Machine Learning for Time Series Forecasting

Machine Learning ID Meet Up #3 Surabaya
March 10, 2018
DILo Surabaya

***Novri Suhermi**
Suhartono

Department of Statistics
Faculty of Mathematics, Computation, and Data Science
Institut Teknologi Sepuluh Nopember

novri@statistika.its.ac.id

Novri SUHERMI

Master of Science, Applied Mathematics in Finance, Insurance, and Risk, Sept 2014 – Sept 2016
Université Paris Diderot – Paris 7, Paris, France
Master of Science, Statistics, September 2013 – September 2014
Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
Bachelor of Science, Statistics, September 2010 - March 2014
Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
September 2017 – present

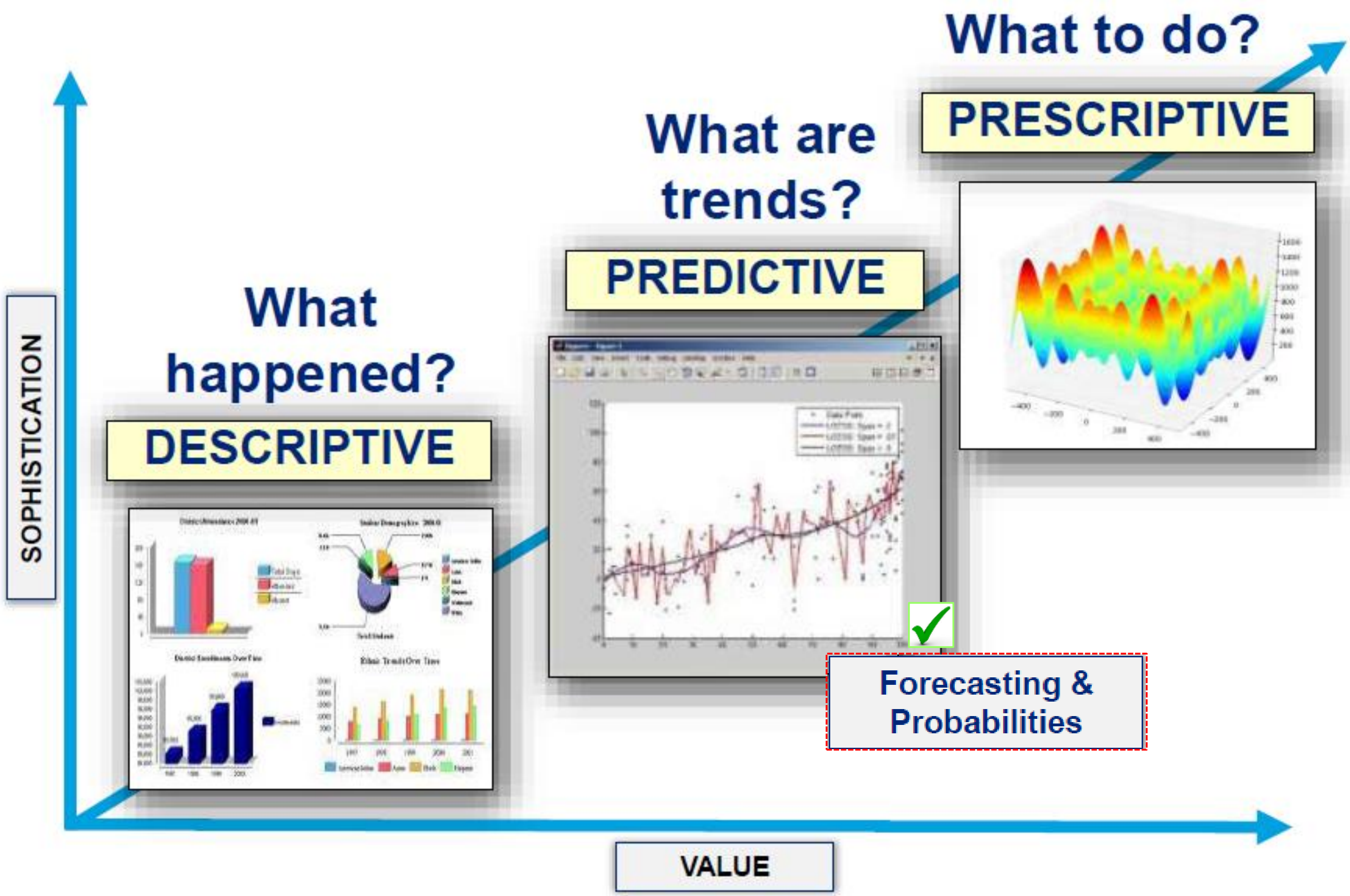
Lecturer, Department of Statistics, ITS**,** September 2017 – present
Member of Statistics for Business and Industry Laboratory
Research area interest: Time Series, Forecasting, Machine Learning, Reliability, Computational Statistics

Data Analyst, GDP Labs & Kaskus, Jakarta, Indonesia, December 2016 – September 2017

Research Assistant, Laboratoire d'Informatique, Université Paris Descartes, Paris, France, April 2016-October 2016
Developed a method to identify Near-infrared Spectroscopy (NIRS) data using time series management and analysis techniques.

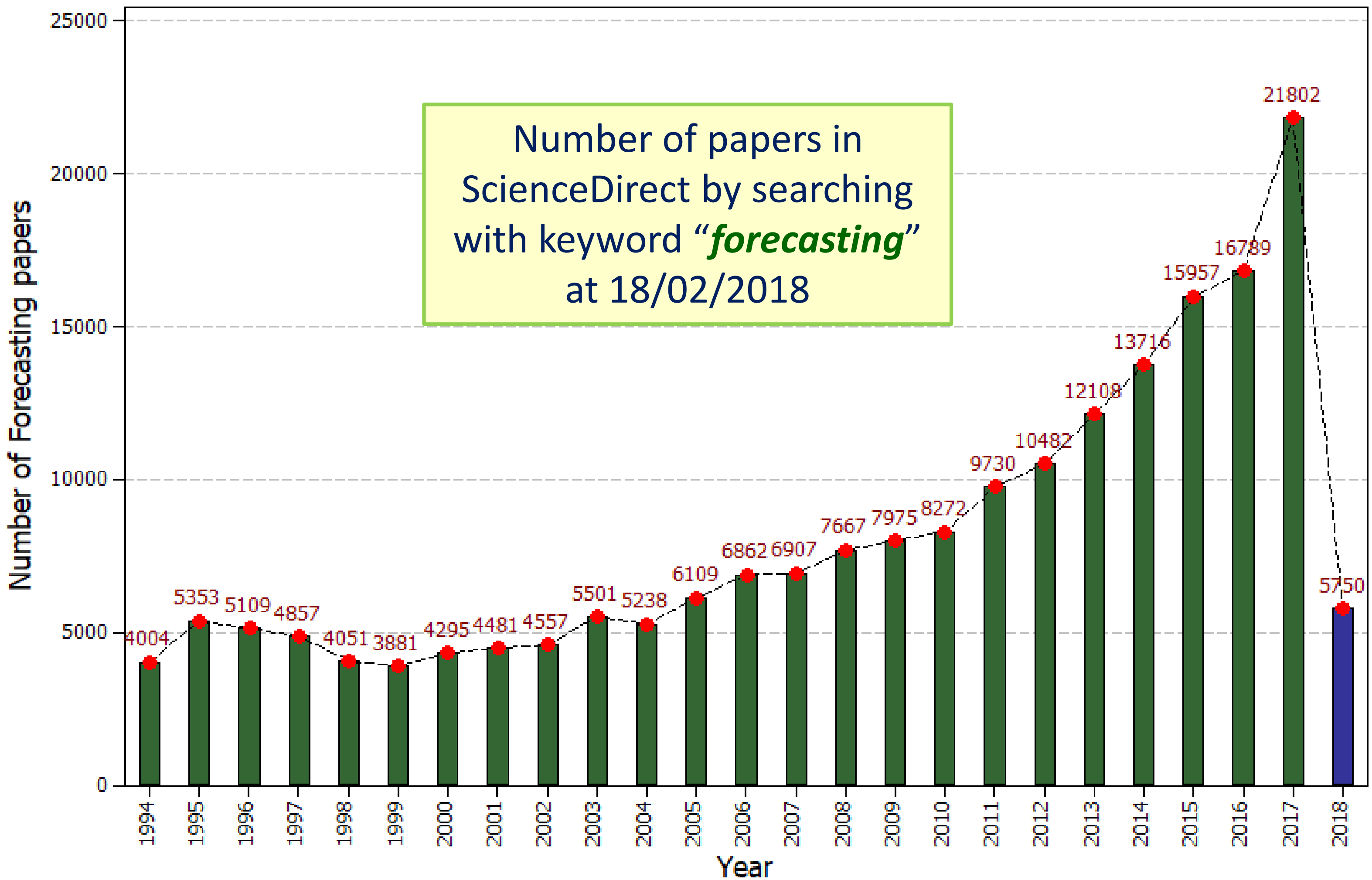
Research Assistant, Laboratoire Jean Kuntzmann, Grenoble-INP, Grenoble, France, April 2015-September 2015
Developed a program of exact monte carlo goodness-of-fit test for imperfect maintenance models using R with Rcpp package (Seamless R and C++ Integration).

Motivation

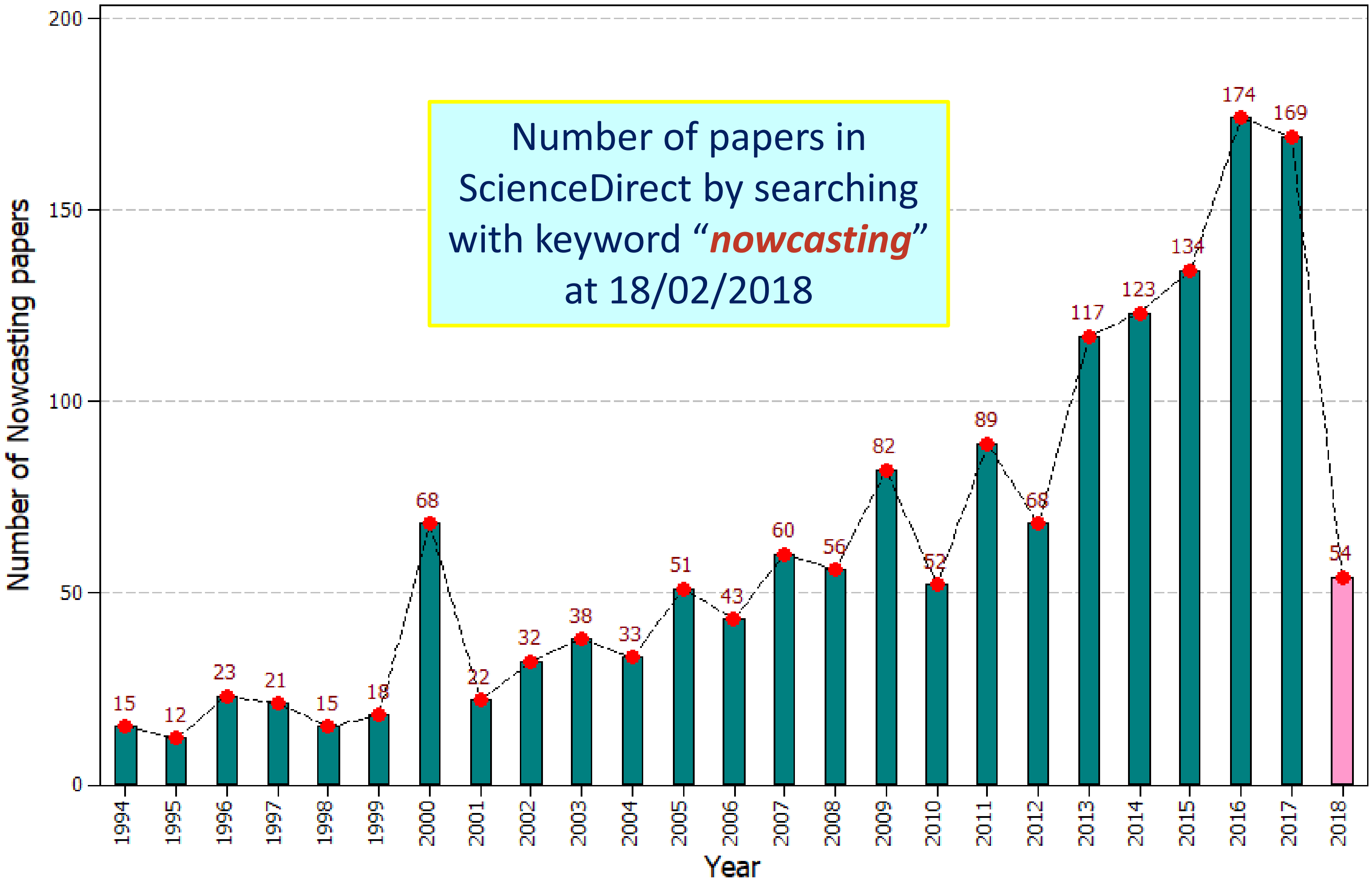


What to do?
PRESCRIPTIVE
What are trends?
PREDICTIVE
What happened?
DESCRIPTIVE
SOPHISTICATION
Forecasting & Probabilities
VALUE

Motivation


Number of papers in ScienceDirect by searching with keyword "*forecasting*" at 18/02/2018
Number of Forecasting papers
Year
25000
20000
15000
10000
5000
0
4004
5353
5109
4857
4051
3881
4295
4481
4557
5501
5238
6109
6862
6907
7667
7975
8272
9730
10482
12108
13716
15957
16789
21802
5750
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018

Motivation



Number of papers in ScienceDirect by searching with keyword "*nowcasting*" at 18/02/2018
Number of Nowcasting papers
Year
200
150
100
50
0
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
15
12
23
21
15
18
68
22
32
38
33
51
43
60
56
82
52
89
68
117
123
134
174
169
54

Motivation

☞ ***The M3-Competition***: results, conclusions and implications

(1) Statistically sophisticated or complex methods do not necessarily provide more accurate forecasts than simpler ones.

(2) The relative ranking of the performance of the various methods varies according to the accuracy measure being used.

(3) The accuracy when various methods are being combined outperforms, on average, the individual methods being combined and does very well in comparison to other methods.

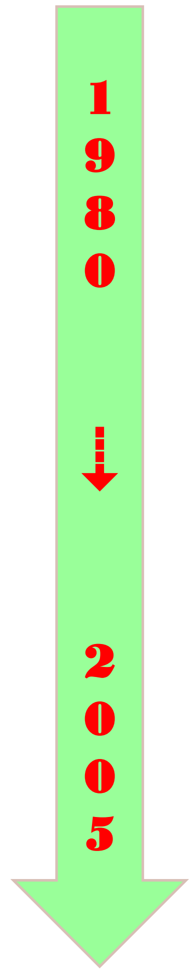
(4) The accuracy of the various methods depends upon the length of the forecasting horizon involved.

Makridakis & Hibon (International Journal of Forecasting, 2000)

25 years of time series forecasting

De Gooijer & Hyndman (International Journal of Forecasting, 2006)

- 25 years of time series forecasting
  - Introduction
  - Exponential smoothing
    - Preamble
    - Variations
    - State space models
    - Method selection
    - Robustness
    - Prediction intervals
    - Parameter space and model properties
  - ARIMA models ✓
    - Preamble
    - Univariate
    - Transfer function
    - Multivariate
  - Seasonality ✓
  - State space and structural models and the Kalman filter



- Nonlinear models
  - Preamble
  - Regime-switching models
  - Functional-coefficient model
  - Neural nets ✓
  - Deterministic versus stochastic dynamics
  - Miscellaneous
- Long memory models
- ARCH/GARCH models
- Count data forecasting
- Forecast evaluation and accuracy measures
- Combining ✓
- Prediction intervals and densities
- A look to the future
- Acknowledgments
- References

Forecasting Methods

1. **Exponential Smoothing:** Holt-Winter's method, Holt-Winter-Taylor's method
2. **Decomposition Method**
3. **Trend Analysis & Time Series Regression**
4. **ARIMA & ARIMAX:** Intervention Analysis, Transfer Function model
5. **Neural Networks: FFNN, RBFN, GRNN, RNN**
6. **Adaptive Neuro Fuzzy Inference Systems (ANFIS)**
7. **Multiresolution Autoregressive (MAR)**
8. **Wavelet Neural Networks (WNN)**
9. **Support Vector Regression (SVR)**
10. **Fuzzy Time Series (FTS)**
11. **Quantile Regression Autoregressive & ARIMAX**
12. **Quantile Regression Neural Networks (QRNN)**
13. **Singular Spectrum Analysis (SSA) - Linear Recurrence Relations (LRR)**
14. **Deep Learning**

Individual Methods

Forecasting Methods

1. Winter's model & ARIMA
2. Winter's model & Neural Networks
3. Winter's model & Fuzzy Time Series
4. Decomposition Method & ARIMA
5. Decomposition Method & Neural Networks
6. Time Series Regression & Neural Networks
7. Time Series Regression & ANFIS
8. Time Series Regression & Support Vector Regression
9. ARIMAX & Neural Networks
10. ARIMAX & ANFIS
11. ARIMAX & Support Vector Regression
12. ARIMAX & Quantile Regression Neural Networks
13. ARIMAX & Deep Learning

Hybrid Methods

Forecasting Methods

1. Decomposition method & ARIMA
2. Decomposition method & Neural Networks
3. Decomposition method & Fuzzy Time Series

4. Wavelet Transform - MODWT & Autoregressive Model
5. Wavelet Transform - MODWT & Neural Networks
6. Wavelet Transform - MODWT & ANFIS

7. Singular Spectrum Analysis & Time Series Regression
8. Singular Spectrum Analysis & ARIMA
9. Singular Spectrum Analysis & Neural Networks
10. Singular Spectrum Analysis & ANFIS
11. Singular Spectrum Analysis & Support Vector Regression
12. Singular Spectrum Analysis & Deep Learning

Hybrid Decomposition Methods

Forecasting Methods

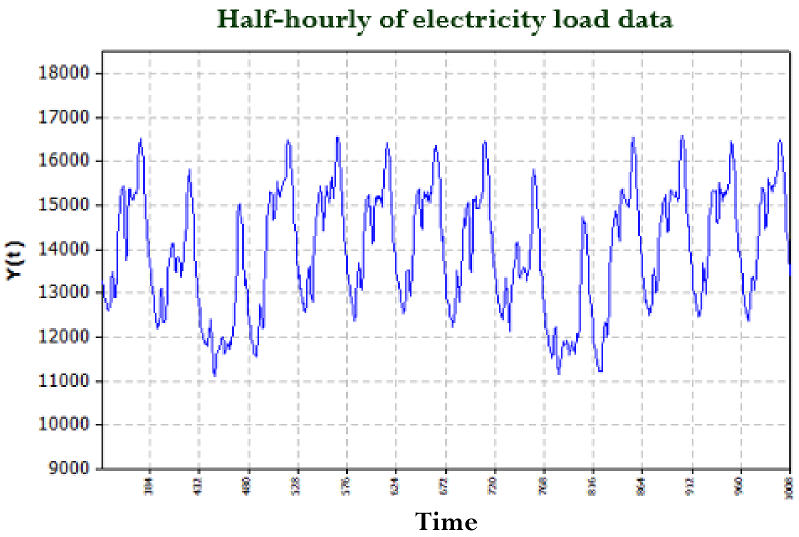
☞ **The main patterns and problems of time series forecasting**

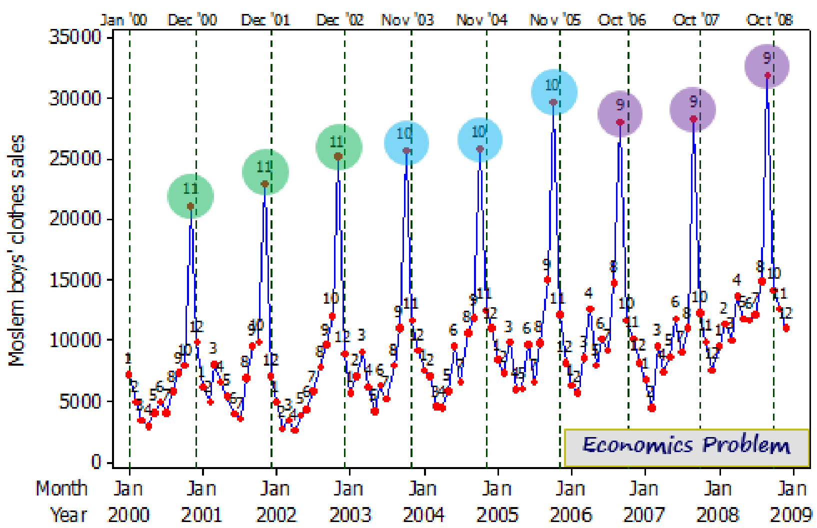
1. **TREND**: linear vs non linear
2. **SEASONAL**: single vs multiple period
3. **STATIONER**: linear vs non linear
4. **CALENDAR VARIATION** effect
5. **OUTLIER & INTERVENTION** analysis
6. **ADDITIVE** vs **MULTIPLICATIVE** pattern

Outlines

1. **Time Series Regression (TSR) model**
   - ☞ Trend, Seasonal & Calendar Variation pattern
   - ☞ Multiple Seasonal models.
2. **ARIMA & ARIMAX Model**
   - ☞ Seasonal Model: Multiplicative, Additive, Subset
   - ☞ Multiple Seasonal models.
   - ☞ ARIMAX: Calendar Variation models.
3. **Neural Network**
   - ☞ Feedforward Neural Network (FFNN)
   - ☞ Deep Feedforward Network.

Time Series Regression for Trend, Seasonal & Calendar Variation





Introduction

- ✓ General time series **"PATTERN"**
  - ✍ Stationary
  - ✍ Trend: ***linear*** & ***nonlinear***
  - ✍ Seasonal: ***additive*** & ***multiplicative***
  - ✍ Cyclic
  - ✍ **<u>Calendar Variation</u>**

Introduction cont'

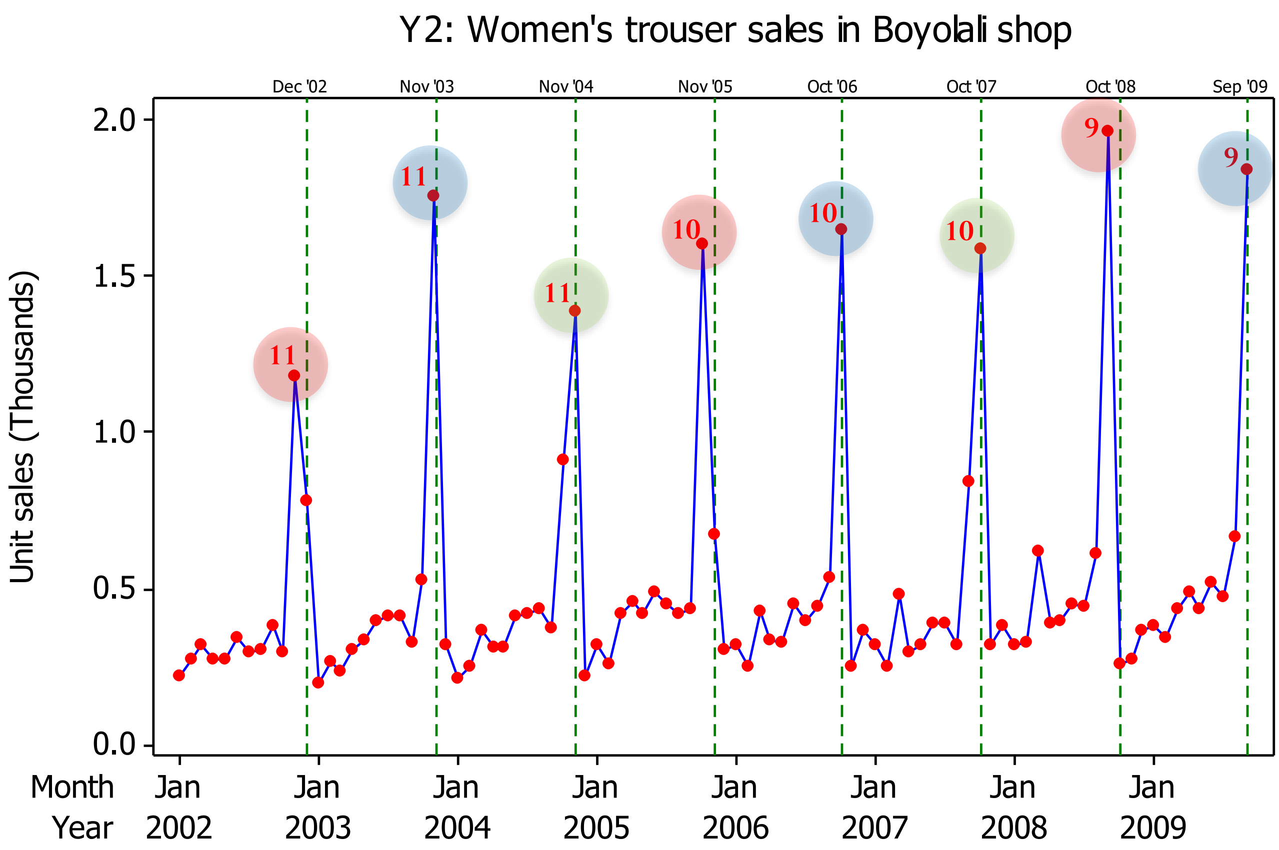
- Two kinds of calendar variation effects:
  1. Trading day effects

     The levels of economics or business activities may change depending on the day of the week. The composition of days of the week varies from month to month and year to year.
  2. Holiday (traditional festivals) effects

     Some traditional festivals or holidays, such as Eid ul-Fitr, Easter, Chinese New Year, and Jewish Passover are set according to lunar calendars and the dates of such holidays may vary between two adjacent months in the Gregorian calendar from year to year.

Introduction

cont'



Introduction

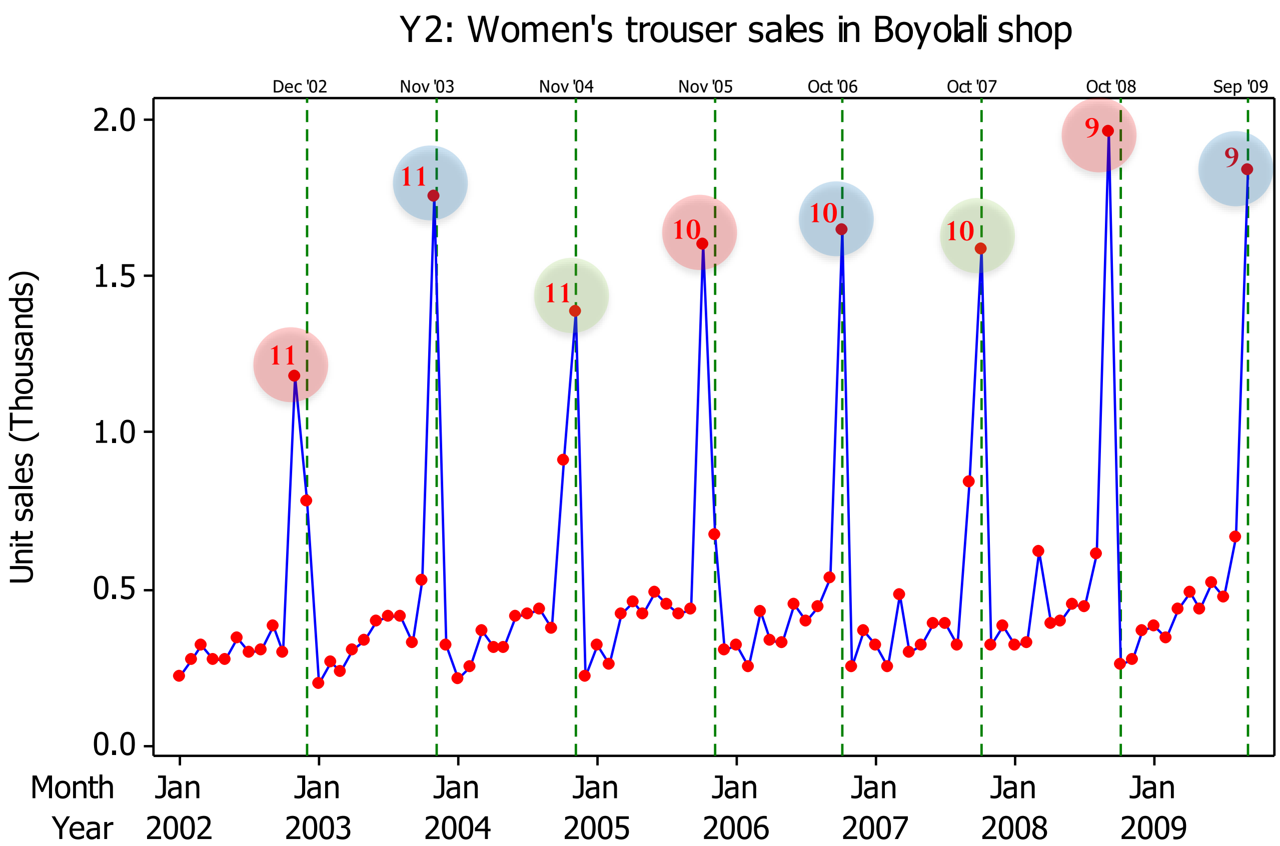
cont'

Eid holidays for the period 2002 to 2011

Year	Date	Explanation
2002	06-07 December	There are 5 days before Eid in December
2003	25-26 November	There are 24 days before Eid in November
2004	14-15 November	There are 13 days before Eid in November
2005	03-04 November	There are 2 days before Eid in November
2006	23-24 October	There are 22 days before Eid in October
2007	12-13 October	There are 11 days before Eid in October
2008	01-02 October	There is 0 day before Eid in October
2009	21-22 September	There are 20 days before Eid in September
2010	10-11 September	There are 9 days before Eid in September
2011	30-31 August	There are 29 days before Eid in August

Introduction

cont'



Introduction

cont'

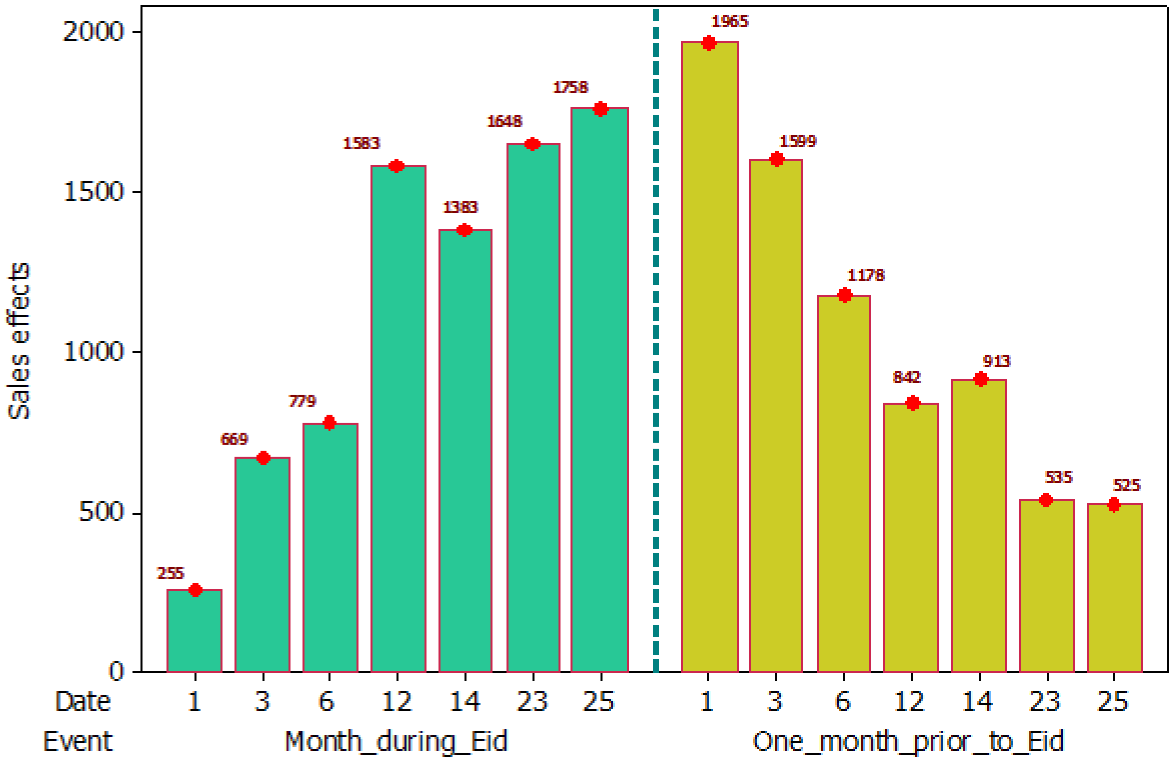


Fig. 2. Bar chart of Eid effects on the women's trouser sales in the month during and one month prior to the Eid celebration in Boyolali shop.

Modeling method

- Model for linear **trend**:

$$y_t = \beta_0 + \beta_1 t + w_t \qquad \ldots (1)$$

- Regression with dummy variable for **seasonal** pattern:

$$y_t = \beta_1 S_{1,t} + \beta_2 S_{2,t} + \cdots + \beta_s S_{s,t} + w_t \qquad \ldots (2)$$

- Regression for **calendar variation** effects:

$$y_t = \beta_0 + \beta_1 V_{1,t} + \beta_2 V_{2,t} + \cdots + \beta_p V_{p,t} + w_t \qquad \ldots (3)$$

The Proposed Model

- **Time Series Regression Model:**

$$Y_t = \boxed{\delta_1\, t} + \boxed{\beta_1 S_{1,t} + \beta_2 S_{2,t} + \cdots + \beta_s S_{s,t}} + \sum_j \alpha_j D_{j,t} + \sum_j \gamma_j D_{j,t-1} +$$

$$\boxed{\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p}} + \varepsilon_t \, .$$

TREND

STATIONER AUTOCORRELATED

SEASONAL

CALENDAR VARIATION

The proposed procedure

Step 1: Determination of **dummy variable** for calendar variation period.

Step 2: Determination of **deterministic trend** and **seasonal** model.

Step 3: **Simultaneous estimation** of **calendar effects** and other patterns.

Step 4: **Diagnostic checks** on error model. If error is **not white noise**, add significant lags (**autoregressive order**) based on ACF and PACF plots of error model.

Step 5: **Re-estimate** calendar effect, other pattern (trend, seasonal), and appropriate lags (autoregressive order) **simultaneously** for the model.

Simulation Data using R

- **TREND**

$$T_t = \beta t$$

In this simulation, $\beta = 0.1$.

- **SEASONAL**

$$M_t = \gamma_1 M_{1,t} + \gamma_2 M_{2,t} + \ldots + \gamma_{12} M_{12,t}$$

The values for γ are:

$$M_t = 20M_{1,t} + 23M_{2,t} + 25M_{3,t} + 23M_{4,t} + 20M_{5,t} + 15M_{6,t} + 10M_{7,t} + 7M_{8,t} + 5M_{9,t} + 7M_{10,t} + 10M_{11,t} + 15M_{12,t}$$

Simulation Data using R

CALENDAR VARIATION

$$V_t = \delta_1 V_{1,t} + \delta_2 V_{2,t} + \delta_3 V_{3,t} + \delta_4 V_{4,t} + \omega_1 V_{1,t-1} + \omega_2 V_{2,t-1} + \omega_3 V_{3,t-1} + \omega_4 V_{4,t-1}$$

where the value of δ and ω are:

$$V_t = 23V_{1,t} + 37V_{2,t} + 44V_{3,t} + 48V_{4,t} + 56V_{1,t-1} + 42V_{2,t-1} + 34V_{3,t-1} + 30V_{4,t-1}$$

NOISE

- Linear: AR(1) models

$$N_t = 0,7N_{t-1} + a_t, \text{ with } a_t \sim IIDN(0,1).$$

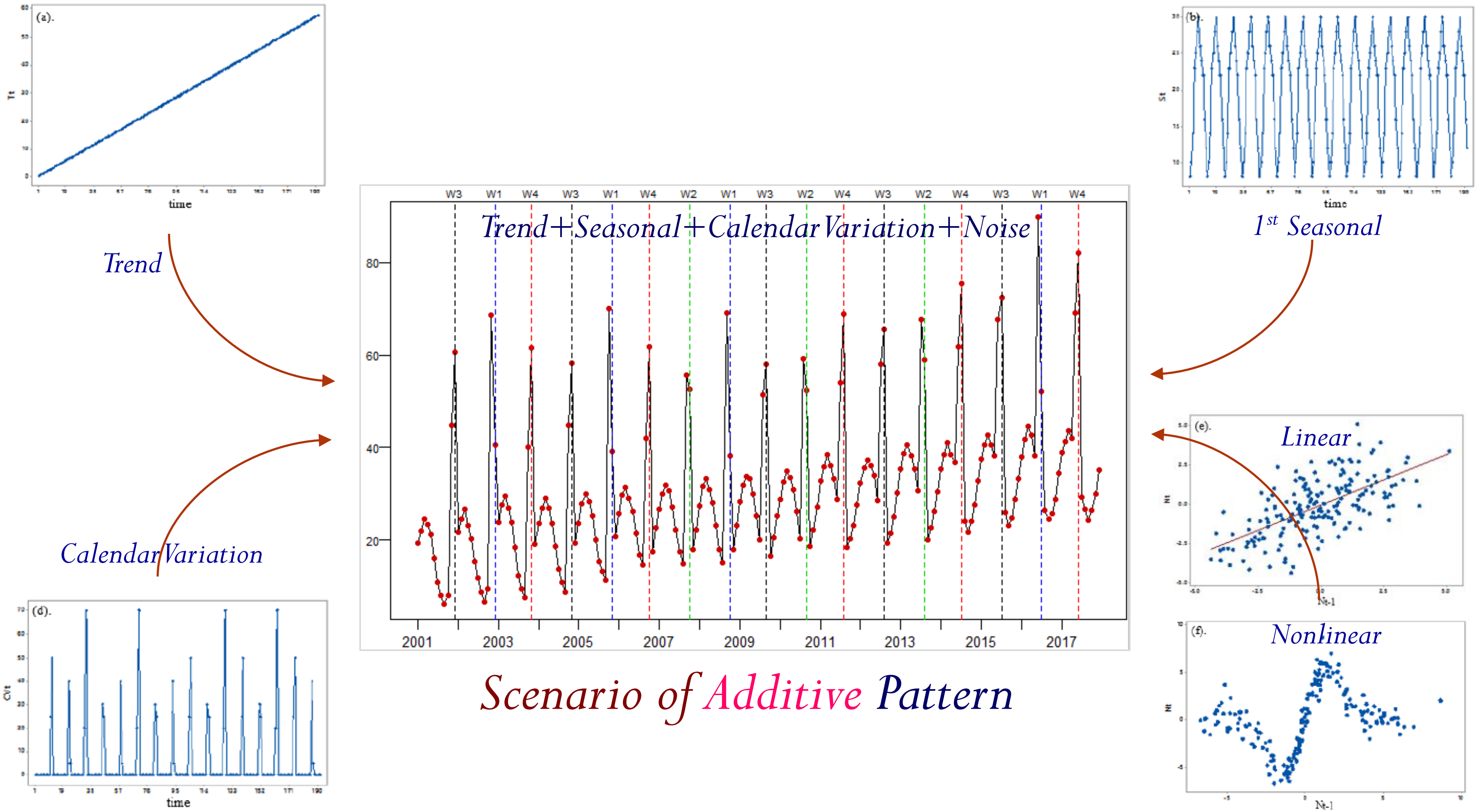
- Nonlinear: ESTAR(1) models

$$N_t = 6,5N_{t-1}.\exp\left(-0,25N_{t-1}^2\right) + a_t, \text{ with } a_t \sim IIDN(0,1).$$

Dummy for Calendar Variation

Year	Date of Eid ul-Fitr	Week of Eid ul-Fitr	Month of Eid ul-Fitr
2001	17-18	3rd week	December
2002	06-07	1st week	December
2003	25-26	4th week	November
2004	13-14	2nd week	November
2005	03-04	1st week	November
2006	23-24	4th week	October
2007	12-13	2nd week	October
2008	01-02	1st week	October
2009	20-21	3rd week	September
2010	09-10	2nd week	September
2011	30-31	4th week	August
2012	18-19	3rd week	August
2013	08-09	2nd week	August
2014	28-29	4th week	July
2015	19-20	3rd week	July
2016	06-07	1st week	July
2017	25-26	4th week	June

Simulation data



Scenario of Additive Pattern

Step 1

- Based on the time series plot, two dummy variables are used for evaluating calendar variation effect, i.e.
  - The months prior to Eid ul Fitr,

    $D_{j,t-1}$ = dummy variable for ONE month prior to Eid ul-Fitr celebration.
  - During the month of Eid ul-Fitr celebration,

    $D_{j,t}$ = dummy variable for during the month of Eid ul-Fitr celebration.
  - j = week of Eid ul-Fitr celebration

Step 2 - 5

- Model for linear trend:

$$y_t = \beta_0 + \beta_1 t + w_t$$

- Regression with dummy variable for seasonal pattern:

$$y_t = \beta_1 S_{1,t} + \beta_2 S_{2,t} + \cdots + \beta_s S_{s,t} + w_t$$

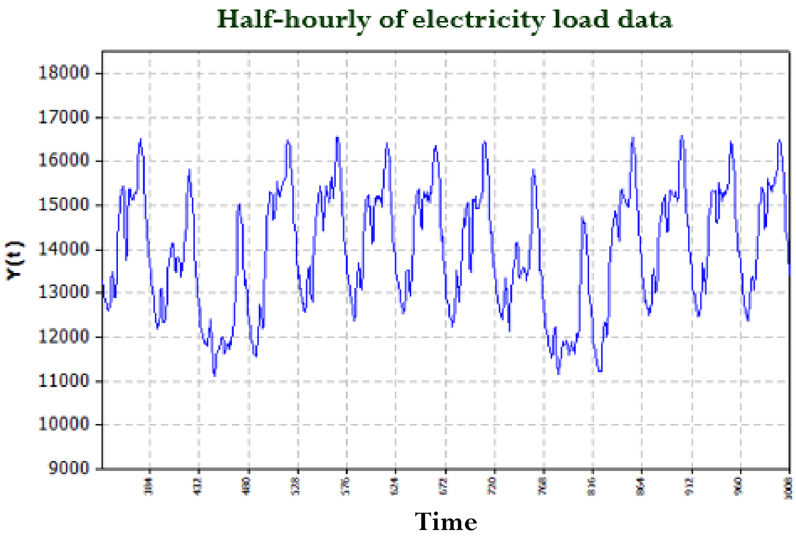
- Regression for calendar effects and other patterns:

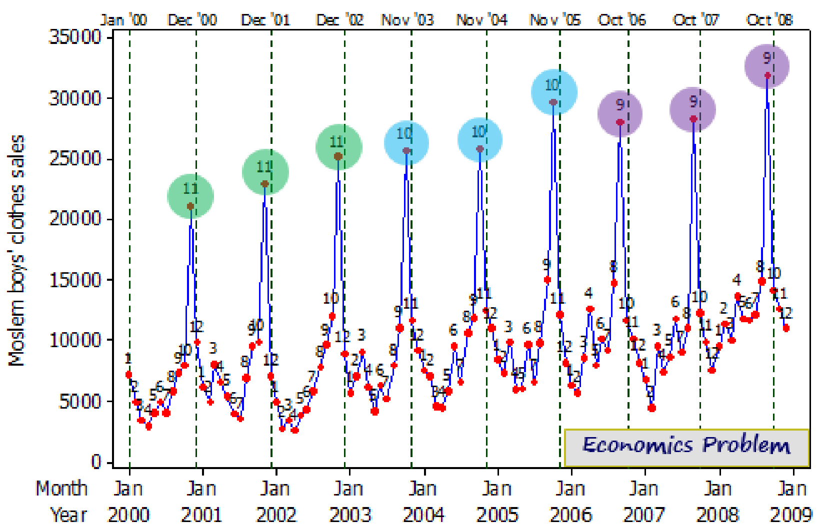
$$Y_t = \delta_1 t + \beta_1 S_{1,t} + \beta_2 S_{2,t} + \cdots + \beta_s S_{s,t} + \sum_j \alpha_j D_{j,t} + \sum_j \gamma_j D_{j,t-1} + N_t \,.$$

- **Final Regression model is:**

$$Y_t = \delta_1 t + \beta_1 S_{1,t} + \beta_2 S_{2,t} + \cdots + \beta_s S_{s,t} + \sum_j \alpha_j D_{j,t} + \sum_j \gamma_j D_{j,t-1} + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t \,.$$

ARIMAX Model for Trend, Seasonal & Calendar Variation





The Proposed ARIMAX Model

- The 1st proposed Model → **ARIMAX-1**: **stochastic** TREND-SEASONAL

$$Y_t = \sum_j \alpha_j D_{j,t} + \sum_j \gamma_j D_{j,t-1} + \frac{\theta_q(B)\Theta_Q(B^S)}{\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D}\varepsilon_t$$

CALENDAR VARIATION

STATIONER AUTOCORRELATED

TREND

SEASONAL

The Proposed ARIMAX Model

- The 2nd proposed Model → **ARIMAX-2**: **deterministic** TREND-SEASONAL

$$Y_t = \underbrace{\delta_1 \, t}_{\text{TREND}} + \underbrace{\beta_1 M_{1,t} + \beta_2 M_{2,t} + \cdots + \beta_s M_{s,t}}_{\text{SEASONAL}} + \underbrace{\sum_j \alpha_j D_{j,t} + \sum_j \gamma_j D_{j,t-1}}_{\text{CALENDAR VARIATION}} + \underbrace{\frac{\theta_q(B)}{\phi_p(B)} \varepsilon_t}_{\text{STATIONER Autocorrelated}}$$

The Proposed ARIMAX Model

- The 3rd proposed Model → **ARIMAX-3**: **deterministic** TREND-SEASONAL,and stochastic TREND-SEASONAL

$$Y_t = \overbrace{\delta_1 t}^{\textbf{TREND}} + \overbrace{\beta_1 M_{1,t} + \beta_2 M_{2,t} + \cdots + \beta_s M_{s,t}}^{\textbf{SEASONAL}} + \underbrace{\sum_j \alpha_j D_{j,t} + \sum_j \gamma_j D_{j,t-1}}_{\textbf{CALENDAR VARIATION}} + \underbrace{\frac{\theta_q(B)\Theta_Q(B^S)}{\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D}\varepsilon_t}_{\textbf{Stochastic Trend, Seasonal \& Stationer}}$$

The Proposed Procedure

Step 1: Determination of dummy variable for calendar variation period.

Step 2: Remove the calendar variation effect form the response by fitting

$$Y_t = \beta_0 + \sum_j \alpha_j D_{j,t} + \sum_j \gamma_j D_{j,t-1} + N_t$$

for model with stochastic trend and seasonal model, or fitting

$$Y_t = \delta_1\ t + \beta_1 M_{1,t} + \beta_2 M_{2,t} + \cdots + \beta_s M_{s,t} + \sum_j \alpha_j D_{j,t} + \sum_j \gamma_j D_{j,t-1} + N_t$$

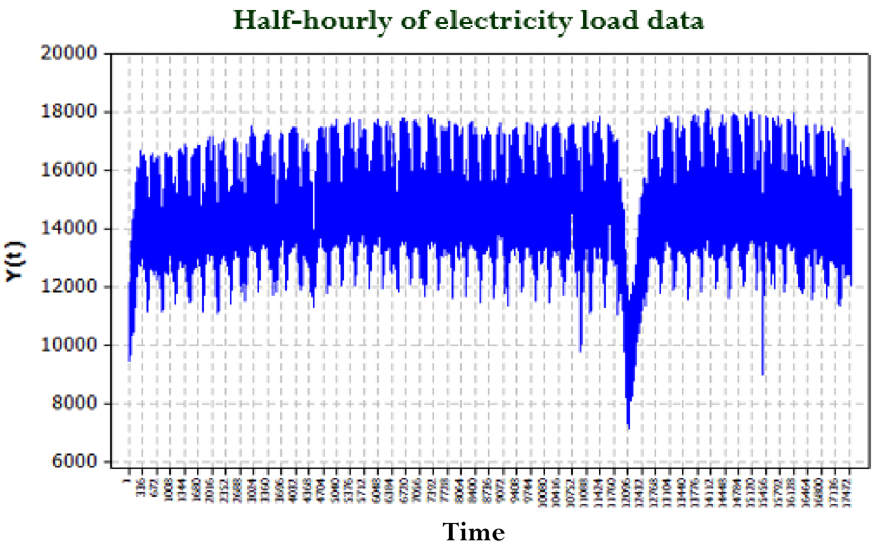
simultaneuously for model with deterministic trend and seasonal, to obtain the error, N_t.

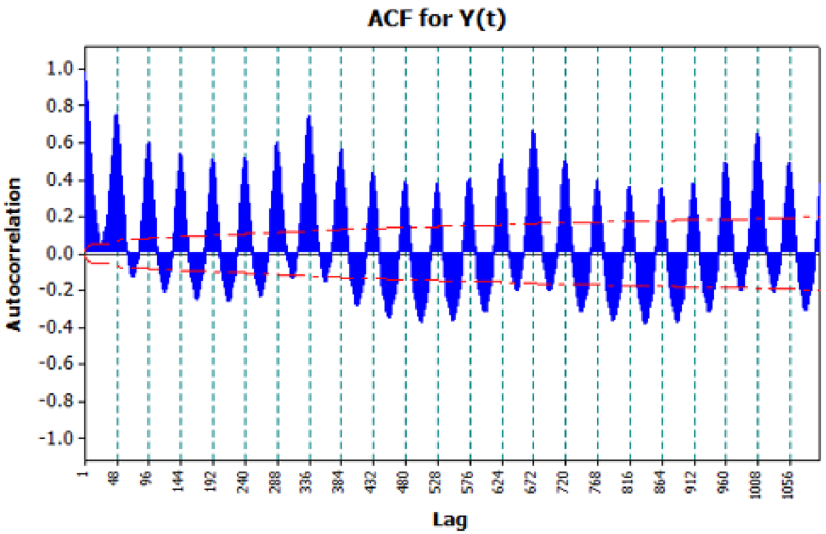
Step 3: Find the best ARIMA model of N_t using Box-Jenkins procedure.

Step 4: Simultaneously fit the model from step 2 and 3. This model is the calendar variation model based on ARIMAX method.

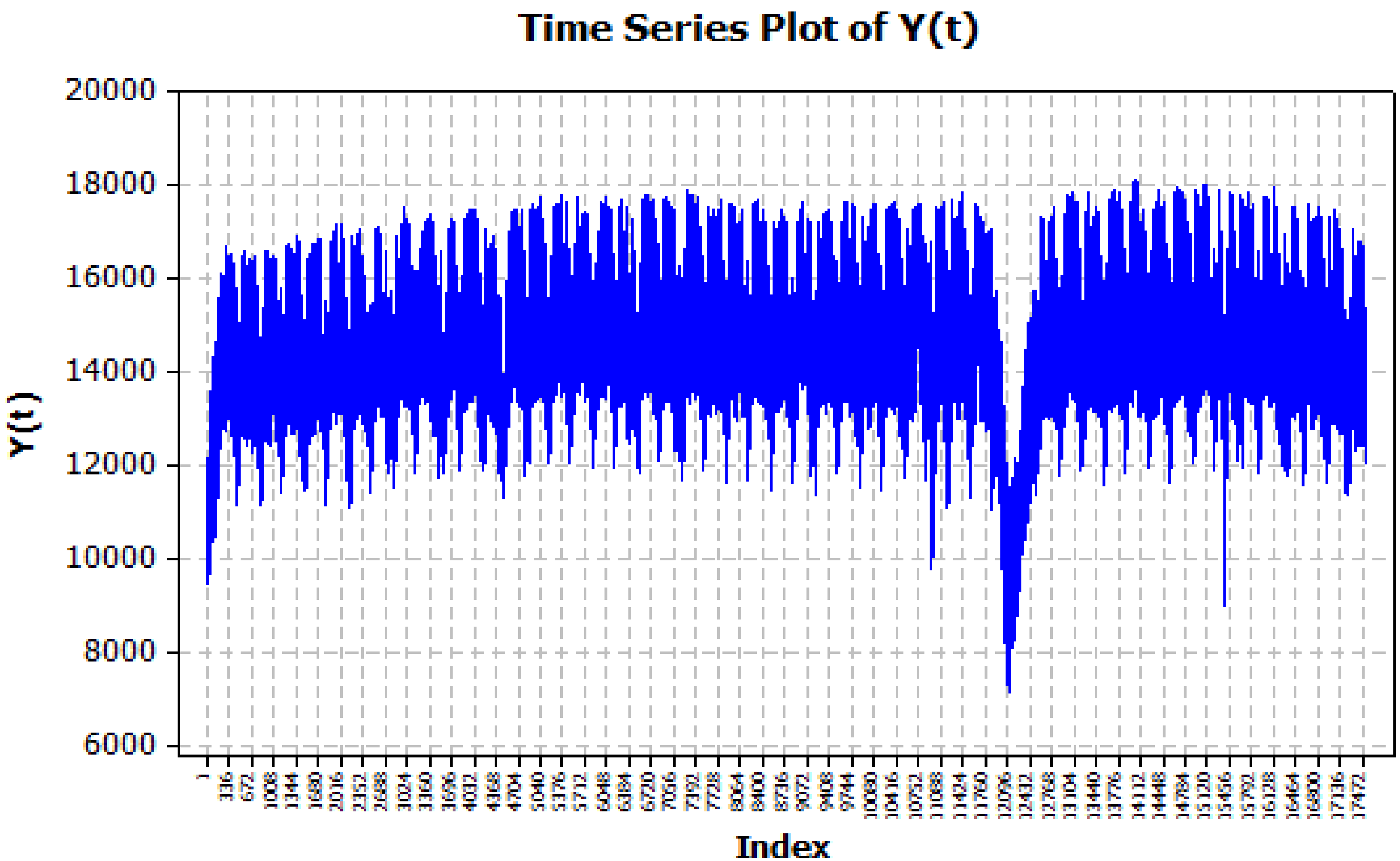
Step 5: Test the significance of parameter and perform diagnostic check.

ARIMA model
Multiple Seasonal: multiplicative – subset

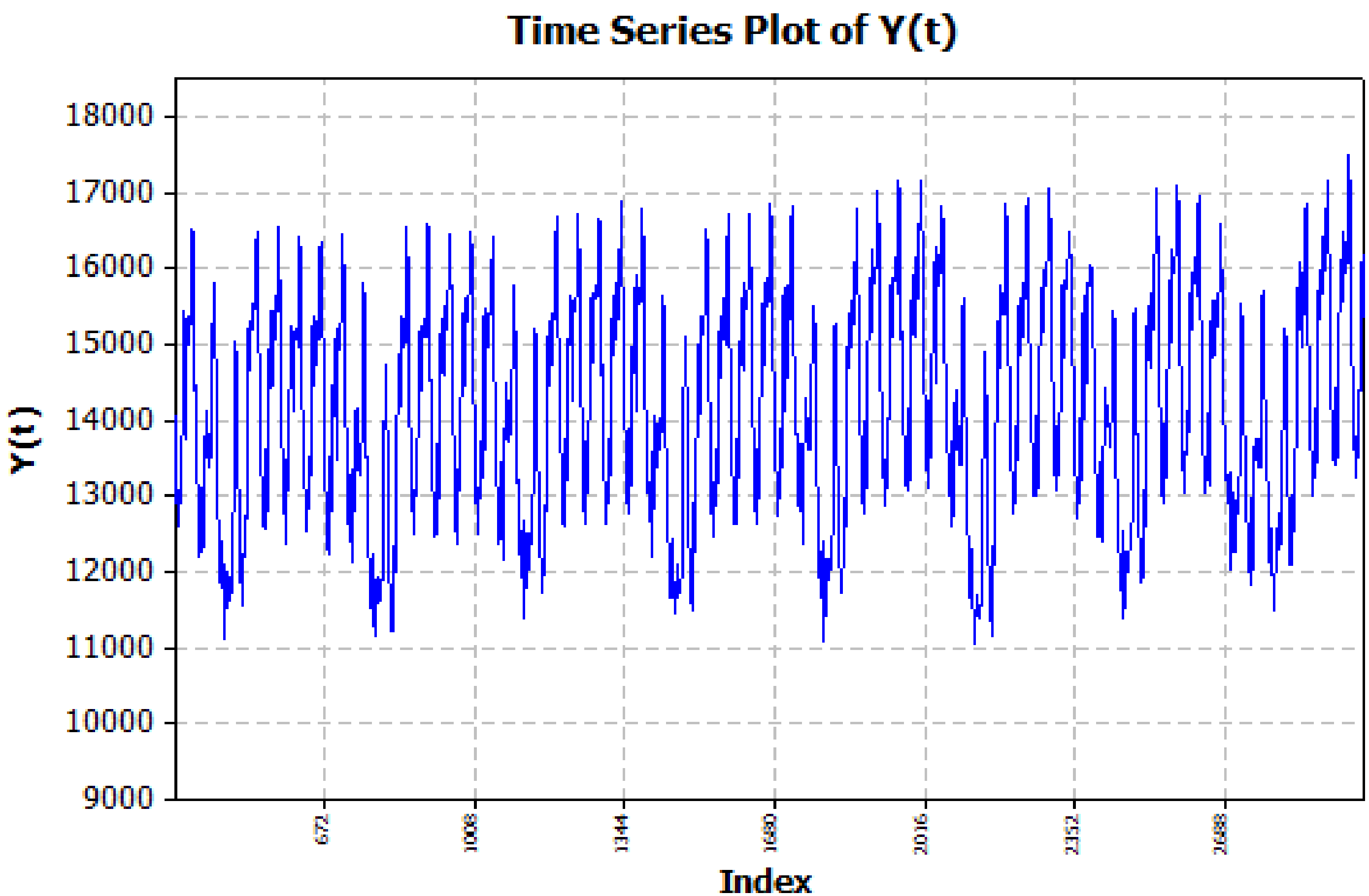




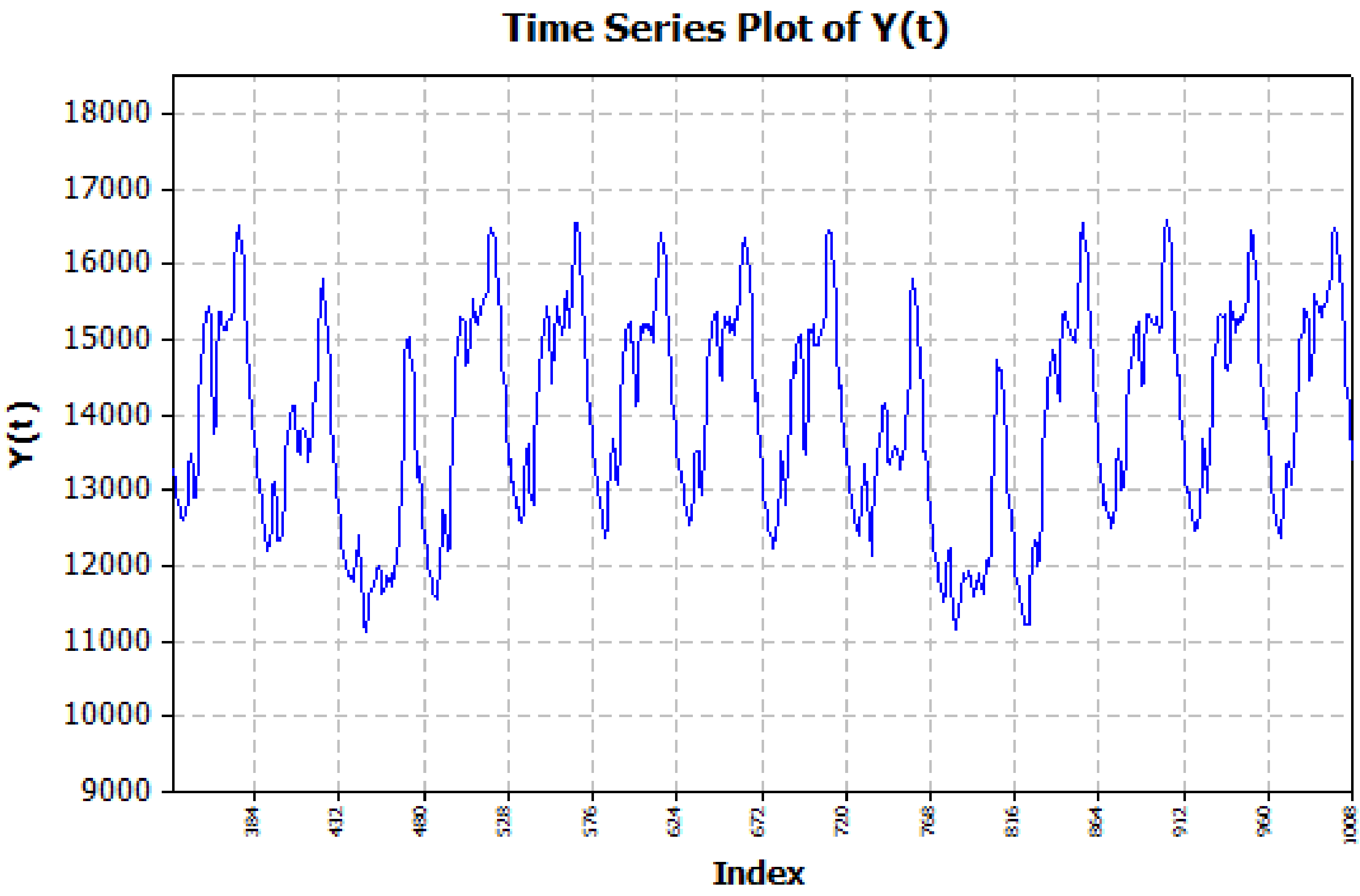
Problem: Prediction of half hourly load data



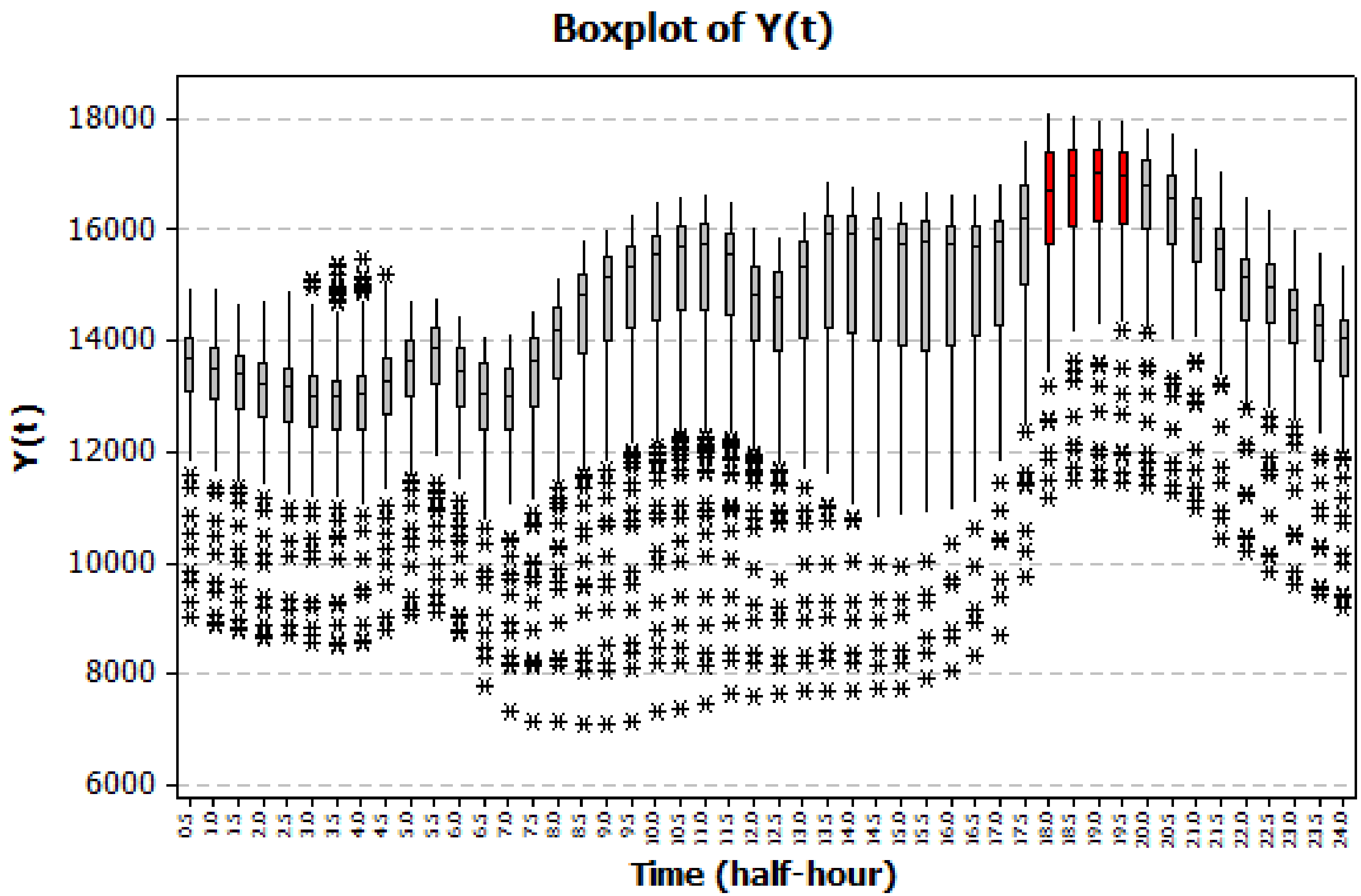
MINITAB Descriptive: half hourly load data



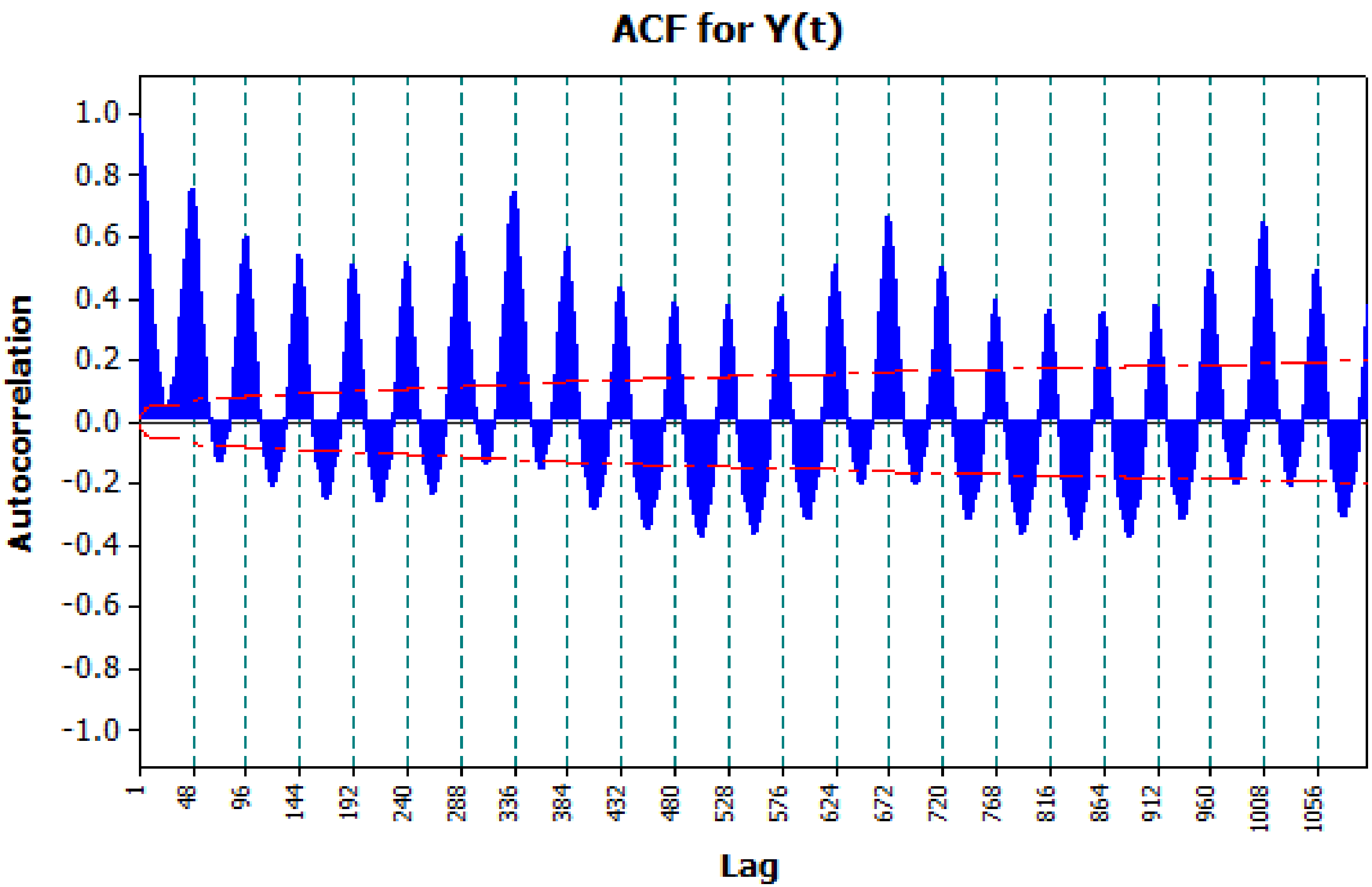
MINITAB Descriptive: half hourly load data



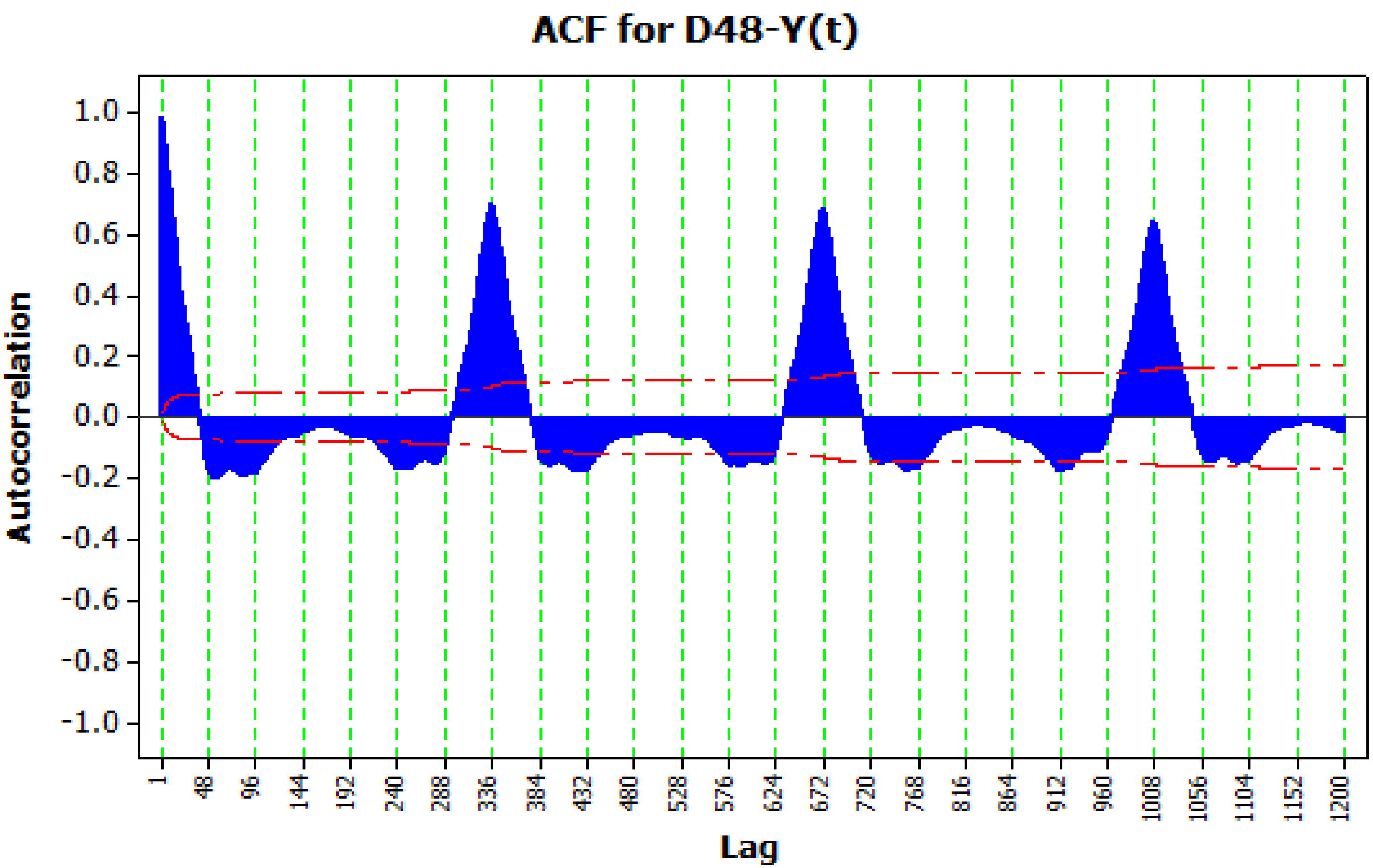
MINITAB Descriptive: half hourly load data



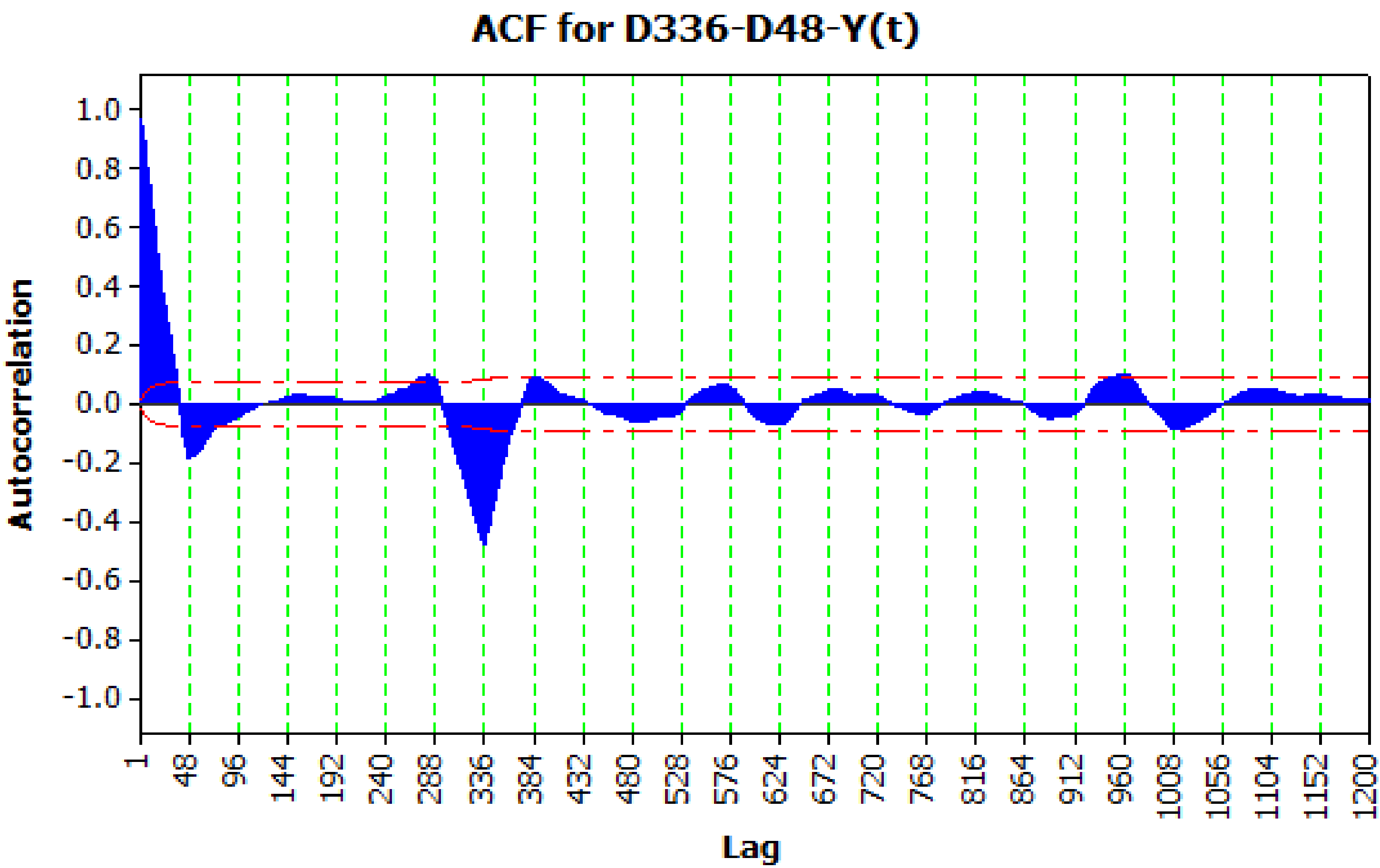
MINITAB Identification: half hourly load data



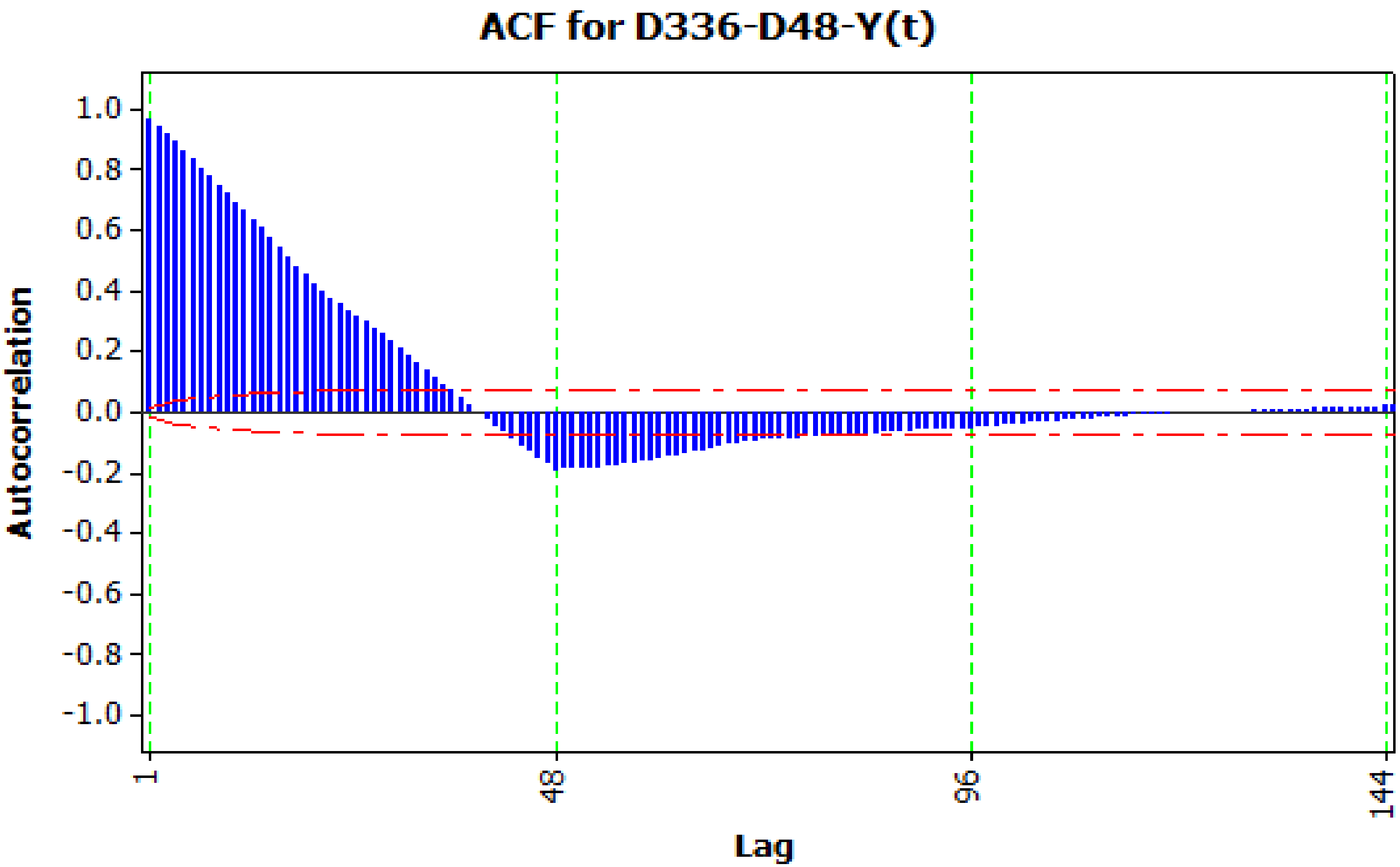
MINITAB Identification: half hourly load data



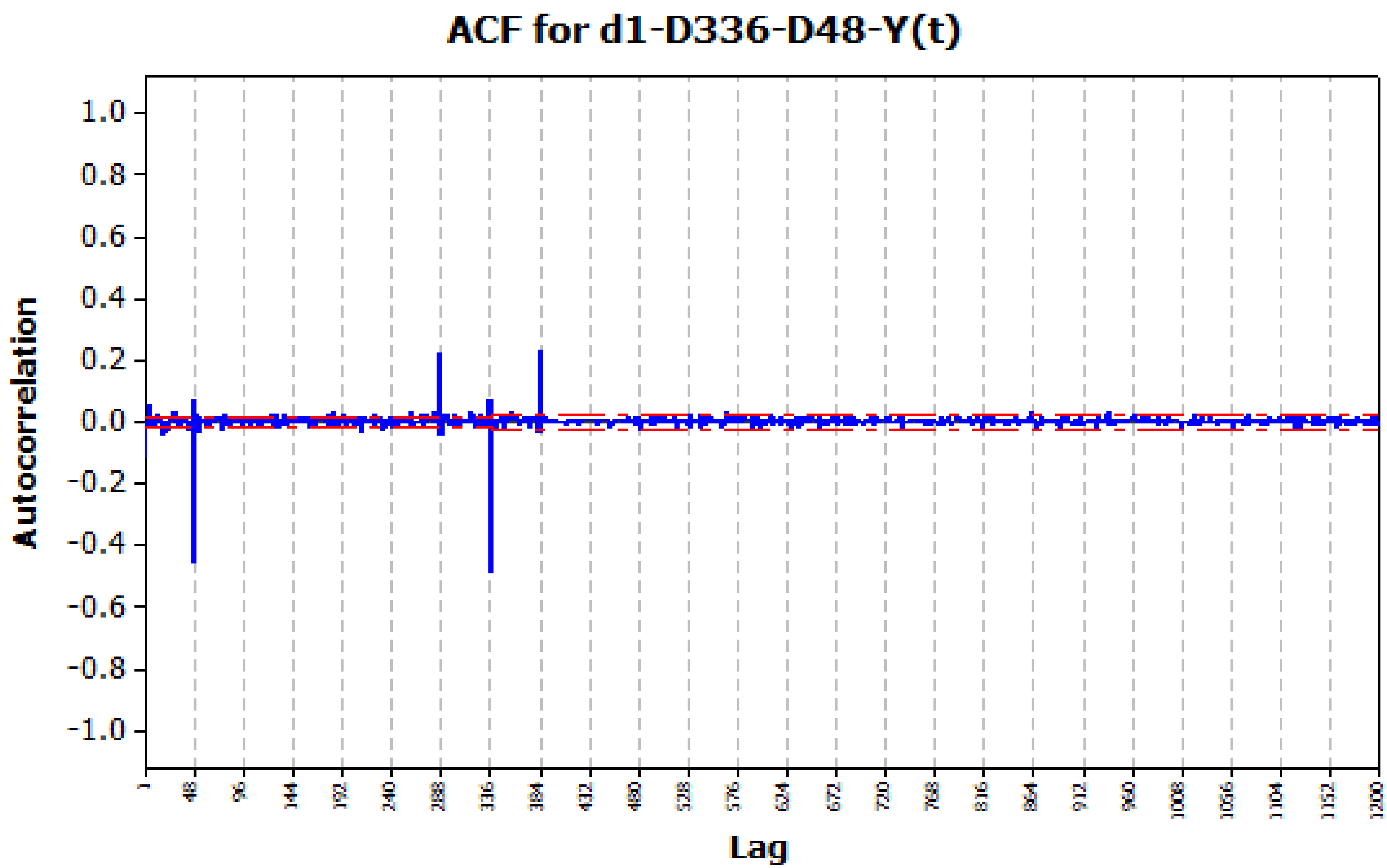
MINITAB Identification: half hourly load data



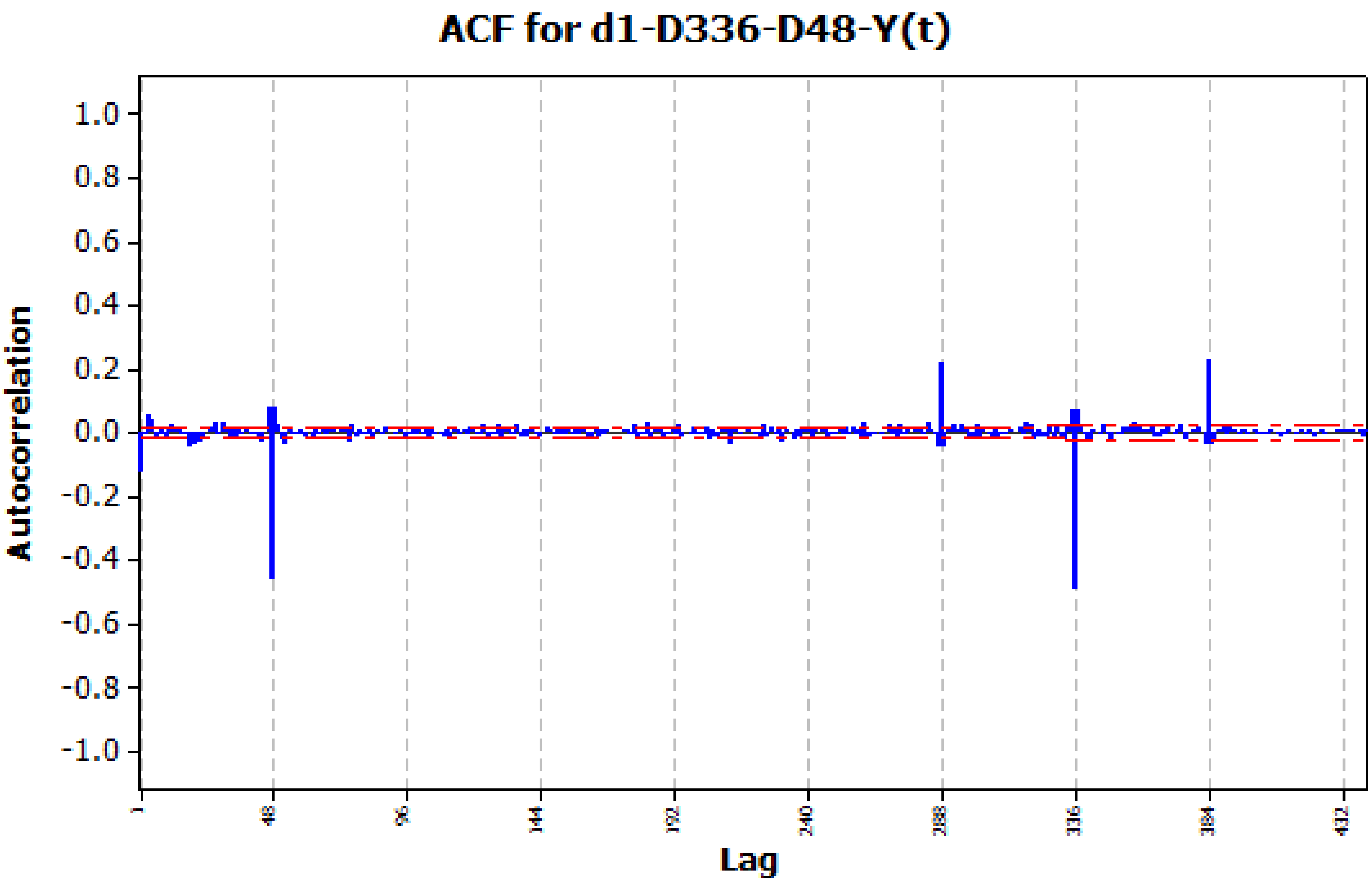
MINITAB Identification: half hourly load data



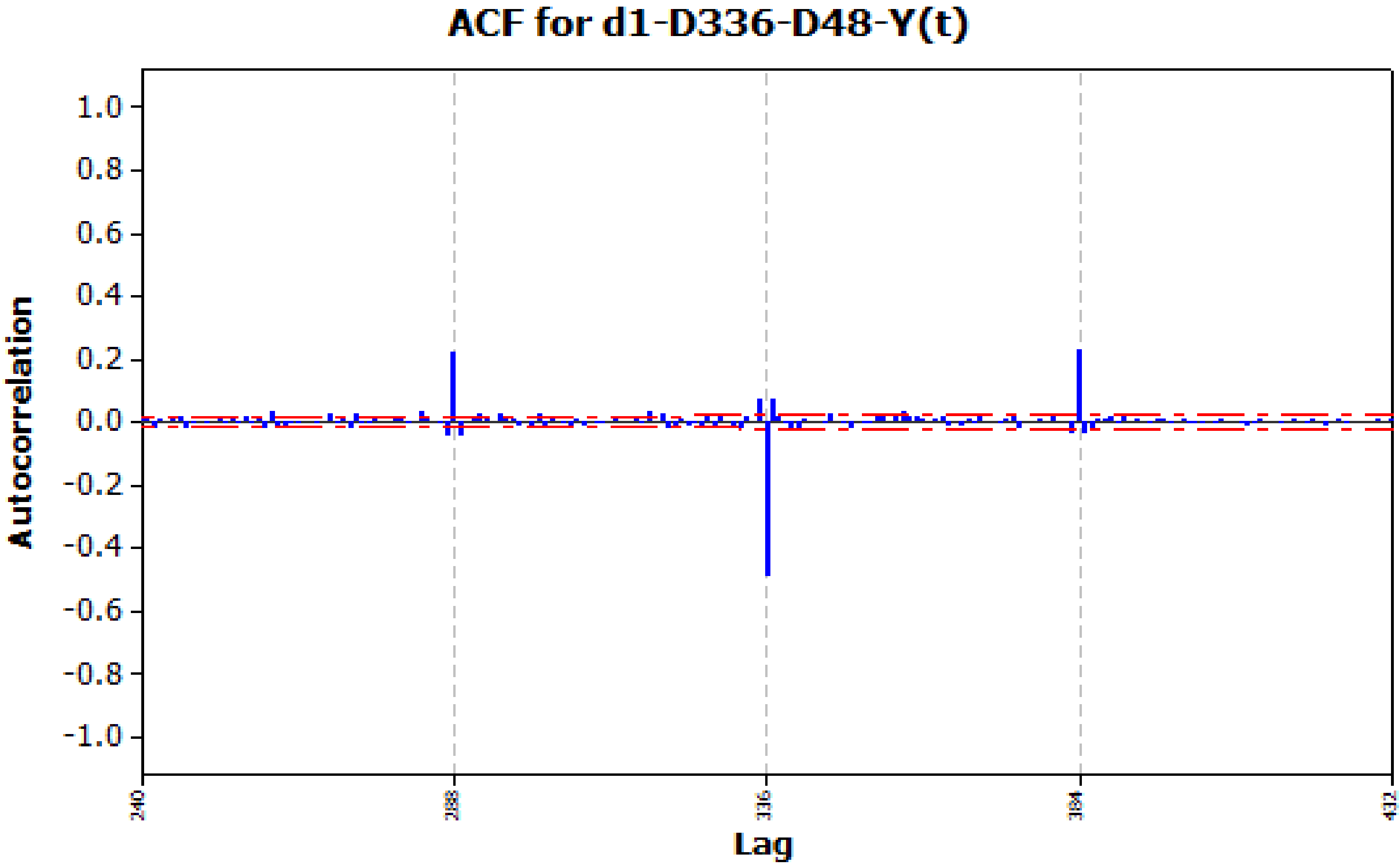
MINITAB Identification: half hourly load data



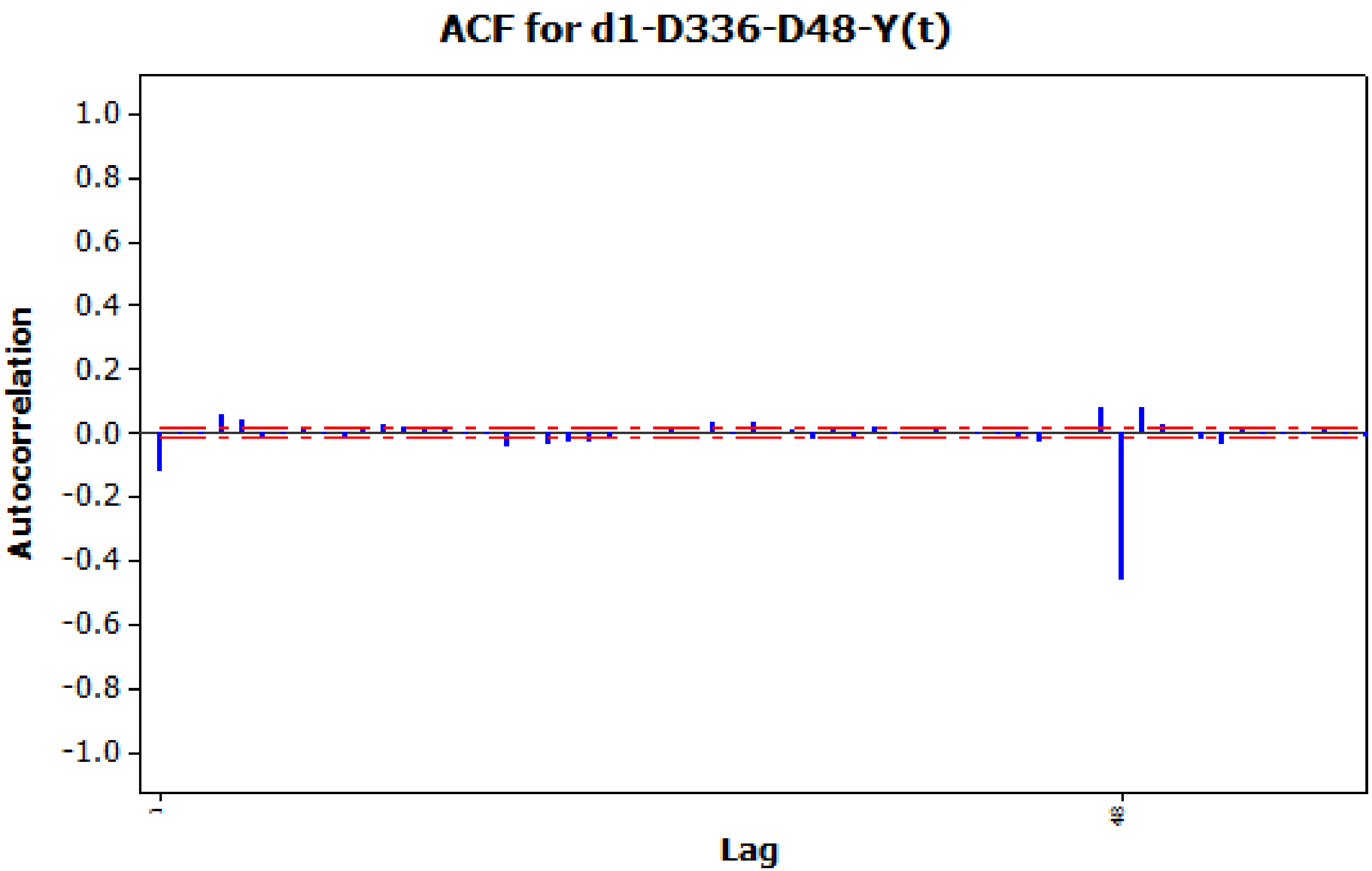
MINITAB Identification: half hourly load data



MINITAB Identification: half hourly load data



MINITAB Identification: half hourly load data



ARIMA, SARIMA, DSARIMA model

- ARIMA model

$$\phi_p(B)(1-B)^d Z_t = \theta_0 + \theta_q(B) a_t$$

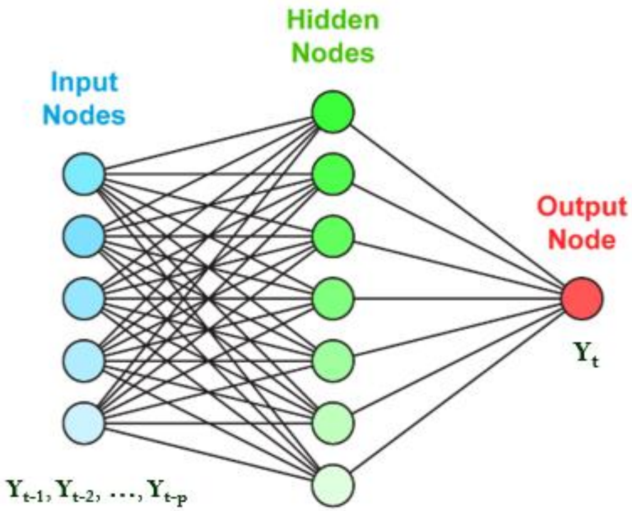
- SARIMA model

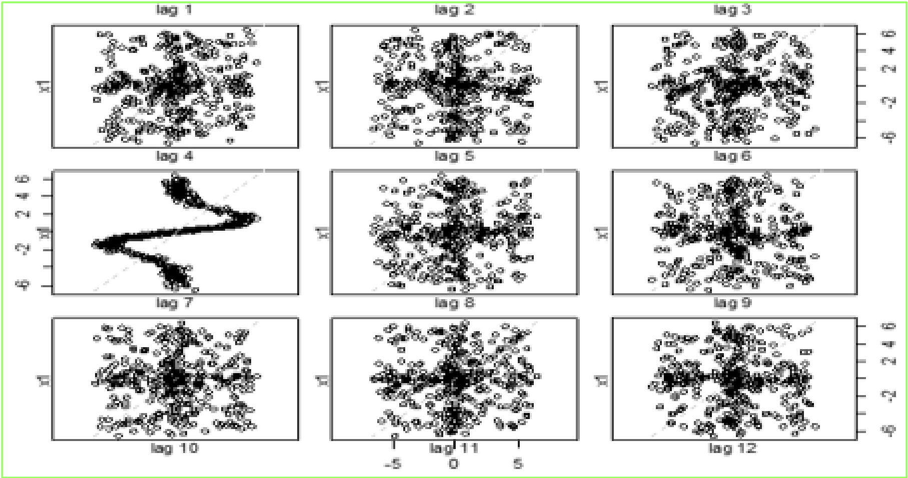
$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D \dot{Z}_t = \theta_q(B)\Theta_Q(B^s)a_t$$

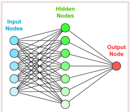
- DSARIMA model

$$\phi_p(B)\Phi_{P_1}(B^{s_1})\Phi_{P_2}(B^{s_2})(1-B)^d(1-B^{s_1})^{D_1}(1-B^{s_2})^{D_2} Z_t$$
$$= \theta_q(B)\Theta_{Q_1}(B^{s_1})\Theta_{Q_2}(B^{s_2})a_t$$

Neural Network for Trend, Seasonal & Calendar Variation

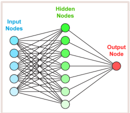




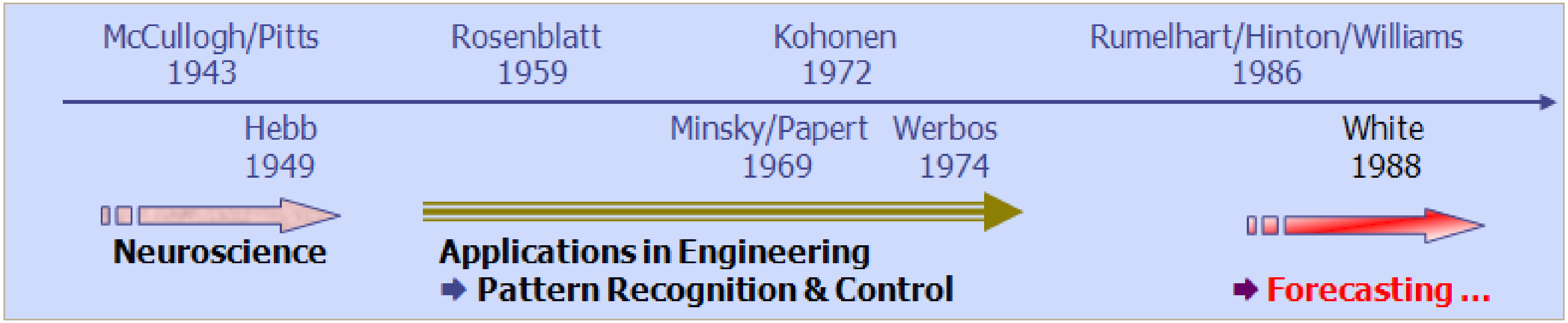


Outline

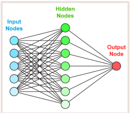
- **Introduction**: *Background*, *Motivation*, *Jargons*, *Goals*.
- **Architecture of Neural Networks:** ***Supervised* & *Unsupervised networks***
- **Model selection in Neural Networks**: ***Inputs*, *Number of hidden neurons*, *Activation function*, *Preprocessing method*.**
- **Application and Development:** ***Forecasting and Classification problems*.**



Neural Networks – NN



- **Sven F. Crone:**
  http://www.sven-crone.com/presentations.htm
  http://www.neural-forecasting.com/
- **Halbert L. White:**
  http://weber.ucsd.edu/~hwhite/
- **Warren S. Sarle:**
  ftp://ftp.sas.com/pub/neural/FAQ.html

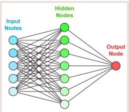


General Background

- During the last few decades,
  - ❶ modeling to explain nonlinear relationship between variables, and
  - ❷ some procedures to detect this nonlinear relationship

  have grown in a spectacular way and received a great deal of attention.
  - Granger, C.W.J. and Terasvirta,T., (1993)
  - Terasvirta, T., Tjostheim, D. and Granger, C.W.J., (1994)

- Due to computational advances and increased computational power, nonparametric models that do not make assumptions about the parametric form of the functional relationship between the variables to be modelled have become more easily applicable.



Motivation of NN Research

- Today's research is largely motivated by the possibility of using NN model as an instrument to solve a wide variety of application problems such as:

  pattern recognition (classification), signal processing, process control, and forecasting.

- The use of the NN model in applied work is generally motivated by a mathematical result stating that under mild regularity conditions, a relatively simple NN model is capable of approximating any Borel-measureable function to any given degree of accuracy.

  (see e.g. Hornik, Stichombe and White (1989, 1990), White (1990); Cybenko (1989))

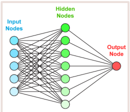
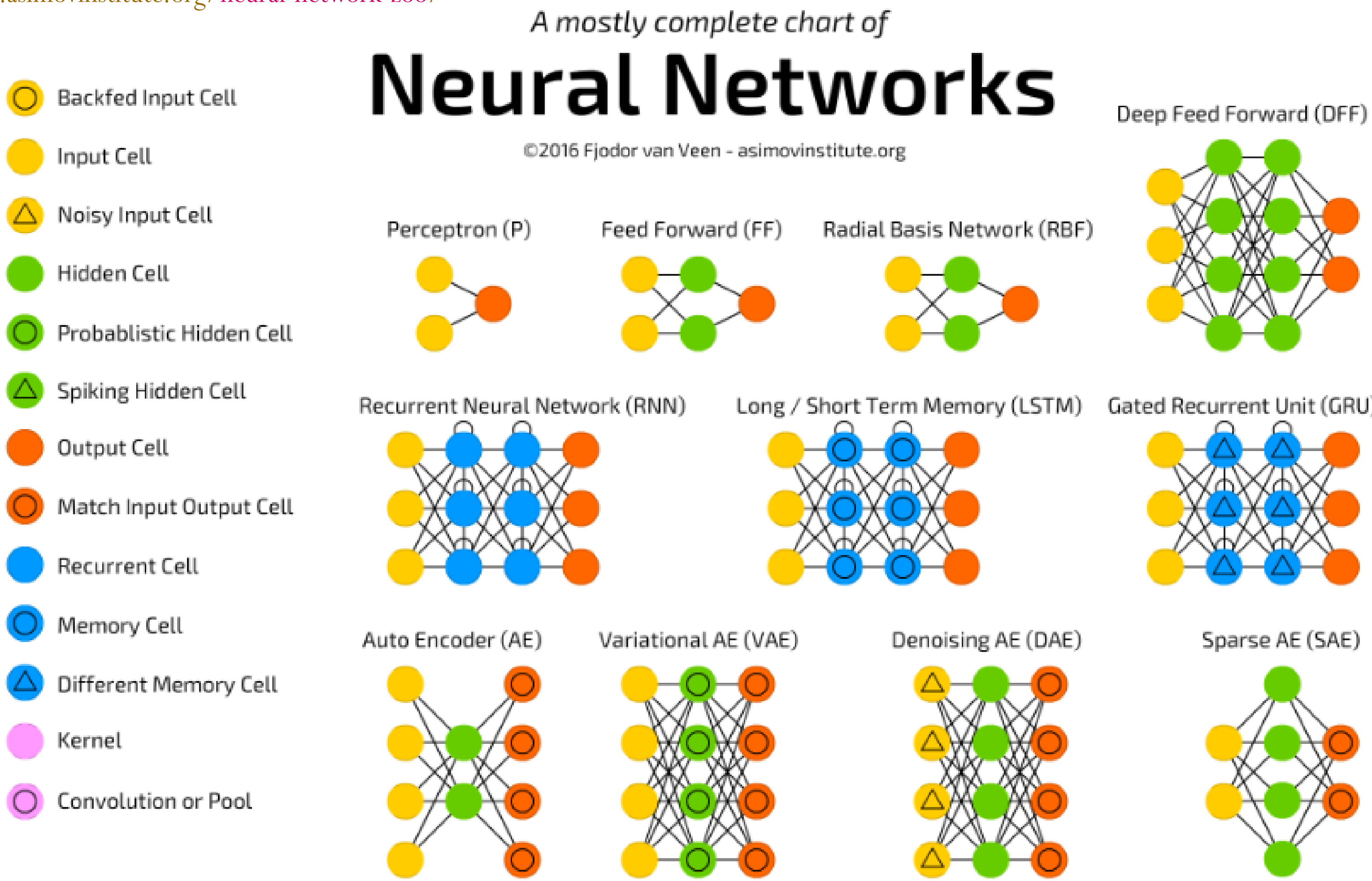


Chart of Neural Networks

http://www.asimovinstitute.org/neural-network-zoo/



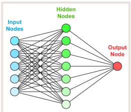
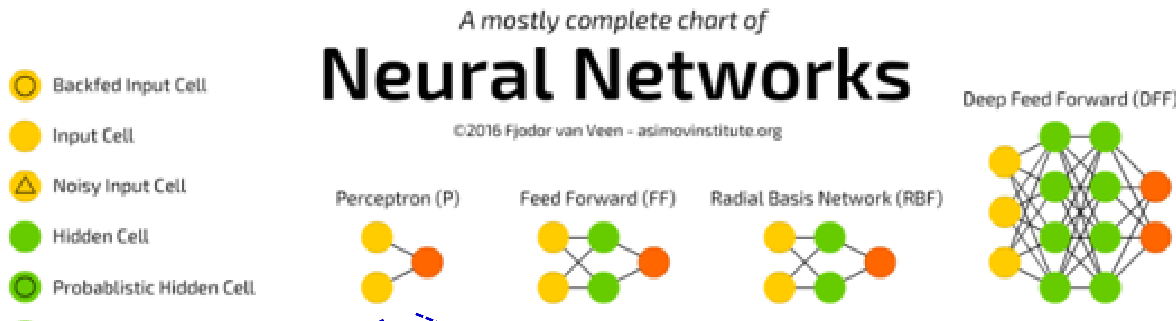


Chart of Neural Networks

http://www.asimovinstitute.org/neural-network-zoo/



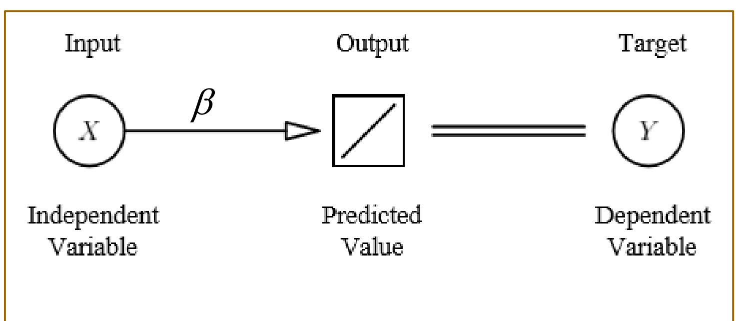


Figure 1: Simple Linear Regression

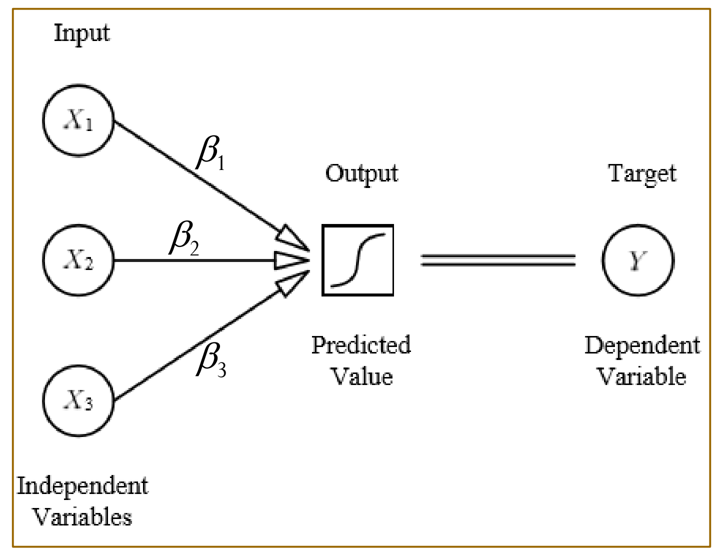
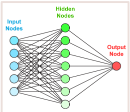


Figure 2: Simple Nonlinear Perceptron = Logistic Regression

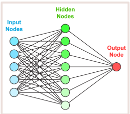
Source: Sarle (1994)



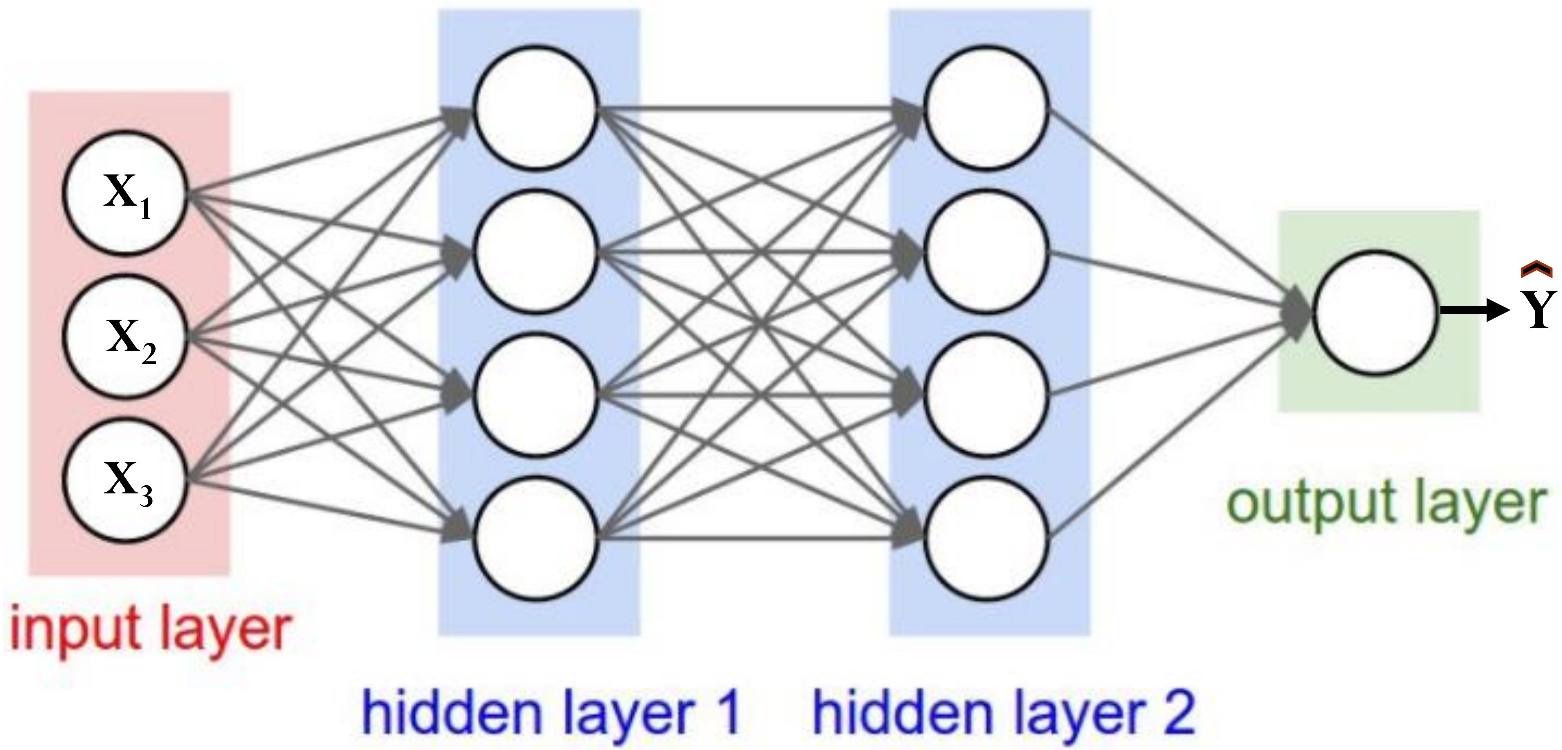
Feed Forward Neural Networks

- **Multilayer perceptron** (**MLP**), also known as **feedforward neural networks** (**FFNN**), is probably the **most commonly** used NN architecture in engineering application.

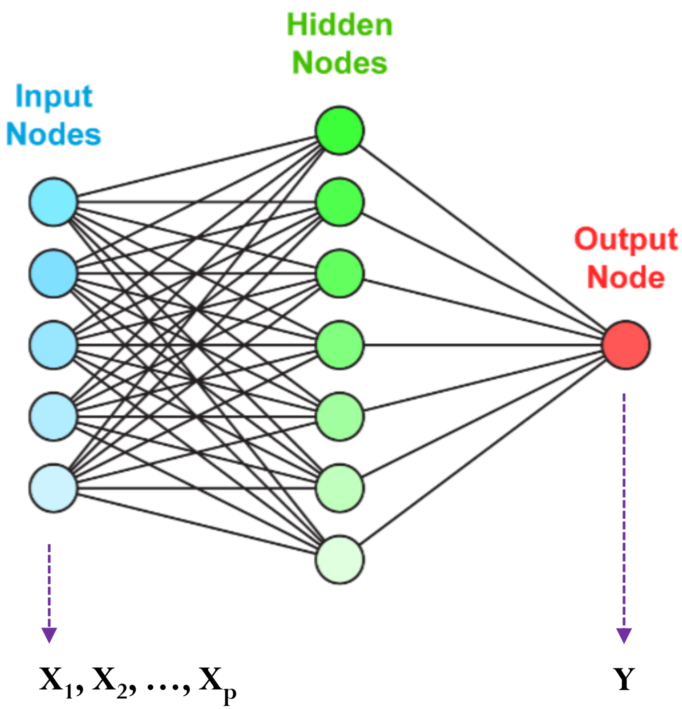
- Typically, applications of NN for regression, time series modeling and classification (discriminant analysis) are **based on the FFNN architecture**.



Deep Feedforward Network

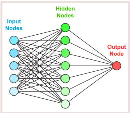


FFNN as Nonlinear regression

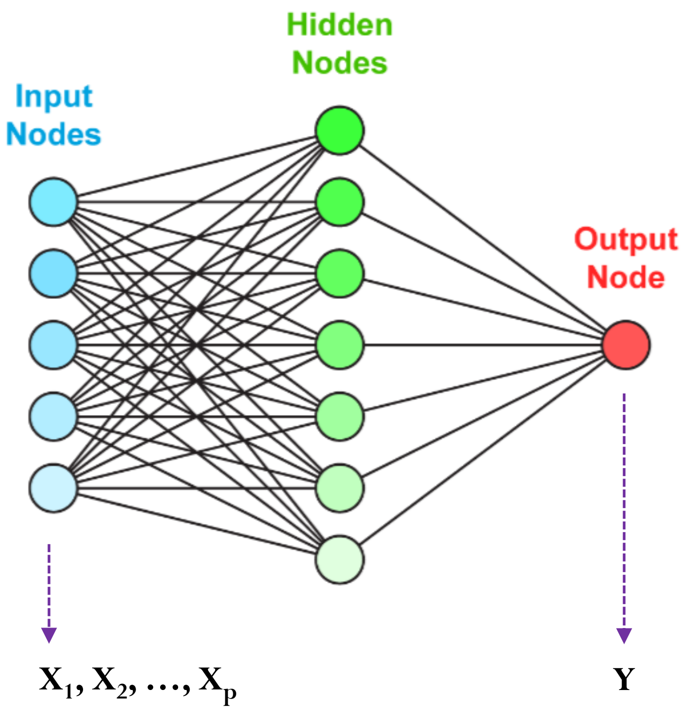


- FFNN includes estimated weights between the inputs and the hidden layer, and the hidden layer uses nonlinear activation functions such as the logistic function, the FFNN becomes genuinely nonlinear model, i.e., nonlinear in the parameters.

➲ In this case, FFNN can be seen as nonlinear regression. FFNN can have multiple inputs and outputs (This figure is multiple inputs with single output), and this architecture is similar to **multiple nonlinear regression**.



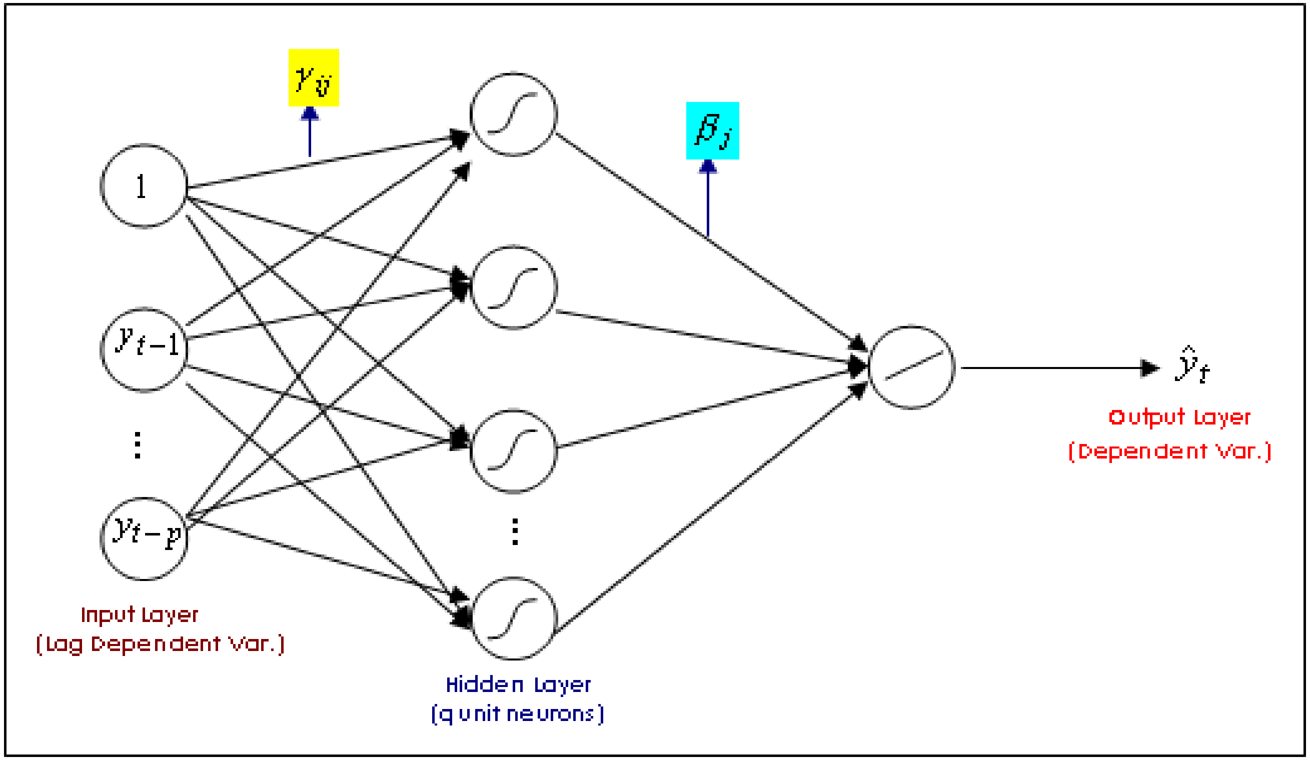
FFNN as Logistic Regression and Discriminant Analysis



- FFNN with nonmetric data (dichotomous/polycothomus) in target values is identical to **logistic regression** and **nonlinear discriminant analysis**.

➲ In this case, FFNN often use a multiple logistic function to estimate the conditional probabilities of each class. A multiple logistic function is called a ***softmax*** activation function in the NN literature.

FFNN as Nonlinear AR(p) model

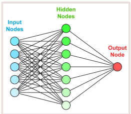


$$y_t = \beta_0 + \sum_{j=1}^{q} \beta_j f\left(\sum_{i=1}^{p} \gamma_{ij} y_{t-i} + \gamma_{oj} \right) + \varepsilon_t$$

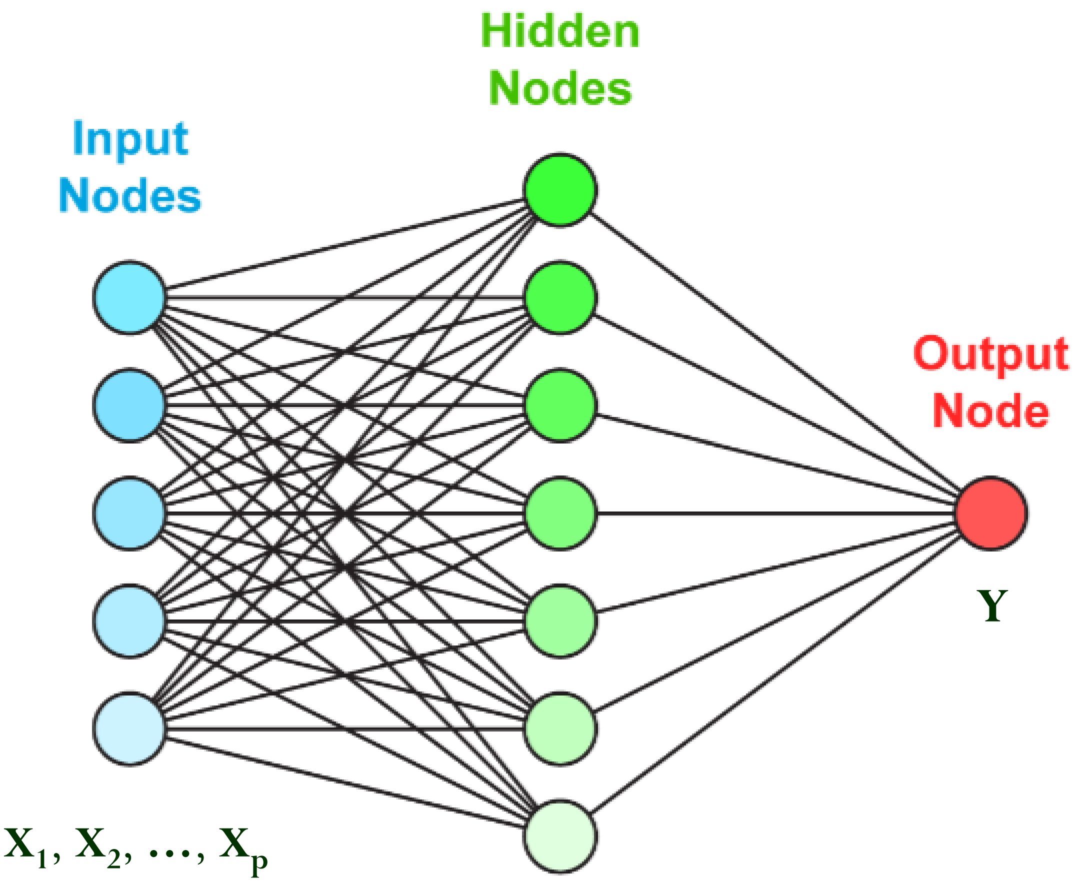
FFNN as Nonlinear AR(p) model

- Model building strategy that proposed by Terasvirta *et al.* (1994)

    1. Test Y_t for linearity, using linearity test (neglected nonlinearity).
    2. If linearity is **<u>rejected</u>**, consider a small number of alternative parametric models and/or nonparametric models.
    3. These models should be estimated in-sample and compared out-of-sample.



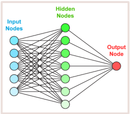
FFNN: the main problems !!!



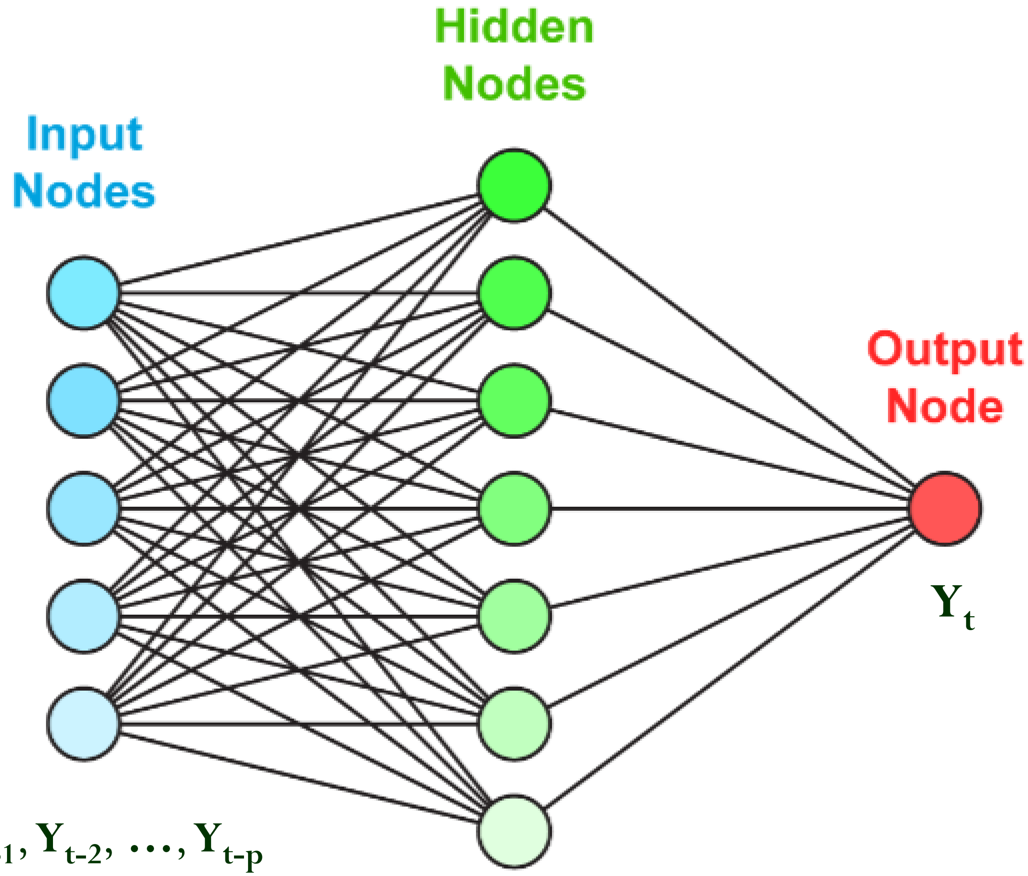
In Classification & Regression

1. How many **nodes** (**neurons**) in hidden layer?
2. What is the best **inputs** (**features selection**)?
3. What is the best **pre-processing** method?
4. What is the best **activation function** in **hidden** and **output** layer?

Model selection in Neural Networks



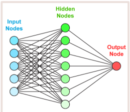
FFNN: the main problems !!!



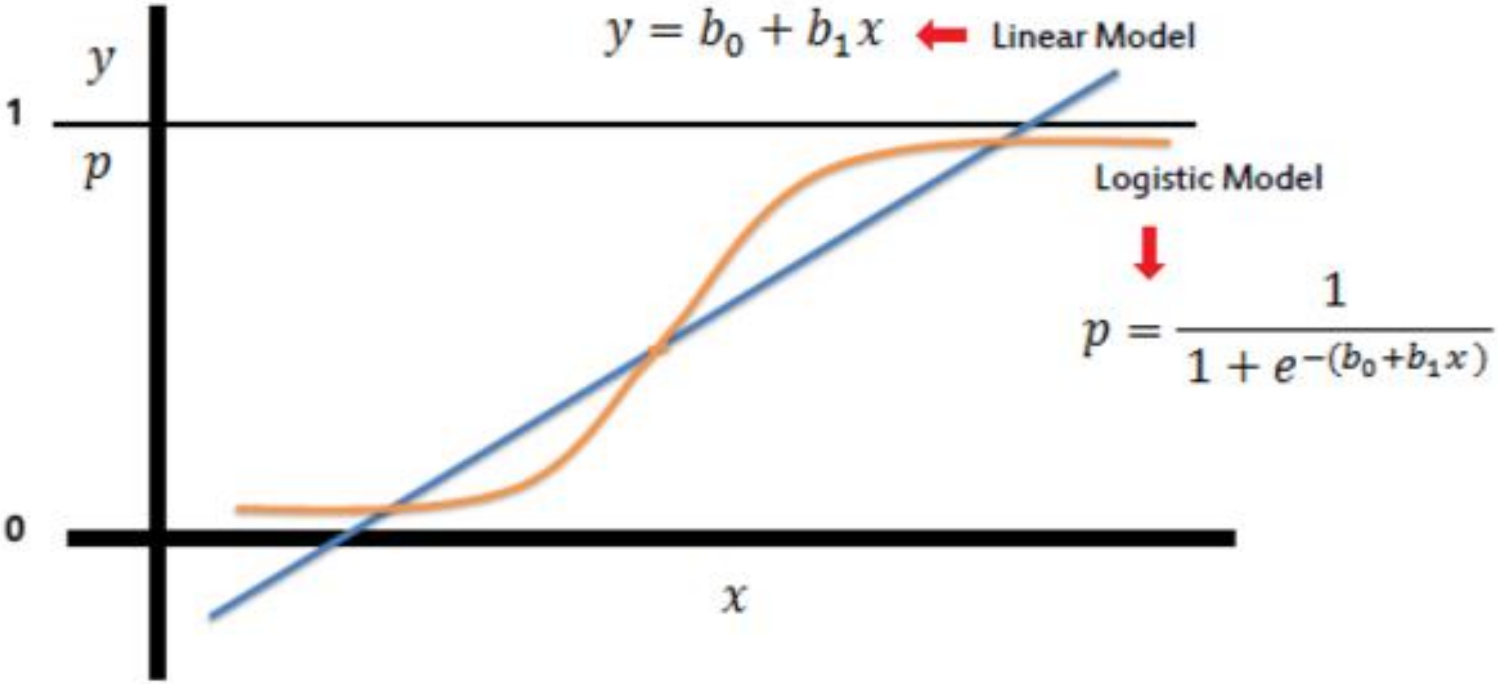
In **Time Series Forecasting**

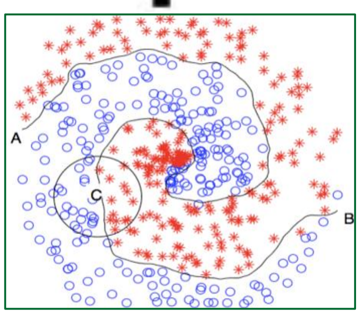
1. What is the best **inputs** (**features selection**)?
2. How many **nodes** (**neurons**) in hidden layer?
3. What is the best **pre-processing** method?
4. What is the best **activation function** in **hidden** and **output** layer?

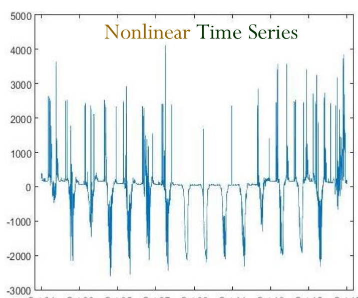
Model selection in Neural Networks

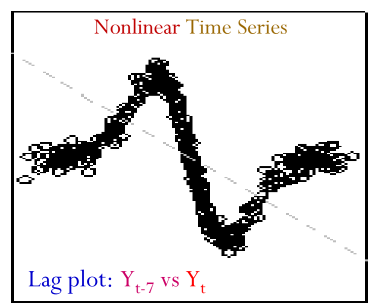


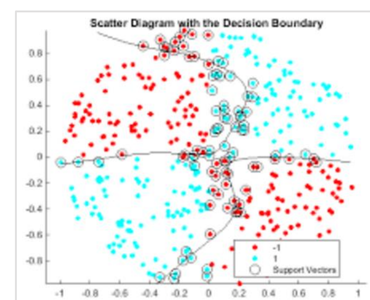
Nonlinear relationship Concept

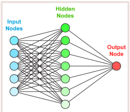












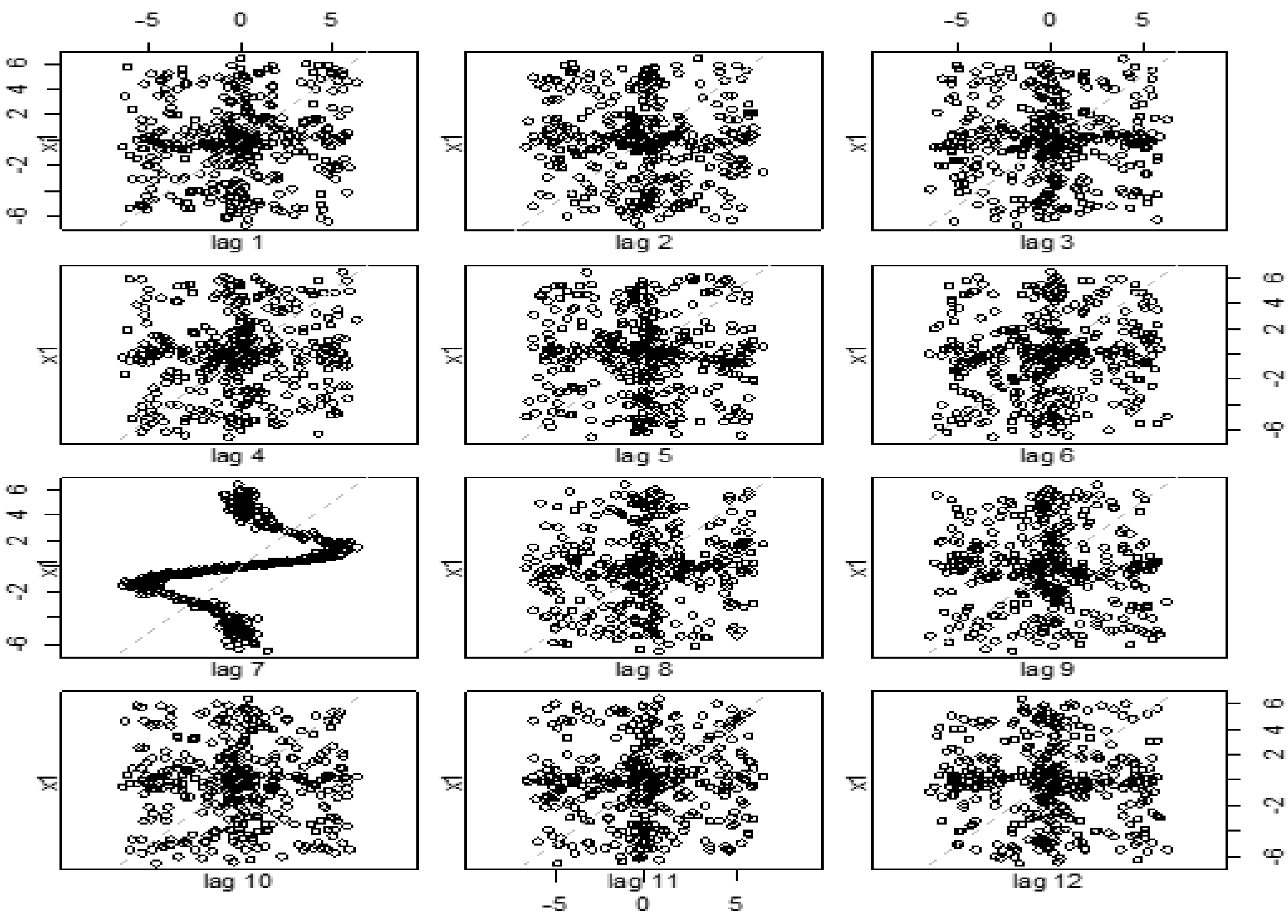
Model Selection in Neural Network

- In general, there are **two procedures** usually used to find **the best FFNN model** or **the optimal architecture**, those are “general-to-specific” or “top-down” and “specific-to-general” or “bottom-up” procedures.

➲ “Top-down” procedure is started from complex model and then applies an algorithm to reduce number of parameters (number of input variables and unit nodes in hidden layer) by using some stopping criteria, whereas “bottom-up” procedure works from a simple model.

Application: NN for Time Series Forecasting

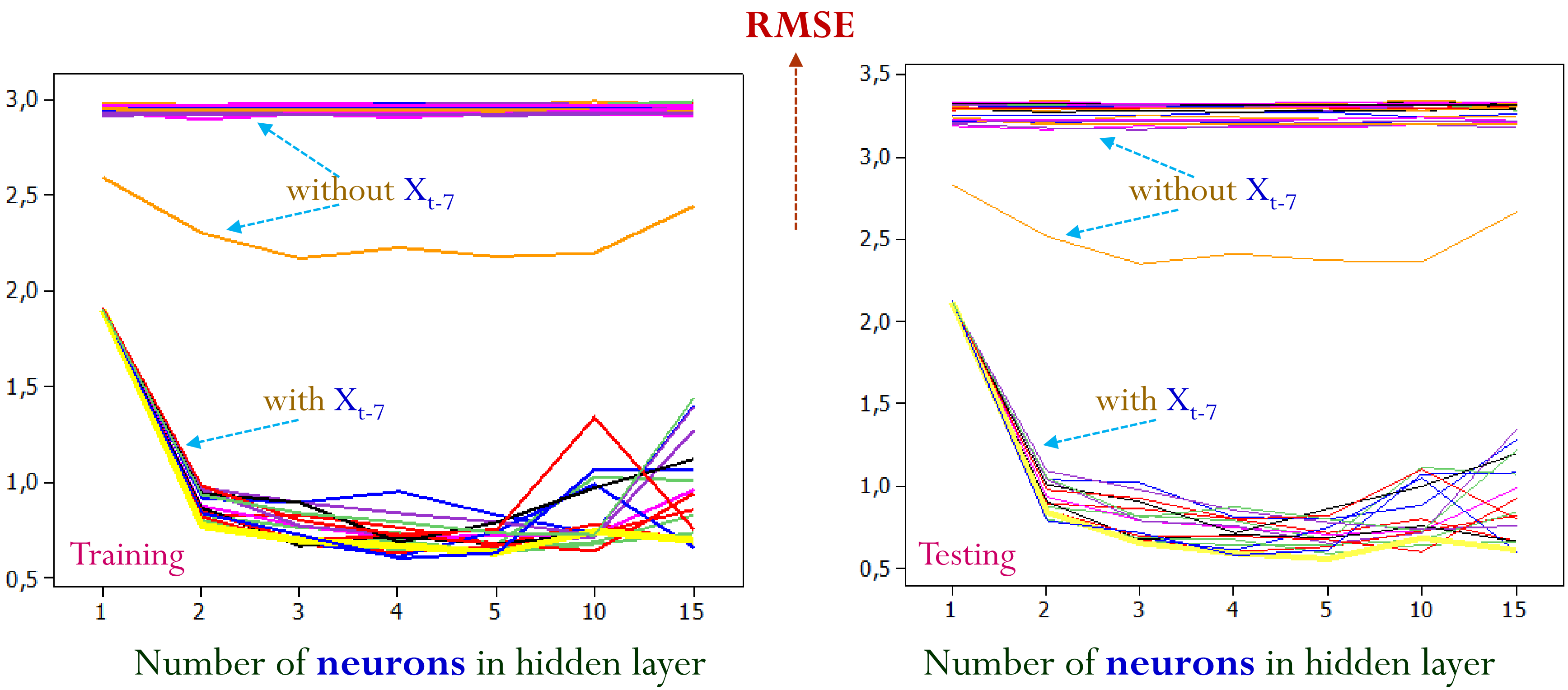
- <u>Source</u>: **simulation study** using ESTAR(1)[7] model



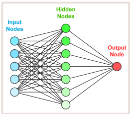
➲ Identification the appropriate lag inputs: use **LAG PLOT** in ***R***

Application: NN for Time Series Forecasting

- Source: **simulation study** using ESTAR(1)^7 model



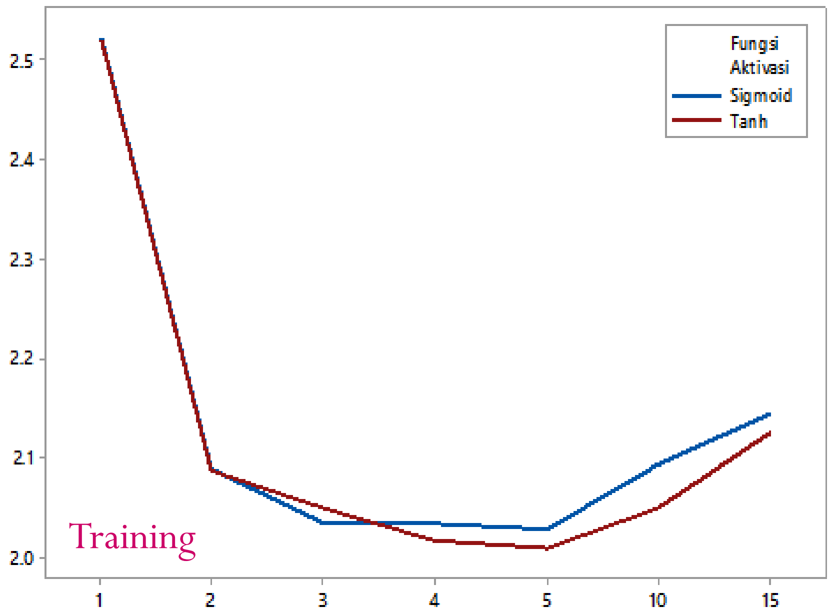
➲ The effect of INPUTS and number of NEURONS in hidden layer



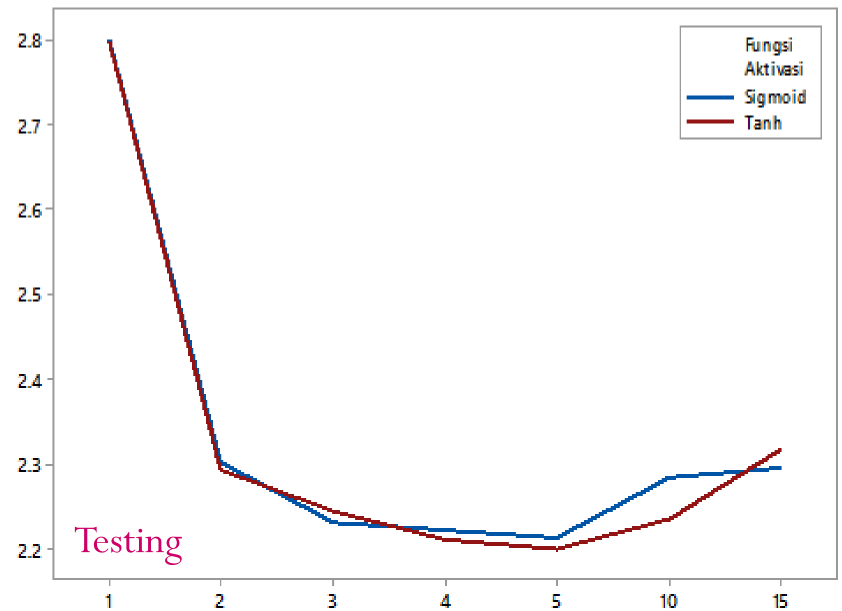
Application: NN for Time Series Forecasting

- <u>Source</u>: **simulation study** using ESTAR(1)[7] model

RMSE

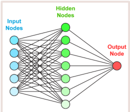


Number of **neurons** in hidden layer



Number of **neurons** in hidden layer

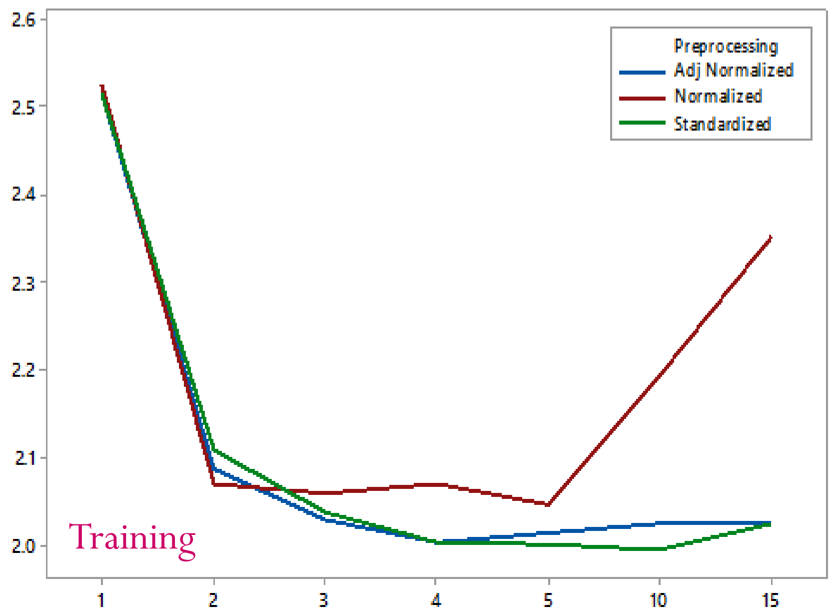
➲ The effect of ACTIVATION function and number of NEURONS



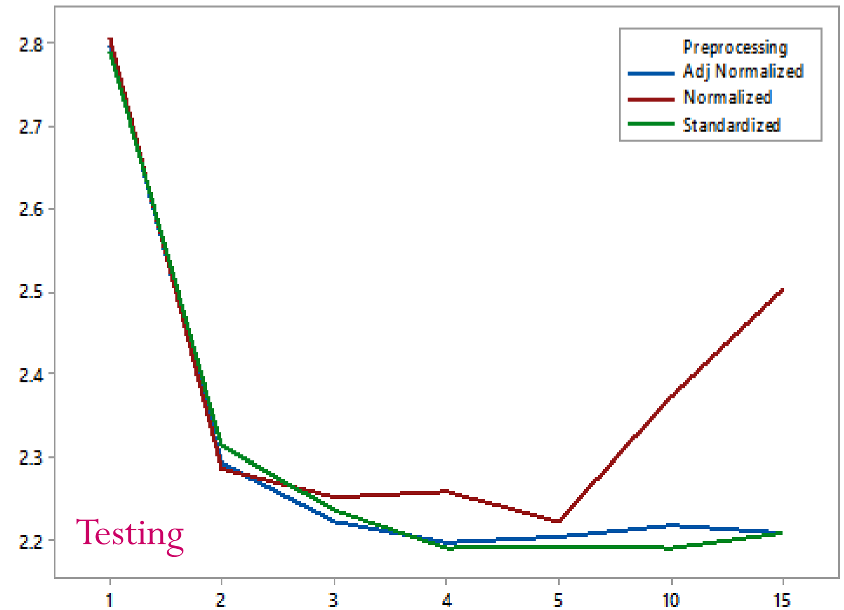
Application: NN for Time Series Forecasting

- Source: **simulation study** using ESTAR(1)[7] model

RMSE

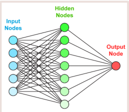


Number of **neurons** in hidden layer



Number of **neurons** in hidden layer

➲ The effect of PREPROCESSING method & number of NEURONS



Application: NN for Time Series Forecasting

- Source: **simulation study** using ESTAR(1)[7] model

☞ ***Summary*** of the results:

(1) More sophisticated or complex methods do not necessarily provide more accurate forecast than simpler ones.

(2) The performance of the various NN methods for time series forecasting problem depends upon :

Inputs or **lag variables**,
Number of **neurons** in hidden layer,
Pre-processing method.