, ACF and PACF, etc).

- **Step 2** (estimation and selection) involves estimation of the parameters of the different models (using step 1) and proceeds to a first selection of models (using information criteria).

- **Step 3** (checking) involves determining whether the model(s) specified and estimated is adequate. Notably, one uses residual diagnostics.

Box-Jenkins Methodology
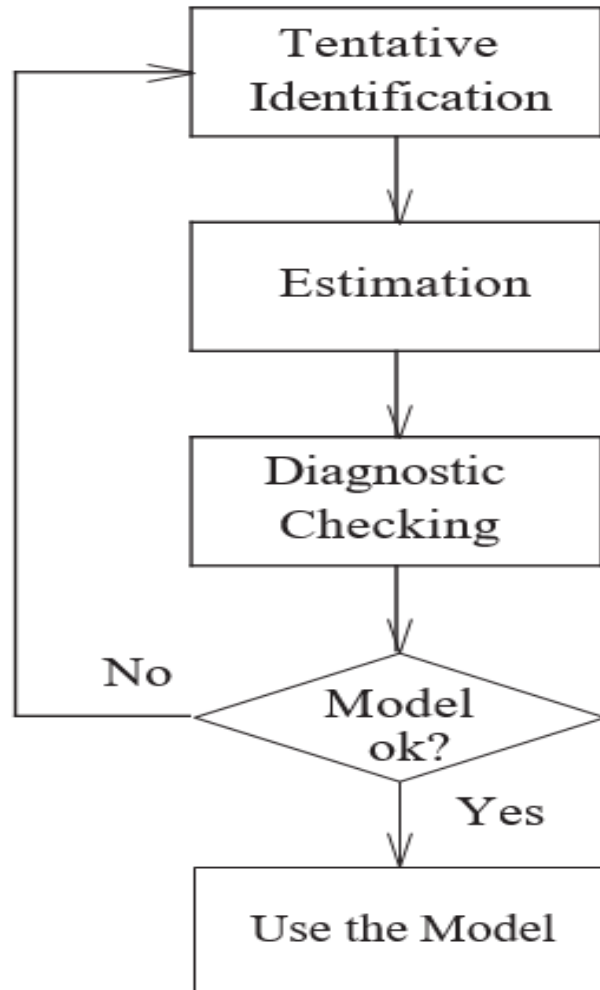
Time Series Plot Range-Mean Plot ACF and PACF

Least Squares or Maximum Likelihood

Residual Analysis and Forecasts

Forecasting Explanation Control

Figure 1:Box-Jenkins Methodology

Identification Overview

- Three objectives

- Stationarity, non-stationarity ? What is the order of differentiation (d) ?

- Seasonal component (ARIMA *versus* SARIMA model) ?

- Identification of the ARMA order(p,q).

Identifying d

Two different approaches

1. Graphical procedure : this involves plotting the data over time and the corresponding (partial) autocorrelation function.

   • If the ACF does not decrease to zero or at a very slow decay : this suggests non-stationarity (or long-memory effects).

   • Box and Jenkins (1976) recommend using the following differencing approach :
   I.   Plot the autocorrelation function of the first-difference series
   II. Iterate the previous step until the ACF looks like the one of a stationary series
   III. Check the inverse autocorrelation function to avoid over-differencing.

   2. Test Procedure : Unit root tests.

Identifying Seasonality

Figure 2: Determining d…
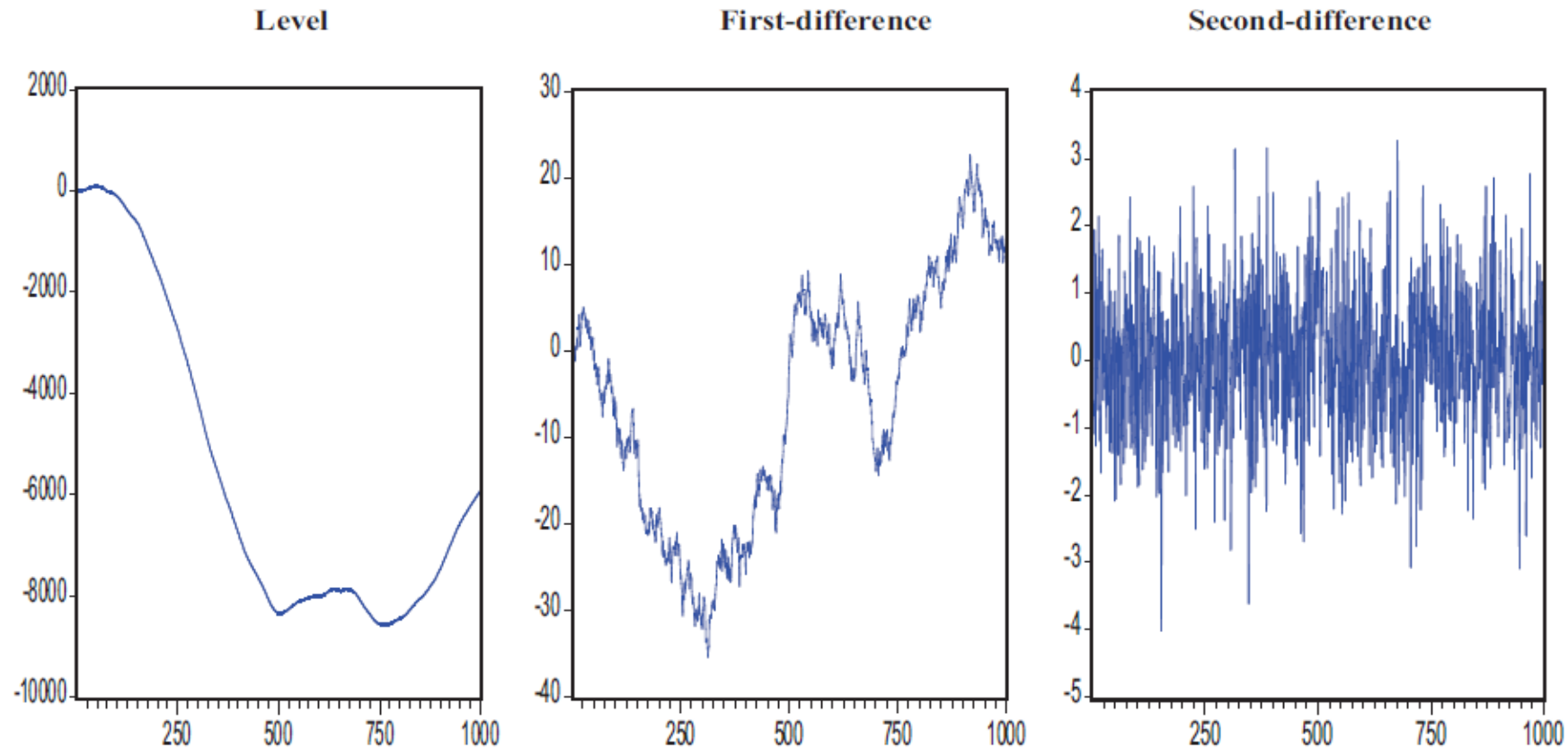
Figure 3: Determining d…
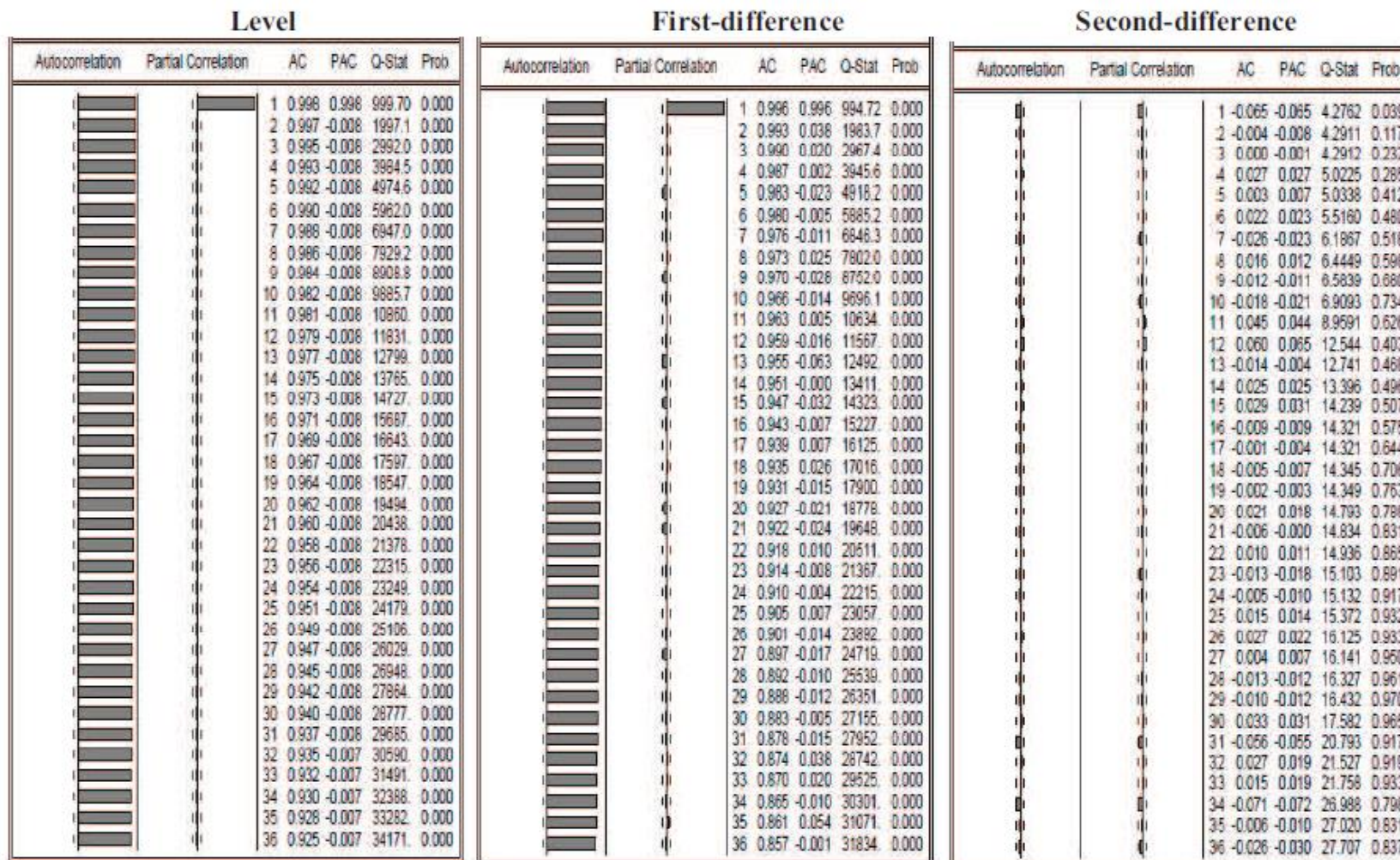
Seasonality

- A stochastic process is said to be a seasonal (or periodic) time series with periodicity s if Z_t and Z_{t+ks} have the same distribution.

- Such seasonal series are common in business, economics and finance

  - Business : The series of monthly sales of a department store, etc ;

  - Economics : Disaggregate price series, unemployment, components of GDP, monetary aggregates, stocks, etc ;

  - Finance : Intraday data, month-effect, day-effect, etc.
- Seasonal variations can constitute a large part of total variation in certain (macroeconomic) time series.

Identifying Seasonality

Figure 4: Determining d…Retail and Food Service Sales: 1992. 1-2008.8
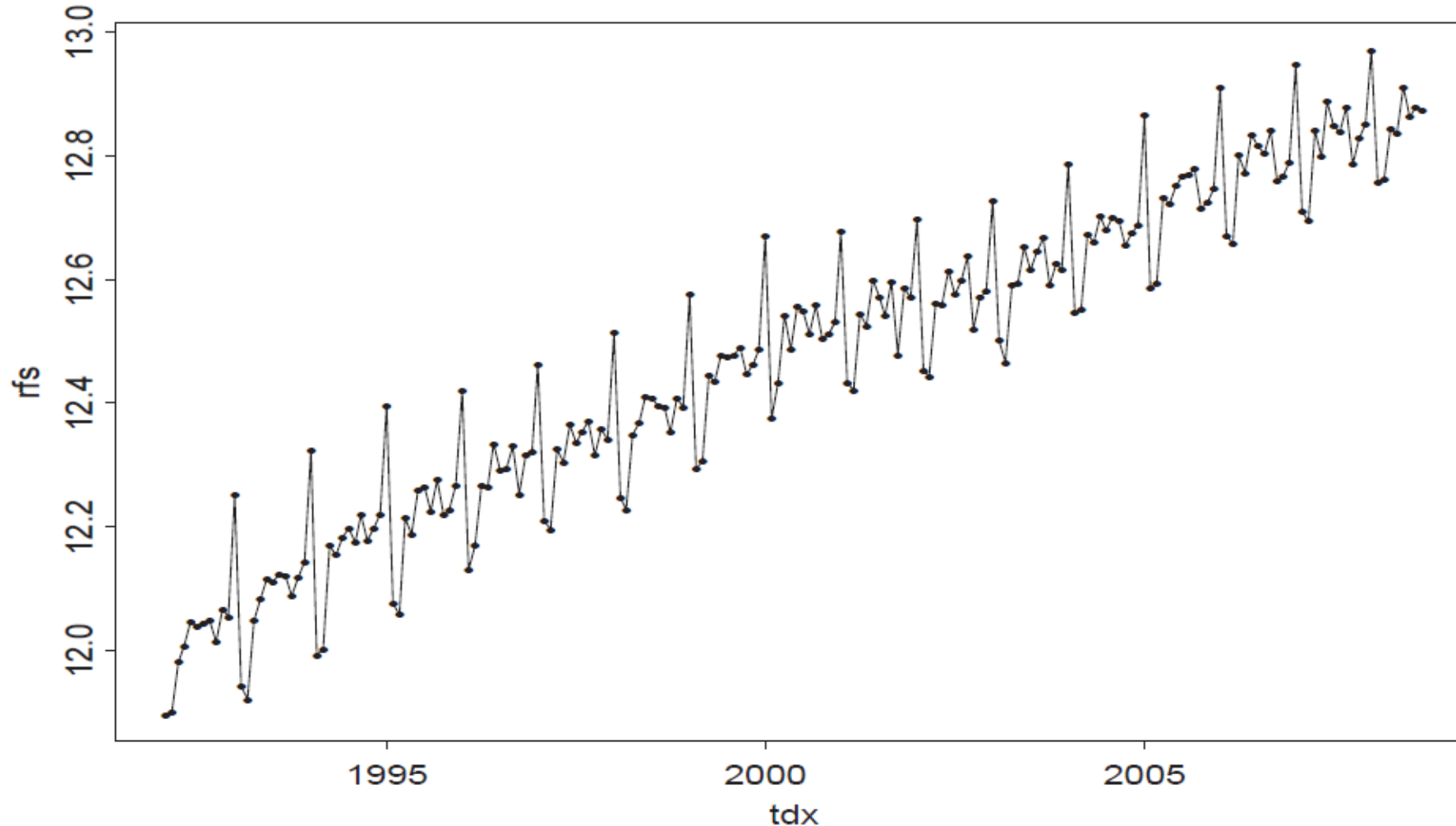
Figure 5: Determining d…Residential and Commercial Energy Consumption 1982-1993

Identifying Seasonality
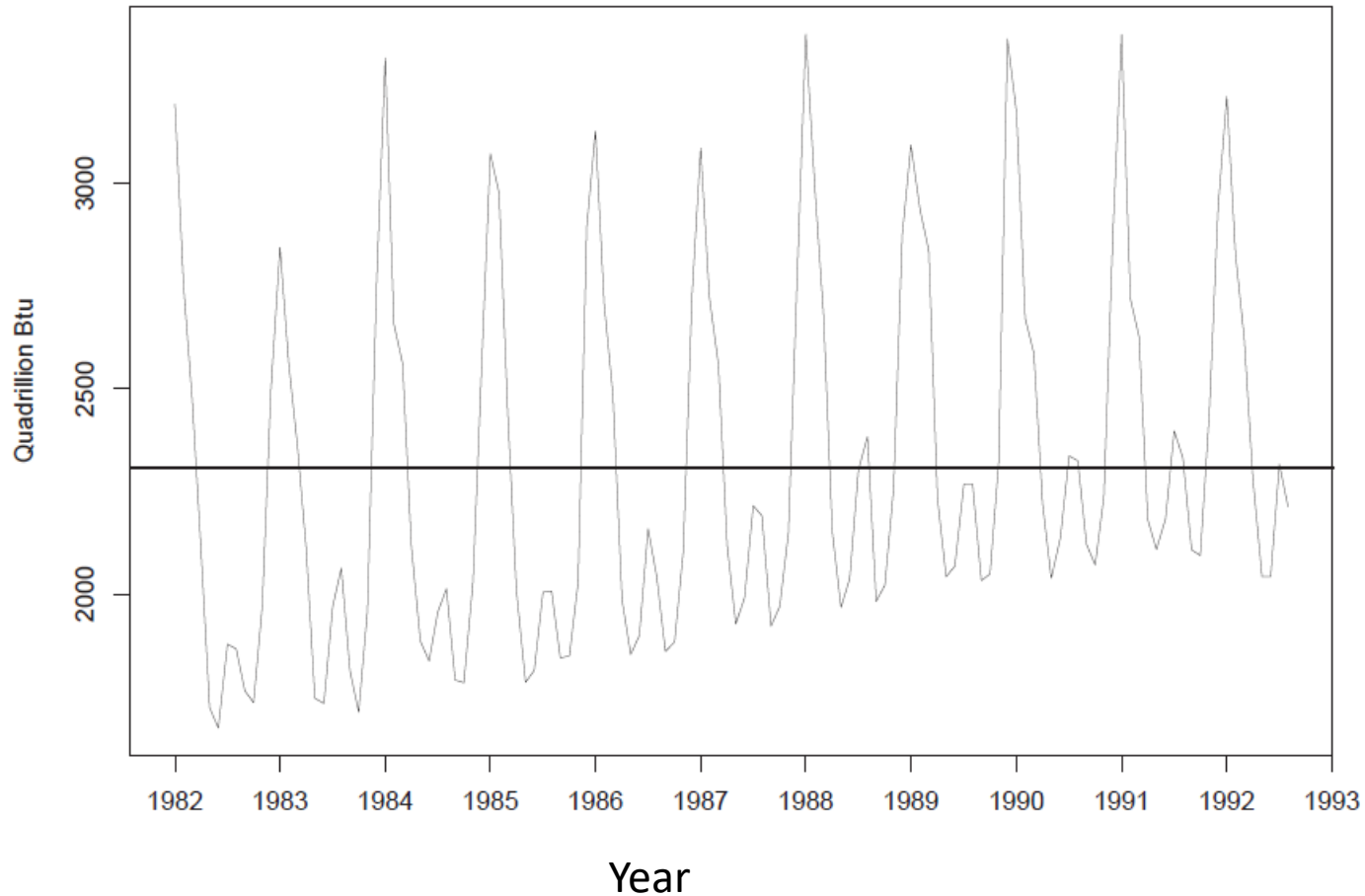
- Seasonality can be assessed using graphical procedures :

  - Plot the series ; The ACF ;

  - The spectral density

- Which treatment(s) for seasonality ?

  - Working with seasonally-adjusted series ?

  - Joint modeling of the seasonal and nonseasonal component ?

Identifying Seasonality

- Using seasonally-adjusted series does not mean that the seasonal pattern is completely removed. This is particularly true if the entire span of data is not used.

- Seasonally adjusting (or pre-filtering) a time series (e.g., the well-known Census X-12 makes extensive use of such filters) can
- distort sone of its important properties and may complicate further analysis.

- For a joint modeling, two approaches :

    - Models with seasonal dummy variables ; SARIMA models (see Appendix).

Identifying Seasonality

Figure 6: Example Seasonality

Identifying p and q

The model order (p, q) can be determined by using graphical plots of the ACF and PACF.

Table 1: Main characteristics of ARMA(p,q) models :

	AR(p)	**MA(q)**	**ARMA(p,q)**
ACF	Tails off	Cuts off after q	Tails off
PACF	Cuts off after p	Tails off	Tails off

Identifying p and q

Different shapes of the autocorrelation function :

- **Exponential decay to zero** : Autoregressive model (use the partial autocorrelation plot to identify the order p)

- **Damped oscillations decaying (exponentially) to zero** : Autoregressive model

- **One or more spikes, the rest is essentially zero** : Moving average model (order q identified by where autocorrelation plot becomes zero)

- **Exponential decay starting after a few lags** : Mixed autoregressive and moving average model

Identifying p and q

- Different shapes of the autocorrelation function (cont'd) :

  - **No significant autocorrelations (zero or close to zero)** : White noise

  - **High values at fixed intervals** : Include seasonal autoregressive terms

  - **No decay to zero or very slow decay** : Non-stationarity or long-memory effects...

Identifying p and q

In practice, identifying p and q using the ACF and PACF involves a trial and error approach...with more or less subjectivity in interpreting these functions : "real data" rarely exhibit simple patterns !

The practical rule is :

1. To choose an upper bound for p, say p_{max}, and q, say q_{max} ;

2. Estimate all models with $0 \leq p \leq p_{max}$ and $0 \leq q \leq q_{max}$;

3. Use information criteria (or other procedures) to discriminate among the competing models.

Estimation and Information criteria

1. Estimation

This step involves the estimation of the parameters of the models identified

(specified) in Step 1 (with or without constraints). This can be done using

- (Nonlinear) Linear least squares method ;

- Maximum likelihood estimation

- (Generalized) Method of moments Etc

Estimation and Information criteria

1. Information Criteria

The goodness-of-fit of the model can be assessed with the residual variance

$$\hat{\sigma}^2(k) = \frac{1}{T} \sum_{t=1}^{T} \hat{\epsilon}_t^2$$

where T is the number of observations used for the estimation, k is the total number of parameters estimated (e.g., $k = p + q + 2$), $\hat{\epsilon}_t$ is the adjusted residual at time t.

- The residual sum of squares is inversely proportional to the number of degrees of freedom : "large" models with many parameters are often chosen when the sample size is large ;

- Choosing over-parameterized (profligate) models tends to fit to data specific features, which would not be replicated out-of-sample $\Rightarrow$ the models may appear to fit the data very well (with a large R^2) but would give inaccurate forecasts !

Estimation and Information criteria

- Broadly speaking, an information criterion measures the goodness of

    fit of the model by assigning an informational cost to the number of parameters to

    estimate (making clear the idea that it is always possible to more or less trade off

    p versus q in selecting the ARMA orders)

- Information criteria embody two factors: A term which is a function of the residual

    sum of squares ;

Estimation and Information criteria

- A term which penalizes for the loss of degrees of freedom from adding extra parameters.

- Using an additional lag leads to two opposite effects

- ↓ residual sum of squares ;

- ↑ value of the penalty term

- ⇒ IC ↓ if and only if $|\Delta SSR| > |\Delta PT|$

- The object is thus to choose the number of parameters which **minimizes** the value of the information criteria.

Akaike information criterion

The Akaike information criterion (AIC) is defined to be

$$AIC(k) = \log \cdot \hat{\sigma}^2(k)^{\Sigma} + \frac{2}{T}(k)$$

where k is the total number of parameters estimated.

Remarks :

- AIC may give more than one minimum (p, q) ;
- AIC depends on the normality assumption ;
- AIC is not consistent : it will deliver on average too large a model
- (even with $T \to \infty$)—AIC tends to over-parameterize.
- AIC is generally efficient (small average variation in selected model orders from different samples within a given population).

Schwarz bayesian information criterion

The Schwarz bayesian information criterion (SBIC) is defined to be

$$\text{SBIC}(p,\ q) = \log \hat{\sigma}^2(k) + \frac{k}{T}\log(T)$$

where k is the total number of parameters estimated.

Remarks :

- SBIC embodies a much stiffer penalty term than AIC, i.e. SBIC penalizes larger models more than AIC and thereby tends to select lower-order models than the AIC.

- SBIC is strongly consistent in selecting p and q : if the data are truly generated by an ARMA(p, q) process, the SBIC picks the true model with probability 1 as the sample size approaches infinity.

- SBIC is less efficient than AIC.

Hannan-Quinn information criterion

The Hannan-Quinn information criterion (HQIC) is defined to be

$$HQIC(p, q) = \log \hat{\sigma}^2(k) + \frac{2k}{T} \log(\log(T))$$

where k is the total number of parameters estimated.

Remarks :
- HQIC contains a penalty term which is somewhere between AIC and SBIC.
- HQIC is strongly consistent in selecting p and q.

Practical implementation

General methodology

Set upper bounds, p_{max} and q_{max}, for the AR and MA order, respectively ;

Fit all possible ARMA(p,q) models for $p \leq p_{max}$ and $q \leq q_{max}$ using a **common sample size** ;

The best models (possibly more than one !) satisfy :

$$\min_{p \leq p_{max}, q \leq q_{max}} AIC(p, q)$$

$$\min_{p \leq p_{max}, q \leq q_{max}} SBIC(p, q)$$

$$\min_{p \leq p_{max}, q \leq q_{max}} HQIC(p, q).$$

Remark : Finite sample corrections are available !

Diagnostic checking

Overview

- The main objective is to check the adequacy of the model(s) selected in the previous step.

- Notably all of the relevant information from the data should be extracted by the model. For instance, the part of the data unexplained by the model (residuals) should be small and no systematic or predictable patterns should be left in the residuals (i.e., weak white noise).

- Diagnostic testing in the Box-Jenkins methodology essentially involves the statistical properties of the error terms (normality assumption, weak white noise assumption).

Residual diagnostics

Residual diagnostics

(a) Procedures

(s_t) is a (weak) white noise process, i.e. the residuals of an estimated model should exhibit white noise-like behavior : departure from this assumption means that some information can still be exploited in the modeling...

Two methods

Graphical procedure

(a) Plot of the residuals : do they exhibit systematic patterns ? Use the SACF and SPACF (as in the identification step) of the residuals : Do they have significant elements ?

(b) Testing procedure : autocorrelation tests.

Autocorrelation tests

- One problem with checking the significance of individual (partial) autocorrelation is that each element might be individually insignificant, but all (or a subset) of the elements may be jointly significant.

- The Box-Pierce Q-statistic (or portmanteau test) tests the joint hypothesis that the first K autocorrelations of the adjusted error terms are jointly zero :

    - $H_0 : \rho_{\hat{s}}(1) = \rho_{\hat{s}}(2) = \cdots = \rho_{\hat{s}}(K) = 0.$

Diagnostic checking Residual

- The test statistic is given by :

$$Q = T \sum_{k=1}^{K} \hat{\rho}_S^2(k)$$

- where $\rho^2(k)$ is the k-th order sample autocorrelation of the estimated residuals, T is the sample size, and K is chosen sufficiently large.

- The Q-test has an asymptotic chi-square (χ^2) distribution with $K - p - q$ degrees of freedom. The null hypothesis of uncorrelated (estimated) residuals is rejected if the Q exceeds the tabulated critical value (for a chosen significance level).

- The Q-test is only asymptotically valid and may perform poorly for small and medium size samples.

- Refinements have been introduced in the literature : Ljung-Box test, McLeod-Li test, Monti's test. etc.

Discussion

- Diagnostic testing based on autocorrelation tests could only reveal a model
- that is under parameterized ("too small") and would not reveal a model that is over parameterized ("too large").

- Autocorrelation of residuals can give rise to common factors especially in over fitted models.
- This makes estimation difficult and the statistical tests ill-behaved.

- For example, if the true data generating process is an ARMA(p, q) and one deliberately then fits an ARMA(p+1,q+1) there will be a common factor (all of the parameters in the latter model can be identified).

- Other residual procedures include the tests for the normality assumption (especially for maximum likelihood estimation), etc.

DATA ANALYTICS

Image Courtesy

https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Time+Series+Analysis:+The+Basics

References

Text Book:

"Business Analytics, The Science of Data-Driven Making", U. Dinesh Kumar, Wiley 2017

Chapter-13 Concept of stationarity, DF and ADF test and transforming non stationary process to a stationary one 13.14.1-13.14.3 in text

DATA ANALYTICS

Image Courtesy

https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Time+Series+Analysis:+The+Basics

https://365datascience.com/arimax/
https://machinelearningmastery.com/gentle-introduction-box-jenkins-method-time-series-forecasting/,
https://math.unice.fr/~frapetti/CorsoP/Chapitre_5_IMEA_1.pdf)

THANK YOU

Jyothi R
Assistant Professor, Department of
Computer Science
jyothir@pes.edu