ECONOMETRIC MODELING VS ARTIFICIAL NEURAL NETWORKS

– A SALES FORECASTING COMPARISON

Master's (one year) thesis in Informatics (15 credits)

Dinesh Bajracharya

Autumn 2010:MI17

UNIVERSITY OF BORÅS
SCHOOL OF BUSINESS AND INFORMATICS

Title: Econometric Modeling vs Artificial Neural Networks – A Sales Forecasting Comparison

Year: 2010

Author: Dinesh Bajracharya

Supervisor: Rikard König

Abstract

Econometric and predictive modeling techniques are two popular forecasting techniques. Both of these techniques have their own advantages and disadvantages. In this thesis some econometric models are considered and compared to predictive models using sales data for five products from ICA a Swedish retail wholesaler. The econometric models considered are regression model, exponential smoothing, and ARIMA model. The predictive models considered are artificial neural network (ANN) and ensemble of neural networks. Evaluation metrics used for the comparison are: MAPE, WMAPE, MAE, RMSE, and linear correlation. The result of this thesis shows that artificial neural network is more accurate in forecasting sales of product. But it does not differ too much from linear regression in terms of accuracy. Therefore the linear regression model which has the advantage of being comprehensible can be used as an alternative to artificial neural network. The results also show that the use of several metrics contribute in evaluating models for forecasting sales.

Keywords: Econometrics, Forecasting, ARIMA, Exponential Smoothing, Regression, Neural Network, Ensemble, WMAPE, MAPE, MAE, RMSE, Linear correlation

Acknowledgements

Thanks to my supervisor Rikard König for providing perfect guidance and insights for carrying out this thesis successfully. Without his comments and invaluable suggestions, I would not have able to complete this thesis in time with perfection.

Special Thanks to Anders Hjalmarsson for providing valuable comments during the thesis progress seminars and to my mother Nanu, father Govinda, brother Er. Deepak and his family, my wife Dr. Manju, uncles Prof. Dr. Pushkar and Er. Ishwor and my friend-cum-brother Sumanji for their continuous support and inspiration.

Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

Most organizations: government or non governments, small or big need good forecasting models. Normally they have to forecast several variables like rate of inflation, salary of employee, sales of product etc. Good forecasting processes facilitate organization to estimate the future and optimize their actions accordingly.

Forecasting helps retail companies achieve a competitive advantage by integrating customer-focused sales and marketing plans for both new and existing products in the supply chain (STL Warehousing, 2010).

A forecasting is based on a forecasting model that models the relationship between different variables. Forecasting models can also show cause-and-effect relationship between variables. Several forecasting models are used today; two common types are Econometric and Predictive models (The Decision Makers' Direct, 2001).

According to the Cambridge Advanced Learner's Dictionary a *technique* is "a way of doing an activity which needs skill." So, modeling technique can be considered as a way of training models which needs skill. Similarly predictive or econometric modeling technique can be defined as the way of training models. In this study comparison between predictive modeling technique and econometric modeling techniques is done by considering econometric models and predictive models.

Verbeek M. (2000) stated that the major task of econometrics is to quantify relationship between different variables using available data, interpret results using statistical techniques and apply the results appropriately. For example, it is known fact that sales of product or simply sales are affected by advertising. This can be modeled with model:

$$sales = a + b * advertise + u \qquad (1.1)$$

here, a and b are parameters and u is random factors affecting sales other than advertising. Econometrics will quantify values of parameters a and b using data and helps to see the effect advertise on the basis of parameters.

The difference between econometric- and predictive techniques is that predictive models are not based on any economic theory. Instead the predictive models are general, assumption free and can often be adapted to any function.

Econometric models are based and developed using the economic theories. (Hirchey M., 2009). The economic theories provide insight to understand the cause-and-effect relationships between several variables involved in the econometric models. This helps to make a forecast logically consistent and reliable. On the other hand predictive modeling uses past data, builds structure of the model and generates forecast value as an output. In general predictive modeling searches for the best model for a certain problem and decides the structure, functions and parameters of the model. The Predictive models are used to predict future values of variables using known results obtained from different historical data (Dunham, 2003).

Both econometric modeling and predictive modeling have their own advantages and disadvantages. The major advantage of Econometric models, according to Song and Li (2008), is the ability to analyze causal relationships between variables, i.e., they are comprehensible. These models do more than forecast; they help to understand economies of function, provide framework for progressive research strategy, and reasons for their own failures.

The one of the popular predictive models, artificial neural network (ANN) often has the advantage of accuracy. That is, the difference between the forecasted value by ANN and the actual value is not big. It does not require any prior information about the distribution and probability of data, can learn from past experience, and work with imperfect and non linear data (graph plotted from non linear data shows different shapes than straight line (Kenndy J.J., 2002)). But the results of ANN cannot be interpreted easily, it is a black box which takes input and produces output without clearly showing how its parameters are fitted. ANNs are difficult to understand (Dunham, 2003)

Many researchers have shown ANN outperforming econometrics models in forecasting accuracy. Burger et al (2001) concluded that ANN performs better than exponential smoothing, ARIMA, multiple regression and genetic regression models. They used tourism data to forecast the arrival of tourists from United States to Durban, South Africa, and used metric MAPE for the comparison of forecast accuracy. Cho (2003) also showed ANN outperforming exponential smoothing and ARIMA models in forecasting demand for Hong Kong Tourism sector. Similarly different researchers have shown that ANN is better than ARIMA or exponential smoothing models in other fields also. But according to Kon and Turner (2005); Palmer et al (2006) ANN provides satisfactory performance but systematic procedure for building model for ANN does not exist and the generally reliable forecasting models are achieved by trail-and error experiments. This statement motivates to compare econometric models with predictive models.

Most comparative studies conducted between econometric models and predictive models used only one or two error metrics (error metric is used to calculate difference between two variables) and in most cases root mean square error (RMSE). According to Armstrong (1985), RMSE is one of the worst evaluation metrics. So, trusting result only based on RMSE is not a good idea. Armstrong instead argues that it is necessary to evaluate models using several error matrices. Necessity of using different error metrics for accuracy of models also motivates a comparison of econometric modeling and predictive to some extent. In this study, retail data about the sales of products are used for comparing econometric and predictive models. Retail data often differ from many other problem domains since commercials often have a very large effect on the sales. Hence, the results done on other forecasting problems may not be applicable to the sales forecasting.

1.1 PROBLEM STATEMENT

Forecasting of sales of product is very important for the supermarkets. It helps to enhance performance of supermarket and enhance customer relationships and provide strategic advantages. Several forecasting models existing but it is difficult to choose best one among those existing. Accuracy of forecasting models is influenced by several factors like nature of data, scale of data etc.

Forecasting models can be compared and evaluated for their accuracy. Several error metrics are in use for the evaluation of models. Reliability of only one considered metric for evaluation is in question according to some earlier researchers as metrics are highly influenced by the nature of data. Use of several error metrics may be required for the evaluation of forecast models.

1.2 RESEARH QUESTION

On the basis of above problem statements the main research question of this thesis is:

Can econometric modeling technique be used as alternative to artificial neural networks and ensembles of neural network for sales forecasting?

The techniques will mainly be evaluated on accuracy based on several error metrics since it is suggested by Armstrong (1985) that it is best practice when comparing forecast models. However, since most studies only used one or two metrics this thesis will also try to answer if this is also necessary for sales forecasting. Hence a secondary research question is:

- *Does the use of several metrics contribute to the evaluation of sales forecasting models?*

1.3 MOTIVATION FOR CHOICE OF MODELS

Liu C., Wu K., Tsao M. (2005) have stated that ARIMA model has outstanding performance for forecasting. This motivates for choosing ARIMA model for study. Žilinskas and Žilinskas, (2010) motivate to study regression analysis by stating that the regression is one of the most popular forecasting models for estimating parameters of model. Similarly, Arsham H.(2010) have stated that exponential smoothing is the one of the most popular forecasting time series model used in business and industry.

According to Pindyck R.S. and Rubinfeld D.L. (1991) structural models are those econometric models in which the relationship between variables is specified on the basis of some economic theory. In the real world lots of models are dynamic in nature. Lag-structure and dynamic nature of models are important aspects for the model specification and testing. To represent lag-structure and dynamic nature in the model econometricians must be grounded well with theory. Sometimes it is possible that economic theory may not be sufficient to address dynamic structure of the problem properly in the model; the economic theory may be very complicated for precise specification of the model. To handle this type of situation, data itself should be allowed to specify the dynamic structure of the model instead of an econometrician. Vector autoregressive

(VAR) and other Autoregressive based models like ARMA, ARIMA provide means for specifying models on the basis of data using minimum economic theory. This also motivates to study ARIMA model for the comparison with predictive model.

[No purely structural models were used in this thesis due to the lack of econometric expertise.]

Maier H.R. (2000) mentioned that artificial neural network (ANN) is a very popular predictive model for forecasting in several areas like finance, power generation, medicine, water resources and environment science. Kourentzes N and Crone S.F. (2008) has mentioned that 77% of articles studied out of 84 articles showed that the neural network is better than benchmarks in forecast performance. These statements motivate to study ANN. Rao K.V. G, Chand P.P, Murthy M.V.R (2007) have mentioned that ANN ensemble technique is very popular amongst neural network practitioners. Friedman J.H. and Popescu B.E. (2005) have stated that learning ensembles are one of the most powerful learning methods used for the forecasting process. These statements motives to consider ANN ensemble model in this study.

1.4 TARGET GROUP

This thesis is evaluating econometric and predictive techniques. Hence, the focuses of this thesis is on those people who are studying, working on econometric modeling or predictive modeling or both, people related to the retail domains who are interested in forecasting of sales of product. Also this thesis focuses on people studying error metrics to some extent.

1.5 ORGANIZATION OF ARTICLE

This thesis is organized in the following way: the second chapter presents research strategy; the third chapter presents the theoretical background about econometrics, econometric models, and predictive models, related work, error metrics; the fourth chapter presents data collection and data preparation tasks; the fifth chapter presents experiment; the sixth chapter presents results of experiment; the seventh chapter present analysis and discussion of results; the eighth chapter presents conclusion of this thesis work and the ninth chapter mentions some suggestions and future works.

2. RESEARCH STRATEGY

The design of this thesis is quantitative research design. The quantitative research design focuses on quantifying the relationship between variables (Hopkins W.G., 2000). The research strategy set up for this thesis work can be explained as follows:

- Collection of articles related to this thesis work. Internet is the main source of articles. But it is difficult to find relevant articles. So, constraining of search text is necessary. Mostly texts with word econometric and predictive models are suitable for the search.
- Selection of relevant articles for this thesis work. The relevance of the articles can be decided according to the topic of the thesis and also using intuition of the researcher.
- Selection of most popular econometric models and predictive models, evaluation metrics from relevant articles and related works. Then setting up necessary theoretical background for this study using selected information.
- Collection and preparation of data for the experiment. Separation of datasets into training set and test set. Training set is set of data with both input and output known and is used to train the model. Test set is set of data used to test the trained model (Dunham, 2003).
- Running all considered models using the same set of training and test sets. Noting required values. Calculating values of all considered error metrics and comparing models on the base of calculated value of metrics,
- Analysis of experiment results and discussion of results using information given in the literature review section
- Writing conclusion based on experimental observations.

Above mentioned steps can be formalized in a diagram as follows:

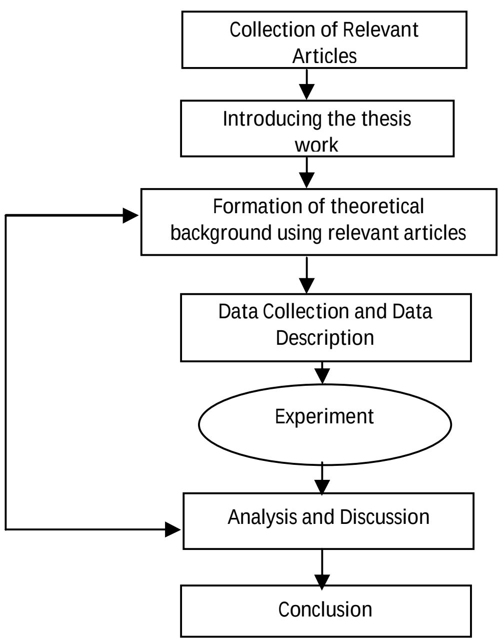


Figure 1 - Research Strategy

3. THEORETICAL BACKGROUND

3.1 FORECASTING

Forecasting is the process of estimating the value of one variable using another variable/s. Forecasting models help to estimate or forecast value of an unknown variable. According to Pindyck R.S. and Rubinfeld D.L. (1991) general categories of forecasting models are: a single-equation regression model in which the variable of interest is described by a single function consisting of explanatory variables; and time-series models in which structural knowledge about the causal relationship between a variable of interest and explanatory variables is not considered. The explanatory variables are those variables which affect the value of another variable called dependent variable. The variable to be forecasted is called the dependent variable, and variable/is used to forecast the value of dependent variable is called explanatory variables. The explanatory variable is known by other names also like independent variable, control variable, predictor variable. Similarly the dependent variable is known by other names like explained variable, response variable, predicted variable. Commercial affects the sales of product X, that is, if commercial is made the amount of sales of product X will change. Here commercial is explanatory variable and the sales of product is dependent variable. Similarly, it can be observed that the amount of fertilizers affecting the yield of crops. The amount of fertilizer is explanatory variable and the yield of crops is dependent variables.

3.2 ECONOMETRICS

Econometrics is about analyzing and explaining relationships between variables of the economic model. Generally, economic models show cause-and-effect relationship between dependent and explanatory variables. Econometrics are used directly or indirectly in lots of areas answering questions like: what will be the effect of increasing advertising budget by 10% on the sales of product, or how do job training, education and experience of employee affect the salary of the employee etc. Econometrics helps to forecast the future value of dependent variable on the basis of cause-and-effect relationship between variables.

According to Watson P.K., and Teelucksing S.S. (2002) the term “econometric” came to use with the start of Econometric society in the 1930s. Visionaries like Ragnar Frisch and Jan Tinbergen were part of that society. At that time econometrics was anything which included mathematical and statistical methods for economic analysis. But the fundamental tool for economical analysis at that time and even today is regression analysis.

In Economics, models are generally expressed in general terms of economic variables and parameters. Econometric methods provide precise meaning to such models. It gives numerical values to the parameters involved in the models and those models are used for numerical forecasting of the economic variables.

The econometric models used in its early days were not successful. According to Zellner A. And Palm F.C. (2004) the structural econometric models used for forecasting was of little success in its early days because of following reasons: economic theories were not able to construct perfect causal structure particularly for short-run dynamics; some of data required for model building were not available; and in some situations models were developed as a function of some unknown explanatory variables. And among these reasons, the third problem, models being function of unknown explanatory variables, was the major reason for the failure of structured econometric models at that time. To cope with this problem, economists and others started looking towards other kinds of models. The result was Time series models which were capable of showing trends in data and relation between past and present data. Exponential smoothing, ARIMA, Vector Autoregressive (VAR) are examples some of the time-series models.

3.3 TYPE OF ECONOMETRIC MODELS

Econometrics is divided into ***classical*** and ***modern*** econometrics. Econometric models represented using single equation and simultaneous equations are included in the classical econometric and models which show the relation between present and past values of a variable like time series models is included in the modern econometrics. (Watson P.K. and Teelucksing S.S., 2002).

3.3.1 Classical Econometrics

In classical econometric relationship between dependent and the explanatory variables are described by mathematical equations. They are used to study cause-and-effect relationship between variables and to forecast the value of one variable using other variables.

3.3.1.1 Single equation model

Single equation model based on economic theory is a model in which relationship between dependent and explanatory variables is represented by a single equation. A hypothetical example is presented here to illustrate single equation model. Sales of product X depends on explanatory variables like commercial made about the product, distribution of childsupport money, and salary of person. In some countries government support children by providing fund, childsupport, for the children. Salary of person is salary provided to him/her by organization for which he/she works. This scenario can be represented by following economic model:

$$SX = f(C, CS, S)$$

Where,

SX = sales of product X,
C = Commercial,
CS = ChildSupport

S = Salary

This economic model can be converted into econometric model using econometric methods into something like following:

$$SX = 31C + 5\,CS + 8\,S$$

To use econometric model for forecasting, data for all variables in the model are required. If values of parameters *C*, *CS*, and *S* are known then value of *SX* can be forecasted. If *n* number of observations is available then the above model can be written as:

$$SX_t = f(C_t, CS_t, S_t)$$

This model is converted into the linear equation form to apply econometric methods:

$$SX_t = \beta_0 + \beta_1(C_t) + \beta_1(CS_t) + \beta_1(S_t) \qquad (3.1)$$

Where, all variables are subscripted with *t* to show time period.

The equation (3.1) represents single equation econometric model. The values of β coefficients in the model are parameters to be determined. After determining values of β coefficients, model can be used for forecasting.

3.4 REGRESSION ANALYSIS

Regression analysis is the process of estimating value of dependent variable on the basis of explanatory variable/s (Kinney J.J., 2002). What will be the sales of product A if commercial about it is made? This is an example of simple linear regression, where sales of product (*y*) is dependent variable and commercial is explanatory variable (*x*). In the real world situation several examples can be found, where more than one explanatory variable is used to estimate dependent variable. What will be the effect of job training, experience and education on the salary of employee? This is an example of multiple regression where more than one explanatory variables: job training, experience, education are used. The relationship between dependent and explanatory variables is shown by the regression analysis.

Regression model is a model which shows how explanatory variable/s affect dependent variable. These models are established through sampling and are based on only possible observations (Kinney J.J., 2002).

The relationship between dependent and explanatory variables can be shown and studied using following regression model:

$$y = f(x) + u \qquad (3.2)$$

In this model y is dependent variable and x is explanatory variable. In real world situation it difficult to see affect of only explanatory variable on dependent variable as several other factors affects dependent variable. So, to model the real world situation properly term u known as error term or disturbance or noise is added in the above model (3.2). The error term u has mean zero and constant variance. The function f(x) is called regression function. The variable x with only one component is called simple regression and variable x with more than one component is called multiple regression. The regression function can have several forms such as:

$$a + bx, a + bx + cx^3, a + bx + cx_1x_3, \ldots,$$

These functions are defined by finite set of parameters *a, b, c.*

3.4.1 Simple Regression Model

Simple Regression model shows linear relationship between one dependent variable and one explanatory variable. It can be used to forecast dependent variable and can be represented with the following equation:

$$y = \beta_0 + \beta_1 x + u \qquad (3.3)$$

Where,

y = dependent variable

x = explanatory variable

u = error term

β_0, β_1= parameters to be estimated.

Above equation can be explained as follows: X amount of change in value of x will bring Y amount of change in value of y or if value of x is known, value of y can also be known. It is necessary that the values for parameters must be given. Equation (3.3) can be interpreted by connecting with sales of product and commercial about it as follows: sales of product (y) can be forecasted if commercial (x) made is known.

3.4.2 Multiple Regression Analysis: Estimation

Simple regression analysis can be used for empirical work but it will be difficult to draw ceteris paribus conclusions using it (Wooldridge, 2006). The sales of product is affected by commercial, and number of customers. Studying affect of commercial on sales of product keeping number of customers constant is an example of ceteris paribus. How x affects y: keeping all other factors affecting y fixed except x seems unrealistic. Multiple regression analysis is more useful for drawing ceteris paribus conclusions because it allows explicit control over many other factors that simultaneously affect the dependent variable. Multiple regression models can accommodate several explanatory variables that are correlated.

3.4.3 The model with two independent variables

For example: Wage of employee is affected by education and year of experience of employee. Here the wage is determined by the two explanatory variables education and experience and by other unobserved factors.

$$wage = \beta_0 + \beta_1 education + \beta_2 experience + u \qquad (3.4)$$

Here u is factors influencing wage other than education and experience. Here effect of education on wage, holding other factors fixed, is given by β_1. Here explanatory variable experience is taken out from the error term, u, and kept explicitly in the relation with its coefficient β_2. β_2 measures the ceteris paribus effect of experience on wage holding education fixed.

Equation (3.4) shows how observable factors other than variable of interest can be included in a regression model. Generally, a model with two explanatory variables can be written as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \qquad (3.5)$$

Where β_0 is the intercept, β_1 and β_2 measure the change in y with respect to x_1, and x_2 respectively holding other factors fixed for each other.

3.4.4 The model with *k* explanatory Variables

The general multiple linear regression model with *k* explanatory variables is written as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u \qquad (3.6)$$

Where,

β_0 is the intercept, β_1 parameter associated with $x_{1,}$, β_2 parameter associated with x_2, and so on and u is error term. The parameters $\beta_0, \ldots, \beta_k$ are estimated by solving system of (k+1) equations. It should be remembered that factors which are contained in *u* cannot be included as explanatory variable when applying multiple regression.

3.4.5 Mechanisms and Interpretation of Ordinary Least Squares

The estimated OLS equation for model containing only two explanatory variables can be written as follows:

$$\hat{y} = \widehat{\beta_0} + \widehat{\beta_1} x_1 + \widehat{\beta_2} x_2 \qquad (3.7)$$

Where, $\widehat{\beta_0}$ estimates β_0, $\widehat{\beta_1}$ estimates β_1 and $\widehat{\beta_2}$ estimates β_2. The OLS method is used to obtain values of $\widehat{\beta_0}, \widehat{\beta_1}, \widehat{\beta_2}$ such that the sum of squared residuals is minimized. That is for given *n* observations on y_i and x_i (i = 1,….,n), the estimates $\widehat{\beta_0}, \widehat{\beta_1}, \widehat{\beta_2}$ are chosen to make

$$\sum_{i=1}^{n}(y_i - \widehat{\beta_0} - \widehat{\beta_1}x_{i1} - \widehat{\beta_2}x_{i2})^2 \qquad (3.8)$$

minimum as much as possible. The *i* subscript refers to observation number. So, equation (3.8) is sum over all *i* =1 to *n* observations. Another subscript differentiates between explanatory variables.

In general equations with *k* explanatory variables, $\widehat{\beta_0}, \widehat{\beta_1}, \ldots.., \widehat{\beta_k}$ are estimated from following equation:

$$\hat{y} = \widehat{\beta_0} + \widehat{\beta_1}x_1 + \widehat{\beta_2}x_2 + \cdots . + \widehat{\beta_k}x_k \qquad (3.9)$$

Advantages of Regression Analysis

- 'what if…' kind of problems can be solved.

Disadvantages of Regression Analysis

- Accuracy cannot be guaranteed

3.5 MODERN ECONOMETRICS

3.5.1 Time-series Models

In time series model, the past behavior of a time series is examined to infer something about its future behavior instead of searching for effect of one or more variables on the forecast variable. Howrey E.P. (1980) stated that the time series models are based on a relatively weak nonparametric formulation of the model. They put more emphasis on the data analysis for simplification of the model.

Different patterns or trends can be seen in the time series data. The time series is influenced by several factors like random components, seasonal component, cyclic components etc. The random component in the time series may shed the influence of other components and make it difficult to describe the observed trends or patterns in the data. Removal or reduction of random components from the time series will result in better forecasting or interpretation of series. To remove or reduce effect of the random component from the time series smoothing technique are used. Smoothing of the time series is done before forecasting. Smoothing techniques like moving average and exponential smoothing can remove random components and seasonal components from the time series data (McClave, Benson, Sincich, 1998).

Some of the time series forecasting methods as follows:

- Moving Average
- Exponential Smoothing
- Holt-Winters Smoothing

3.5.2 Exponential Smoothing

Exponential smoothing is weighted average type of smoothing technique used to remove the random components from the time series data. It is used for short-range forecasting, generally one time period ahead (Kalekar P.S., 2004). It assigns positive weights, *w*, known as exponential smoothing constant to both past and current values of the time series. The value of w is taken between 0 and 1 including 0 and 1. Exponential smoothing generates smoothed series as follows:

$$E_1 = Y_1$$

$$E_2 = wY_2 + (1-w)E_1$$

$$E_3 = wY_3 + (1-w)E_2$$

$$\ldots\ldots\ldots\ldots$$

$$E_t = wY_t + (1-w)E_{t-1} \qquad (3.10)$$

Where E_t is the smoothed value, *w* is positive weight, and Y_t is current value in the series. Weight *w* is assigned to the current series data (Y_t) and the (1 – *w*) is assigned to the smoothed series data (E_t). The smoothness of series depends on the value of *w*. The small value of *w* gives more weight to the previous data and results in smoother series than original series. Large value of *w*, near to 1, gives more weight to current data and smoothed series will look like the original series.

The forecasting of next value of the time-series Y_{t+1} is simply the smoothed value at time *t*.

$$F_{t+1} = E_t \qquad (3.11)$$

Where F_{t+1} is the forecast of Y_{t+1}. This formula can be interpreted as follows using equation (3.10) and (3.11):

$$F_{t+1} = E_t = wY_t + (1-w)E_{t-1}$$

$$= wY_t + (1-w)F_t \qquad [F_t = E_{t-1}]$$

$$= F_t + w(Y_t - F_t) \qquad (3.12)$$

Equation (3.12) shows that the forecast for (*t* + 1) is equal to the sum of forecast for time *t* (F_t) and correction for the error in forecast for time *t*, $(Y_t - F_t)$. Exponential smoothing averages past and present values, the smoothed values lags behind the series if long-term trend exists. Therefore, exponential smoothing forecast is useful if the trend and seasonal component in time-series are relatively insignificant. If the time-series,Y_t is free of trend and seasonal components, exponential smoothing results same forecast for all future values of Y_t:

$$F_{t+2} = F_{t+1}$$

$$F_{t+3} = F_{t+2}$$

[Note: Derivation and explanation of above equations are inspired from book "Statistics for business and economics", 7th ed, by McClave, Benson, and Sincich]

3.5.3 The random walk model

Random walk is a simplest model containing stochastic trends given by following equation:

$$y_t = y_{t-1} + u_t$$

Here, u_t is called error term or white noise. In the simplest random walk process, future value of time series is given by its immediate previous (one step back) value.

3.5.4 ARIMA

ARIMA model is subset of univariate model in which time series is expressed in terms of past values of it, current and lagged value of a 'white noise' or error term. ARIMA models do not assume any knowledge about underlying economic model or structural relationships between variables (Meyler A., Kenny G., Quinn T., 1998). ARIMA model is formed by combining two models: Autoregressive model and Moving Average model.

3.5.4.1 Autoregressive Model

Autoregressive model represents current value of time series as combination of one or more previous values of the series. It shows the dependency of one value with its nearest previous values. Autoregressive process is a difference equation determined by random variables (difference equation shows current value of series as function of its previous values). Autoregressive model has order term, *p*, that determines how many previous values are to be included in the difference equation to estimate current value. A difference equation relates a variable X_t at time *t* with its previous values (Horvath Z., Johnston R., 2006).

The autoregressive AR(1), $p = 1$, includes only one previous value. It is a standard linear difference equation and written as:

$$X_t = \rho X_{t-1} + u_t, \qquad t = 0, \pm 1, \pm 2, \ldots \tag{3.13}$$

Where, u_t is error term and ρ is parameter to be estimated.

The p^{th} order AR time series is AR(p) and is given by the following expression:

$$X_t = \rho_1 X_{t-1} + \rho_2 X_{t-2} + \cdots + \rho_p X_{t-p} + u_t, \quad t = 0, \pm 1, \pm 2, \tag{3.14}$$

Where, $\rho_0 \neq 0, \rho_p \neq 0$, and u_t are uncorrelated random variables.

Using difference equation, value of X_t can be obtained from X_{t-1} , value of X_{t-1} is obtained from X_{t-2} and so on.

3.5.4.2 Fitting model

The AR(1) model is fitted with collected data by first estimating value of ρ. To estimate value of ρ least squares estimation method is used. It minimizes the sum of square of errors for the observed values with respect to ρ.

$$\frac{\partial}{\partial x}\sum_{t=2}^{n}(X_t - \rho X_{t-1})^2 = 2\sum_{t=2}^{n}(X_t - \rho X_{t-1})(-X_{t-1}) \qquad (3.15)$$

Equating above equation (3.15) to zero and solving it further gives the value of least square estimator for ρ:

$$\hat{\rho} = \frac{\sum_{t=2}^{n} X_t X_{t-1}}{\sum_{t=2}^{n} X_{t-1}^2}$$

From estimated value of ρ, distribution of error terms can be found.

$$\hat{u}_t = X_t - \hat{\rho} X_{t-1}$$

Now using estimated value of ρ and distribution of error data, the model can be fitted using equation (3.14).

3.5.4.3 Moving Average Process of order q, MA(*q*)

A time series is influenced by random shocks in noisy environment. As a result current value of series is affected by the random shocks appeared in previous values. Moving average terms are used to capture the influence of previous random shocks in the future value.

First order moving average or MA(1) is a simple time series, given by

$$X_t = \mu + u_t + \alpha u_{t-1} \qquad (3.16)$$

This equation says, apart from mean, μ, X_1 is a weighted average of u_1 and u_0, $X_2$is a weighted average of u_2 and u_1 etc. The values of X_t is defined in terms of random shocks u_t.

A Moving average of order q, MA(q) process X_t , is given by

$$X_t = u_t + \theta_t u_{t-1} + \dots + \theta_q u_{t-q} \qquad (3.17)$$

Above equation (3.17) representing MA (q) process is always stationary**.** In fact MA process is inverse of AR model. The MA model is invertible if an MA model can be expressed as autoregressive (infinite order) model

3.5.4.4 Autoregressive moving average process, ARMA(p,q)

Autoregressive Moving Average model is formed by combining terms of AR and MA models. Autoregressive model or Moving Average can be used to approximate any stationary process with any degree of accuracy as desired.

Combining equation (3.14) and (3.17), ARMA model of order p and q is formed,

$$X_t = \rho_1 X_{t-1} + \rho_2 X_{t-2} + \cdots + \rho_p X_{t-p} + u_t + \theta_t u_{t-1} + \ldots . + \theta_q u_{t-q} \qquad (3.18)$$

3.5.4.5 Autoregressive Integrated Moving Average Process of Order p, d, q, ARIMA (p, d, q)

The ARMA model assumes that the time series data is stationary (that is statistical properties of data do not change over time). But the real data are not stationary in nature. Time series data is made stationary by differencing process. The first order differencing process of time series X_t is defined as $X_t' = X_t - X_{t-1}$. ARMA time series which is made stationary by differencing process is known as Integrated Autoregressive Moving Average (ARIMA) model. ARIMA model is represented by three parameters: p order of autoregressive model, d order of differencing, and q order of moving average model.

ARIMA model takes historical data and decomposes that data into an autoregressive (AR) process which maintains memory of past events, an Integrated (I) process which makes data stationary for easy forecast and a Moving Average (MA) process of forecast errors. It does not suffer from existence of serial correlation between the error residuals and their own lagged values.

An ARIMA (*p*,*d*,*q*) model can be checked if it is a good statistical fit for data or not using Akaike Information Criterion (AIC) and Schwarz Criterion (SC) method. Autocorrelation (AC) and partial autocorrelation (PAC) statistics help to determine the right parameters for ARIMA model (Real Options Valuation, 2007).

Box-Jenkins has specified four stages for ARIMA model selection (Kahforoushan E., Zarif M., Mashahir E. B., 2010):-

- Determining values of p, d, q
- Estimate parameters of the model
- Checking whether considered model fits data properly or not, if not other consider another model
- Estimation using the best selected model.

3.5.4.6 Trend, cycles with ARMA model

It is obvious to ask question: How can ARMA model be applied to non-stationary data? Most real World Series show trend, that is up and down movements in the series. For example, sales of item can be seen to increase with advertisement and decrease without advertisement. Cyclic behavior like effect of discount provided on product can also be seen in the series. These up and down trends and cyclic effects can be removed from the series through differencing process. If t is in months and Y_t is a series in which an increasing trends are seen by some constant amount K every month. Then expression $Y_t = K + Y_{t-1} + N_t$ can be written, where, N_t Noise component with expectation zero. The difference between the current data and one step previous data can be written as $Y_t - Y_{t-1} = K + N_t$. Here $K + N_t$ is stationary series without the linear trend. Now ARMA model to series $z_t = K + N_t$ can be applied. Cycles can be removed also and made stationary by differentiating at different lags (**2T Time Series**, 2010).

Advantages of ARIMA

- It requires only data on the time series in question and this is advantage in case of forecasting large number of time series.
- No problem of timelines of data

Disadvantages of ARIMA model:

- Model identification may require skill and experience of the forecaster
- No underlying theoretical model or structural relationships is assumed. According to Kmenta J., Ramsey J.B., (1980), time series models are developed with little or no economic theory, so these models are not good in showing cause-and-effect relationship between different variables of the system under study.

3.6 PREDICTIVE MODELING

Predictive modeling is a technique which considers one field from tabulated dataset as target (or dependent) variable, other fields of tabulated dataset as explanatory variables produces models using target and explanatory variables and produces forecast value for the target variable. The main problem addressed by this technique is to produce accurate values for target variable using noisy data (Apte C. et al, 2002). Neural network, ensemble, decision trees are types of predictive models.

3.6.1 Artificial Neural Networks (ANN)

Artificial neural networks (ANN) are systems that are used for classification and forecasting process. It consists of graph and various algorithms to access the graph. Each element of ANN is independent of each other and functions as an independent unit (Dunham M.H., 2003).

The characteristic of neural network is defined by its three components: architecture, learning algorithm and activation function (Idri A., Khoshgoftaar T. M., Abran A., 2002). The structure of ANN consists of input (source), output (sink) and hidden (internal) layers. The number of hidden layers in ANN can be one or more or sometimes zero (that is no hidden layer). It takes records at the input layer and produces forecast value at the output layer. ANN has one input node per variable and can have more than one output node for the output variables. The structure of ANN can be explained with following figure 2.

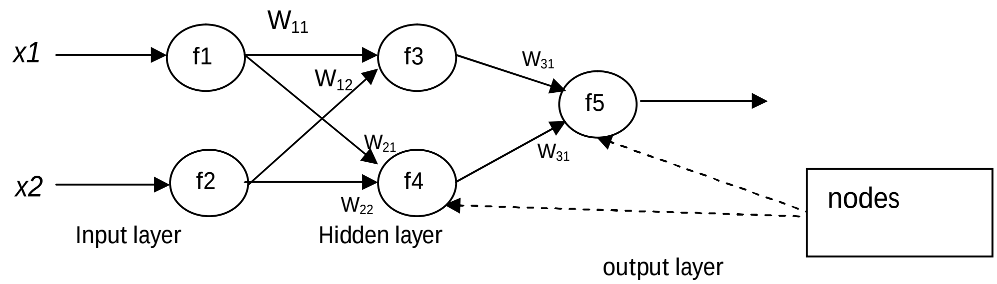


Figure 2 - Neural Network

Layers of ANN are connected to each other by arcs. Values known as weights are assigned to all arcs. Each node of a layer is connected with each node of its successive layer. A function *f*, known as activation or squashing function, is applied to the input of node to produce output. Some of the proposed activation functions are threshold, sigmoid, symmetric sigmoid, Gaussian.

Multilayer Perceptron (MLP): Figure 2 is Multilayer perceptron. It is the simplest feed forward ANN. It has more than one layer. In this type of ANN outputs of nodes always flow towards the output layer (Dunham, 2003).

ANNs are good at solving complex problems like pattern recognition, speech recognition and synthesis, medical applications, robot control, fault detection etc. They have ability to “learn” from the prior applications (Dunham, 2003; Idri A., Khoshgoftaar T. M., Abran A., 2002). If the generated output is not good, it can be improved next time by adjusting its units. It requires less formal statistical training and has ability to find nonlinear relationships between explanatory and dependent variables, and there is also possibility to use different training algorithms. ANNs have become extremely powerful and are also used in areas like: finance, power generation, medicine, water resources and environmental science (Maier H.R., 2000).

ANN is trained using training data. Test set is used to test the trained ANN. The training set accuracy of ANN is increased by continually training it with the training data. But the continual training may decrease test set accuracy. This problem is called overfitting and ANN is prone to it (Rosin P.L, Fierens F., 1995).

According to Shachmurove Y., (2005) , Kline D. K., & Kohers G., (2000) ANN has ability to analyze complex patterns with high degree of accuracy. It does not make any assumptions about the nature of data, so it is not biased in their analysis. It is capable of handling nonlinear data and performs well even with incomplete data set. It is easy to use when forecast is to be made in shorter period of time. But ANN models are difficult to understand and are not of good use if the controlling of possible variables and potential outcomes in the system is necessary. ANNs are not all-purpose problem solvers as there is no structured methodology available for choosing, developing, training and verifying an ANN. They are “black boxes”; it is impossible to known how relations in their hidden layers are estimated.

Ostafe D. and Jeffheaton (2008) state that the selection of number of hidden layers and the number of units or nodes in hidden layer are difficult tasks in ANN. There are no fixed accepted theories for calculating the number of hidden layers or the number of nodes in the hidden layers. If the number of nodes is less than required, problem of under fitting will arise and if the number of nodes is more than required, problem of over fitting will arise. The problem of under fitting means neural networks can’t learn the entire information. The problem of over fitting results in complexity in network and training set with limited amount of information may not be sufficient to train all of the nodes in the hidden layers.

3.6.2 Ensemble

Ensemble (ensemble of ANN) is a set of classification or regression models which are used together for the purpose of forecasting. Since 1990, ensembles techniques have been used for daily numerical weather forecasts. The structural model of ensemble takes following form:

$$F(\mathbf{x}) = a_0 + \sum_{n=1}^{N} a_n f_n(\mathbf{x})$$

Where N is the number of ensemble members (models) and $f_n(\mathbf{x})$ are members of ensemble. Each member of ensemble is a different function of the input variable x taken from the training data. Output of the ensemble is produced by calculating the mean of output produced by each member**.** The mean of forecasts produced by the members of ensemble offers better forecast than individual member (Zhu Y., 2005).

Friedman J.H., and Popescu B.E. (2005) have stated that the forecast results given by the ensemble is often more accurate than the best models. A best model however can sometimes be interpreted which is a big advantage.

Bagging and Boosting are two popular methods for building accurate ensembles (Opitz D., Maclin R., 1999). Both methods use “resampling” technique to obtain different training sets for different classifiers. Boosting method produces series of classifiers. Different training sets are used for different members of each series. The performance of one member of a series influences the choice of training set for another member of the same series. Boosting method tries to produce better classifier by training incorrectly classified training sets again and again.

Bagging is simpler but more robust than boosting. It is highly more parallel technique. It uses as set of training sets and a class of classification models. Multiple models are trained on different samples and average of their forecast is taken.

An ensemble is an important forecast model due to:

- Forecast error: Ensemble produces a set of independent solutions for the future which help to reduce forecast error.
- Predictability: The reduction of forecast error in the ensemble module greatly increases the predictability.

3.7 ERROR METRICS

Forecasting models need to be evaluated from different perspectives like: how much forecasted values deviate from actual value; the model used to forecast is useful or not; strength of linear relationship between dependent and independent variables.

Armstrong J.S. (1985) has found that use of only one metrics for the comparison of forecasting model is not suitable as the result of different metrics differ. He has suggested to use multiple error metrics for testing forecasting models. The result of one metric may not be reliable and above all RMSE is one of the worst metrics. One metric may perform well in one situation but same metric may not perform that well in another situation.

Forecasted values obtained from different forecasting models may differ. The error metrics show how risky the forecast model is. The model is tested by taking difference between actual value and forecast value, less the difference, better the model is. Several criteria can be used to compare different forecasting models. According to (Zeng T., Swanson N.R., 1998), some of methods for evaluating forecasting models are as follows:

- Root Mean Square Error (RMSE)
- Mean Absolute Error (MAE)
- Mean Absolute Percentage Error (MAPE)
- Weighted Mean Absolute Percentage Error (WMAPE)

Rummel R.J. (1976) has considered linear correlation as the workhorse of quantitative research and analysis. This motivates to consider linear correlation in comparing models. In this study five different error metrics are considered for the evaluation of forecasting models. They are root mean square error (RMSE), mean absolute percent error (MAPE), weighted mean absolute percent error (WMAPE), mean absolute error (MAE) and linear correlation.

3.7.1 Root Mean Square Error (RMSE)

Root mean Square Error (RMSE) is square root of average of sum-squared errors and is given by following formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Where

$\widehat{y_i}$ = estimated value of y_i,
y_i = actual value
n = number of observations

There is one problem with RMSE and it is that they may be close to 0 if large positive and negative errors cancel out each other. RMSE gives high weight to the large errors and are generally useful where large errors are not of importance.

RMSE are more sensitive than other metrics to the infrequent large errors as the squaring process gives large weight to very large errors (Decision 411, 2010).

3.7.2 Mean Absolute Error (MAE)

The problem of RMSE, canceling out of large positive and negative errors can be avoided by using Mean Absolute Errors. In average, MAE weights all the differences equally.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i}$$

MAE and RMSE can be used together to study the variation in the errors in a set of forecasts. RMSE are always larger or equal than MAE.

3.7.3 Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE) is the sum of absolute errors divided by actual values. MAPE expresses the value of error relative to actual value for observation *i*. The relative measure is expressed as percent.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| * 100$$

The concept of MAPE is simple and convincing but it has limitations in practical application:

- The possibility of zero value for actual observation results in division by zero
- MAPE is zero in case of perfect fit.
- Large error can unfairly skew the overall error
- A MAPE of 25% means that the forecast is over or underestimated by 25%, it is scale free

3.7.4 Weighted Mean Absolute Percentage Error (WMAPE)

Weighted mean absolute percentage error eliminates the problem in MAPE, problem of large error skewing the overall error. WMAPE is given by:

$$WMAPE = \frac{\sum \left| \frac{y_i - \hat{y}_i}{y_i} \right| * w}{\sum y_i}$$

It does not have problem such as over-skewing due to low or zeros volumes. WMAPE is good even where there are infrequent large values in the series.

3.7.5 Linear Correlation (r)

The Linear correlation shows the strength of relationship between dependent and explanatory variable (MathsBit.com, 2010).

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}}$$

Where,

x = actual value
y = estimated value
Value of r lies between -1 and +1..

The value of r can be interpreted as follows:

- Value of *r* nearly zero or zero implies that there is little or no relationship between *y* and *x*
- Value of *r* closer to -1 or 1 implies that there is strong linear relation between *y* and *x*
- Positive value of *r* means that *y* value increases as *x* value increases
- Negative value of *r* means that *y* value decreases as *x* value increa**ses**.

One point must be remembered that the causal relationship does not imply causality. It is possible that increase/decrease in value of *y* may be caused by some other factors expect *x*. Hence, it will be wise to conclude that when high sample correlation is seen means there may exist linear trend between *x* and *y* (McClave, Benson, Sincich, 1998).

3.7.6 Properties of error metrics

Error metrics can affect the ranking of forecasting methods as they are affected by several factors (Armstrong, 1985). So it is obvious that the rank of one forecasting method differ from rank of another forecast method on the same data set. Some of the factors that influence the error metrics are as follows:

- Scale of data
- Nature of data
- Outliers in the data

Note: If infrequent large values are not a problem in decision situation then the MAE or MAPE be more relevant error metric than RMSE (Decision 411, 2010).

3.8 RELATED WORK

Several comparative studies between predictive modeling and econometric modeling have been done using different data from different domains. In this related work section some of the previous studies about econometric and predictive modeling are included.

Pope R.D. et al. (1979) compared econometric models for US farmland prices with ARIMA model. Herdt and Cochrane's simultaneous equation model and Klinefelter single equation models were considered for the analysis.

The Herdt and Cochrane's simultaneous model used was:

$$N_s = f(P, R, U, Lf) < supply\ equation >$$

$$N_d = f(P, R, T, P_r/Pp, G) < demand\ equation >$$

$$N_s = N_d < identity >$$

Where,

N_s is number of farms supplied,
N_d is number of farms demanded;
P is average value per arce
R is rate of return on nonfarm investment
U is unemployment rate
Lf is amount of land in farms
T is the USDA productivity index
P_r/Pp is ratio of index of prices received by farmers to the index of prices paid by farmers and
G is the wholesale price of index

The considered models were compared with a naïve forecasting model (Box-Jenkins). Root mean square error (RMSE) was the metric used to compare models. The result of study showed that the simple single equation model can perform better than simultaneous equations and ARIMA model in case of simple economic model.

Yoon and Swales (1991) considered multiple Discriminant analysis (MDA) models and ANN to forecast stock price. MDA is a set of simultaneous equations. Data were gathered from two information sources: The Fortune 500 and Business Week's "Top 100". A stock's total return and market valuation were the error metrics used to evaluate models. MDA model was 74% correct on the training set and 65% correct on the test set. Artificial Neural network was 91% correct on the training set and 77.5% correct on the test set. This showed that neural network is better than MDA. In addition to this result the authors have stated that MDA has the capability to explain the feature and significance of each input parameters.

Kudyba S. (1998) compared neural net-based computer algorithm with Semtsa (structural econometric model time series analysis). Data about electricity demand in US from 1945 to 1990 were considered for the comparison of models. Neural network was developed using all the information used to develop the Semtsa model. Measures used were adjusted R-square in conjunction with t-, t^2- and F-statistics. Also root mean square error (RMSE) and the mean absolute percent error (MAPE) were used to compare the accuracy of each model. The result showed that the neural network for electricity demand is better than complex Semtsa. It was also seen that Semtsa was more expressive than neural network in explaining outputs.

Gruca T.S., Klemz B.R., Petersen E.A.F (1999) compared artificial neural network with Multiplicative Competitive Interaction (MCI) models using sales data of coffee and A.C. Nielsen catsup. (MCI model is a set of equations and parameters.) Data were obtained from Sioux Falls, SD. The coffee dataset contained 52 weeks data from March 1981 to April 1982. This data set was split into two samples: 43 weeks data as estimation sample and 9 weeks data as hold-out sample. The catsup data was of 156 weeks from August 1985 to August 1988. This data set was also separated into two samples: 146 week data as estimation sample and 10 week data as hold-out sample. Coffee dataset was small dataset with only few observations and catsup dataset was large dataset with enough data. The mean absolute percentage error (MAPE) was used to compare models. The result showed that ANN is more accurate than MCI.

Moshiri S. and Cameron N. (2000) compared different types of Back Propagation Neural Network (BPN) with econometric models using data about inflation rate. Different BPN considered were: BPN, BPN with ARIMA, and BPN with VAR model. Different time series models were considered as econometric models and they were: an ARIMA model, a vector autoregression (VAR) model, and Bayesian vector autoregression (BVAR) model. Monthly data from 1973:1 – 1994:12 about inflation rate, GDP gap, money supply, and import price inflation were obtained from the CANISM databank. The dataset was divided into two sets: data from 1970:1-1990:12 as training set and data from 1991:1 – 1994:12 as test set. Root mean square Errors (RMSE) and Mean Absolute Errors (MAE) were the metrics used for calculating forecast quality. Information test method was also used to compare usefulness of models according to information content. The test result showed hybrid BPN were similar or better than their equivalent econometric model in dynamic forecasting.

Camargo M. E., et al (2009) performed comparative study between artificial Neural Networks and ARIMA model using 8 years sales data collected from a medium sized enterprise in Brazil. The data were from January 2000 to December 2008. MAPE and residual variation were the performance metrics used. The results of comparison were as follows shown in the table 1.

Forecasting model	**RMSE**	**MAPE**
ARIMA	0.1235	0.6812
Neural Network	0.1027	0.4765

Table 1 - Values of observed metrics

Experiments showed that artificial neural network adjusted well with the sales data and provided satisfactory forecast.

Tjung L.C. et al (2010) compared neural network with the regression model (OLS method was used to estimate parameters of regression model) using financial stock data. Eight explanatory variables were used for forecasting financial stock prices. SPSS program was used to create unique regression model and Alyuda NeuroIntelligence program was executed to create neural network model. The mean and standard deviation of the % error were the evaluation metrics used to compare models. Authors also calculated Adjusted R-square value. The result of comparison was that neural network is more accurate than OLS. The accuracy of neural network was 96% while accuracy of OLS was only 68%. The study also showed some difficulties with neural networks. Neural network is complex, it requires more training time to find the best model. It has problem of over-fitting and it cannot be assured perfectly that the model created is the best because it is a blind search.

Shu Chang and Burn D.H. (2003) forecasted flood frequency using ensemble of ANN and single ANN. The accuracy of ensemble ANN was found better than single ANN and it was found that ensemble is less sensitive to the choice of initial parameters. The evaluation metrics used for the comparison between ensemble of ANN and single ANN were relative squared error (RSError), percent relative error (PRError) and relative bias (RBias). Further they compared ensemble of ANN with multiple regression and found that ensemble is better than it. The major finding of their experiment was that properly designed ensemble of ANN is better than single ANN and multiple regression model.

Armstrong (1985) has mentioned several points about the error metrics. Some of them are:

- MAPE is biased if the data series contains only positive numbers and it favors low forecast
- RMSE is strongly influenced by the scale of series and is unreliable if data contains outliers
- Adjusted MAPE or similar error metrics are more reliable than MAPE

According to Decision 411 forecasting (2010), there are no absolute criteria for a “good” value of RMSE or MAE as they depend on the units of variable and degree of forecasting accuracy. RMSE is always greater than MAE. If the difference is great, then there will be great variance in the individual errors in the data. If RMSE = MAE then all the errors will be same.

The fluctuations in data are taken in account by WMAPE and if the fluctuations in data are small WMAPE simply turns into the ordinary MAPE (Schutz W., Kolassa S., 2011)

In 1981, a survey found RMSE was more popular than MAPE, 48% of 62 academics and 33% of 61 practitioners used RMSE while only 24% of academics and 11% of practitioners used MAPE. But a decade later MAPE was found to be the most commonly used metric (52%) compare to RMSE (10%) (Armstrong, 1985).

4. DATA COLLECTION AND PREPARATION

Data were obtained from the ICA group which is a leading retail company in north Europe. ICA has retail stores in Sweden, Norway, Estonia, Latvia and Lithuania. Products are bought and sold to stores in Sweden and franchise stores in Norway. Services like marketing, logistics, training and in-store technology are provided to the stores by the headquarter of ICA.

4.1 DATA

Sales data of five products were considered for the experiment. The five products were Frozen chicken, Sausages, Frozen vegetables, Frozen fish Gratin, and Sandwich Ham. The data sets contained data for approximately two years and were stored in excel file. There were 100 records for Frozen chicken, 101 records for Sausages, 96 records for Frozen vegetables, 99 records for Frozen fish Gratin, and 104 records for Sandwich Ham.

Sales, i.e. the number of items sold for a certain product is the target or dependent variable. (Target variable is a variable which value is forecasted). Explanatory variables are commercials, child support, salary, priceindex. The variable *commercial* refers to the advertisements made for the product. *Child support* is binary attribute which shows if a governmental child support have been paid for the current week or not. Similarly, *salary* signals if the salaries are paid during the current week or not. *PriceIndex* is variable which shows how the current price differs from the average price for the current year. A value below one signifies a lower price, i.e., a discount.

The sales are weekly sales rolled up from daily sales. Roll-up is the process of displaying data given in lower level to higher level (Han J., Kamber M., 2005). Each seven days data is summed and presented as the sales for the week. Here daily data is in lower level and weekly data is in higher level.

From each explanatory variable another lagged variable was formed and named accordingly. Lagged variables contained values of the previous week for the respective variables. These variables were added to show the possible effects of their previous values on the current sales of the product. The total number of records for each product and number of explanatory variables is shown in following table 2.

Product	Test Set No. Records(n)	Number of explanatory variables (k)
Frozen chicken	100	10
Sausages	101	10
Frozen vegetables	96	12
Frozen Fish Gratin	99	8
Sandwich Ham	104	10

Table 2 - Number of records and explanatory variables for each product

4.2 PATTERN OF DATA

To see the general sales patterns and the relationship between the explanatory and dependent variables the sales were plotted against the other attributes. The plotted graphs showed strong influence of commercials on the sales.

Graph between commercials and sales of product are shown in following figures. Sales of product is shown in y-axis and weeks are shown in x-axis.

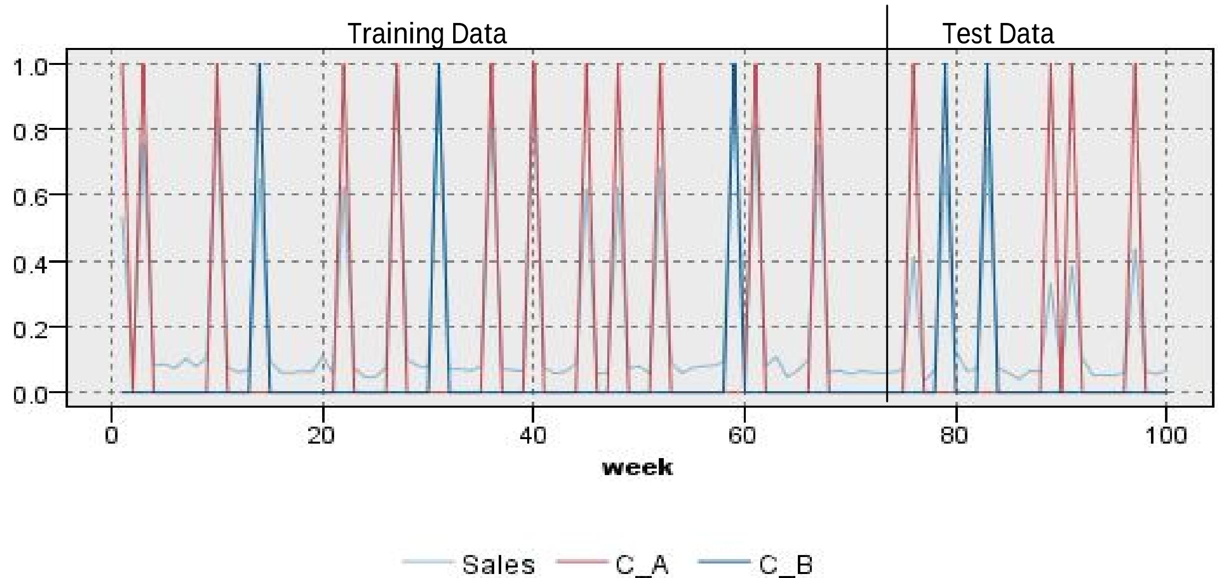


Figure 3 - Sales of Frozen chicken versus Commercial

From the figure 3, it is seen that the sales of frozen chicken are highly influenced by commercials. The sales of the frozen chicken increased whenever commercial was made. Two commercials C_A and C_B are made, and both commercials helped to increase sales. For example, between weeks 40 and 60 commercial C_A helped to increase sales of product.

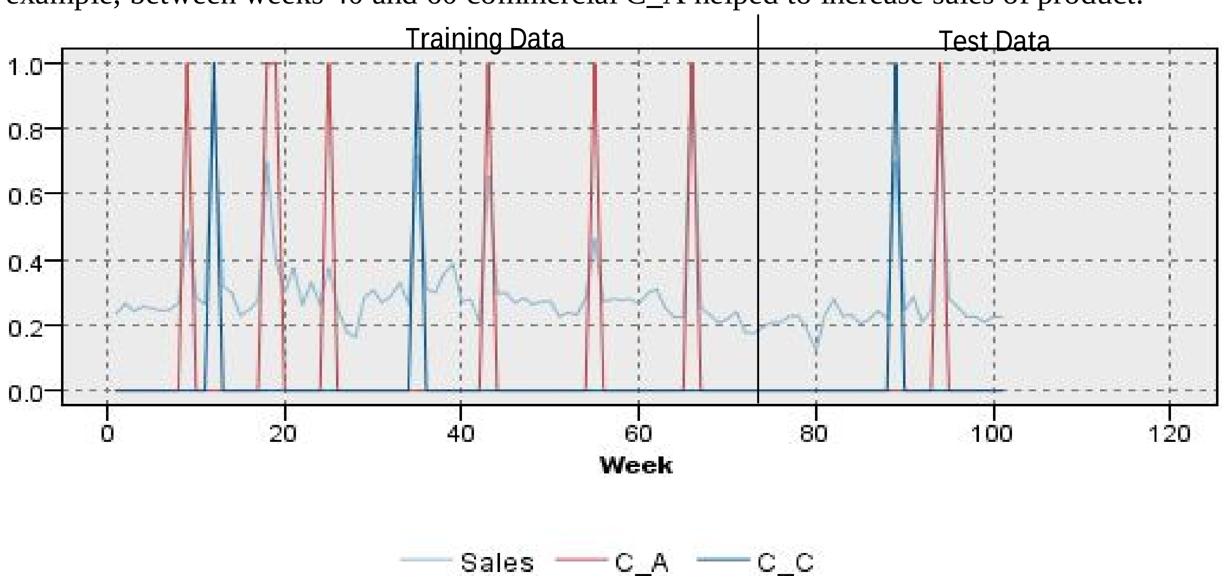


Figure 4 - Sales of Sausages versus Commercials

For Sausages, two commercials were made. The sales of Sausages were increased more when second commercial C_C was made than first commercial C_A. In some places slight rise in sales was also noticed where none of the commercials were made. In ninth week commercial helped to increase sales, but in week 39 sales increased even there were no commercials

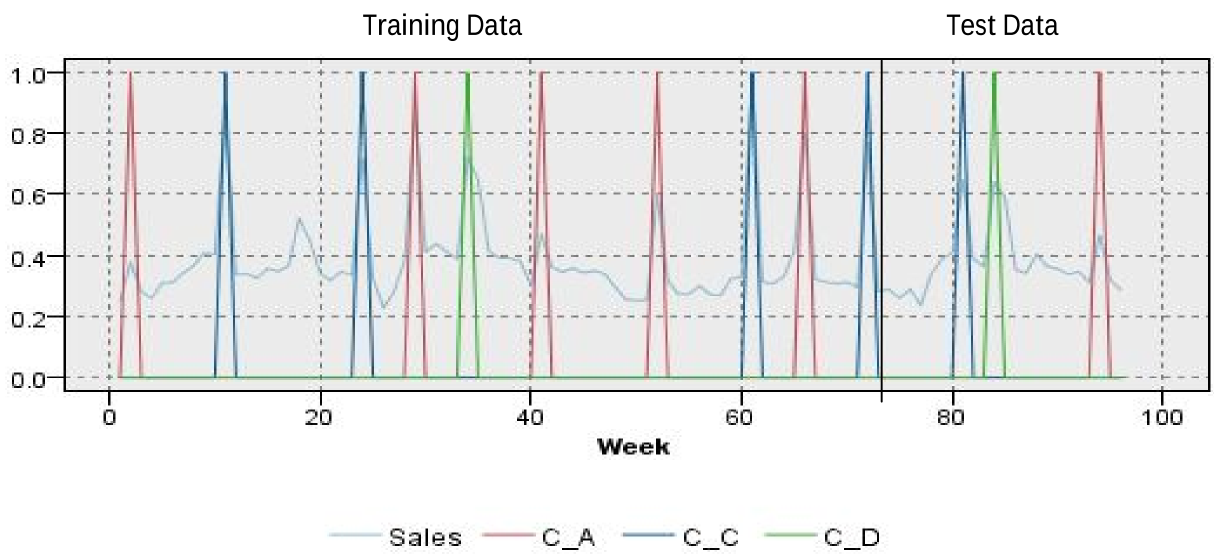


Figure 5 - Sales of Frozen vegetables versus Commercials

Three commercials were made for frozen vegetables. The effect of all three commercials was seen on the frozen vegetables. In some cases commercials were seen to increase sales highly but in some other cases it was not so. Also moderate rise in sales of frozen vegetables was seen where no commercials were made. In week 23 sales was increased with commercial C_A but in 17 week sales increased even no commercials were made.

Figure 6 – Sales of Frozen Fish Gratin versus Commercials

Only one commercial was made for the frozen fish gratin. Here also the commercial helped to increase the sales of product. But in some instances sales raised even there was no commercial. In week 32 commercial was made and huge increase in sales was seen.

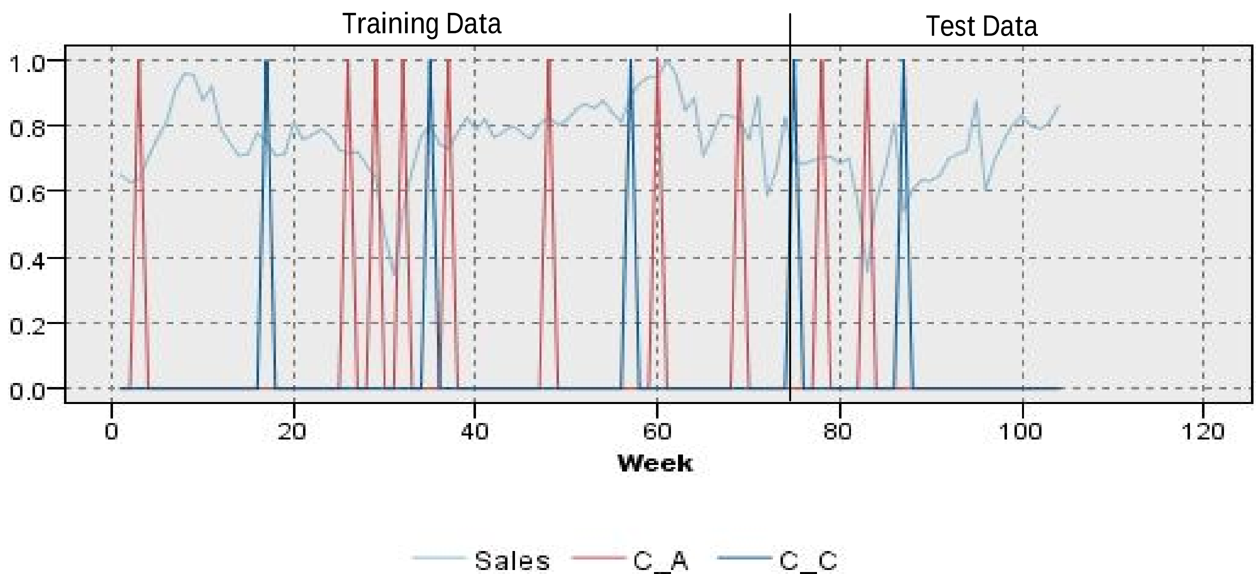


Figure 7 - Sales of Sandwich Ham versus Commercials

Two commercials were made for sandwich ham. For this product affect of commercials was difficult to see as there were lots of variations in the data. In 7th, 8th, 9th weeks sales increased without commercials. Between weeks 20 and 40 several commercials were made but increase in sales was not seen.

The plotted graphs suggest that there is effect of commercials on the sales. Commercials help to increase the sales or this can be said another way also: cause-and-effect relationship between the sales and the commercial made about them is evident.

5. EXPERIMENT

5.1 TITLE

Comparing Econometric modeling and predictive modeling: Artificial Neural Networks using sales data

5.2 INTRODUCTION

This experiment is about the comparison between the econometric models and predictive models to see which model is more accurate in estimating future values of the sales of product. Several metrics are used to evaluate the accuracy of each considered models. The considered econometric models are: Linear regression model, ARIMA and exponential smoothing, and predictive models are: ANN and ensemble of ANN.

5.2.1 Ranking of models

For the comparison purpose models are ranked for each metrics. The purpose of ranking of models for metric was also to see if the rank of model differs according to the metric or not. Ranking was done as follows:

	RW	LR	Expo	ARIMA	ANN	Ensemble
MAPE	220.00	26.90	168.40	32.58	29.25	10.47

Table 3 - Values of models for MAPE

[Where, RW = Random Walk, LR = Linear Regression, Expo = Exponential Smoothing, Rk = Rank, val=value]

Here, values for MAPE for each model are stored in a row. Model with lowest metric value is given rank 1, model with second lowest metric value is given rank 2 and so on. Here as Ensemble has lowest value of 10.47 it is ranked 1. Table 4 shows models with their values and rank for MAPE.

	RW	Rank	LR	Rank	Expo	Rank	ARIMS	Rank	ANN	Rank	Ensemble	Rank
MAPE	220.00	6	26.90	2	168.40	5	32.58	4	29.25	3	10.47	1

Table 4 - Values and Ranks of models for MAPE

The average of ranks of a model according to metrics is used to compare models on overall performance of model on the data set and it is calculated as follows: Ranks of all metrics (given in column) for a model was added and divided by 5 as there were five metrics.

The comparison of models on the basis of products is also necessary. For this average of the average rank of each model for each product on the test set is used.

5.2.2 Number of differences between any two metrics considering all models and products

The rank between two metrics differed in several places. So, to see how many times rank between two metrics differed out of total 70 places (7 models and 10 tables) is counted and difference in percentage is also calculated. The total percentage difference between any two metrics for a model is calculated as follows: Here in table 5 values of MAPE and WMAPE differed for LR, Expo, ANN and Bagging, i.e. in 4 places. There are total 7values to compare between MAPE and WMAPE. So, difference in percentage is 57 (4/7%). In this way difference between any two metrics is calculated considering all the ten tables (both training and test).

	RM	LR	Expo	ARIMA	ANN	Boosting	Bagging
MAPE	6	4	1	7	2	5	3
WMAPE	6	3	4	7	1	5	2
RMSE	6	2	5	7	1	4	3
MAE	6	3	5	7	2	4	1
R	7	1	5	6	3	4	2

Table 5 – Rank of models for different metrics

In ten tables (training and test), there are 70 values for any metric. Hence the total percentage difference between any two considered metrics is calculated as x/70%, where x is number of differences in rank between two considered metrics considering all ten tables.

5.3 TOOLS USED

- SPSS PASW Modeler 14
- MS Excel 2007

5.4 INPUT

Weekly Sales Data about the sales of five products: Frozen chicken, Sausage, Frozen vegetables, Frozen fish Gratin and Sandwich Ham were inputs.

Variable Name	**Type**	**Description**
Sales	Dependent variable	Variable to be forecasted
Commercials	Explanatory variable	Commercials are made for advertising product and it influences Sales,
Child Support	Explanatory variable	Money provided to child by government
Salary	Explanatory variable	Money provided for work done by office to its employee
Price index	Explanatory variable	shows how the current price differs from the average price for the current year

Table 6 – Description of variables

5.5 STEPS

1. New column in the worksheets of each product was inserted and named **TYPE**. The value **TRAIN** was inserted for TYPE column for first 75 records, and the value **TEST** was inserted for the rest of the records. This new column TYPE was inserted to separate training set and test set.
2. Stream with all necessary nodes for a model was created in the SPSS PASW Modeler 14. (See Appendix C for Steps to create stream).
3. This modeling technique was executed to produce model for Training data.
4. Model produced in step 3 was copied and connected with the nodes representing test data.
5. The model of the training set was executed to produce output. From that output:
   a. The actual sales value and estimated sales value were copied
   b. The copied values were stored in the excel sheet
   c. The values for all considered metrics were calculated using actual and estimated sales values.
6. The model of the test set was executed to produce output. From that output: Steps (5.a) (5.b) and (5.c) were repeated for test set.
7. All the steps from step (2) to step 5 were executed for each model.

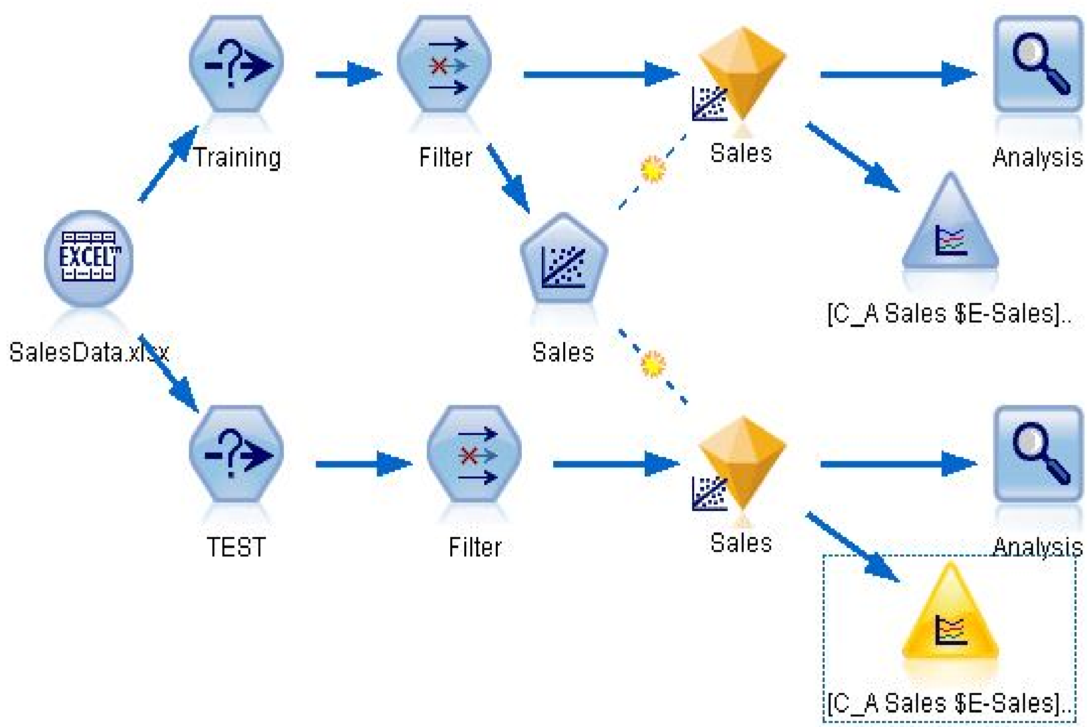


Figure 8 - Stream of nodes created in PASW modeler

8. Each model was ranked for each metric.
9. Average Rank of each model for each data set was calculated
10. Average Rank of each model on the basis of average rank of each product on the test set was calculated.

In this experiment, Multilayer perceptron (MLP) was used as ANN model. ANN was run for 10 times (with training enabled). For ensemble, boosting and bagging methods were used. Mean was used as combining rule for the target variable for ensembles.

For exponential smoothing:

Time Interval: weeks

For ARIMA:

Time Interval: Weeks

Value for parameters of ARIMA were:

Autoregressive (p): 1
Difference (d): 1
Moving Average (q): 1

Models	**Criteria**
Random walk	Difference between previous value and current value
Linear Regression	Mentioning dependent and explanatory variables
Exponential Smoothing	Specifying Time Interval: **Weeks,** Weight
ARIMA	Specifying Time Interval: Weeks, Specifying parameters: p = 1, d = 1, q = 1
Neural Networks	MLP Running for ten times and taking average of ten runs
Ensemble Boosting	Method: Boosting
Ensemble Bagging	Method: Bagging

Table 7 – Models specification

6. RESULTS

6.1 EVALUATION OF ECONOMETRIC TECHNIQUES

The values and ranks of error metrics for each **econometric model** are tabulated in following tables.

Frozen Chicken

	RW		LR		Expo		ARIMA	
	Val	**Rk**	**Val**	**Rk**	**Val**	**Rk**	**Val**	**Rk**
MAPE	220.00	4	26.90	1	168.40	3	32.58	2
WMAPE	1.38	4	0.17	1	1.03	3	0.21	2
RMSE	488.28	4	59.15	1	313.89	3	89.51	2
MAE	316.29	4	38.03	1	236.98	3	41.43	2
r	-0.18	4	0.98	1	-0.27	3	0.98	1
		4		**1**		**3**		**1.8**

Table 8 - Frozen Chicken (Training Set)

	RW		LR		Expo		ARIMA	
	Val	**Rk**	**Val**	**Rk**	**Val**	**Rk**	**Val**	**Rk**
MAPE	182.50	4	45.61	1	179.44	3	140.56	2
WMAPE	1.26	4	0.52	1	1.00	2	1.09	3
RMSE	338.69	4	160.89	1	218.64	2	334.75	3
MAE	230.67	4	94.29	1	183.18	2	188.67	3
R	-0.16	2	0.96	1	-0.01	4	0.12	3
		3.6		**1**		**2.6**		**2.8**

Table 9 - Frozen Chicken (Test Set)

Tables 8 and 9 show that:

- Linear Regression is accurate than other considered econometric models
- Change in rank of same model for different metrics is seen, i.e. exponential smoothing is ranked 3rd by MAPE but 2nd by WMAPE. ARIMA ranked 2nd by MAPE but 3rd by other metrics for test data.
- For the test data set exponential smoothing is seen better than ARIMA and Random walk.
- Ranking of Exponential smoothing and ARIMA differed on MAPE and WMAPE
- On RMSE and MAE none of the models differed in rank
- High correlation is shown by Linear regression
- Time series models (ARIMA, Exponential Smoothing, Random walk) showed very low linear correlation and it is not surprising as these models don't consider cause-and-effect relationship between variables.

[Note: in table models with same values are ranked differently as these values are rounded to 2 decimal places]

Sausages

	RW		LR		Expo		ARIMA	
	Val	**Rk**	**Val**	**Rk**	**Val**	**Rk**	**Val**	**Rk**
MAPE	29.57	4	14.32	1	24.85	3	15.17	2
WMAPE	0.35	4	0.14	1	0.27	3	0.14	2
RMSE	29.51	4	9.09	1	21.10	3	9.65	2
MAE	15.67	4	6.07	1	12.18	3	6.01	2
R	0.00	4	0.90	2	-0.17	3	0.91	1
		4		**1.2**		**3**		**1.8**

Table 10 - Sausage (Training Set)

	RW		LR		Expo		ARIMA	
	Val	**Rk**	**Val**	**Rk**	**Val**	**Rk**	**Val**	**Rk**
MAPE	32.34	3	28.79	2	19.11	1	56.49	4
WMAPE	0.39	3	0.25	1	0.25	2	0.54	4
RMSE	30.13	3	12.22	1	21.86	2	30.75	4
MAE	15.35	3	9.70	1	14.90	2	20.90	4
r	0.01	4	0.88	1	0.27	2	0.13	3
		3.2		**1.2**		**1.8**		**3.8**

Table 11 - Sausage (Test Set)

- Linear regression model is more accurate
- Change in rank of models for different metrics is observed
- Exponential smoothing is better than ARIMA and random walk
- Ranking of Linear regression, Exponential smoothing differed on MAPE and WMAPE
- Here Time series models showed low value for *r* (Test Set)

Frozen Vegetables

	RW		LR		Expo		ARIMA	
	Val	**Rk**	**Val**	**Rk**	**Val**	**Rk**	**Val**	**Rk**
MAPE	29.70	4	11.98	1	24.85	3	13.60	2
WMAPE	0.33	4	0.11	1	0.27	3	0.13	2
RMSE	11.95	3	3.32	1	21.10	4	3.93	2
MAE	6.87	4	2.41	1	5.58	3	2.60	2
r	0.02	4	0.92	1	-0.11	3	0.91	2
		3.8		**1**		**3.2**		**2**

Table 12 - Frozen Vegetable (Training Set)

	RW		LR		Expo		ARIMA	
	Val	Rk	Val	Rk	Val	Rk	Val	Rk
MAPE	22.48	3	14.10	1	17.73	2	35.85	4
WMAPE	0.24	3	0.14	1	0.19	2	0.34	4
RMSE	6.95	3	3.92	1	6.12	2	10.17	4
MAE	5.02	3	3.02	1	4.08	2	6.77	4
r	0.09	3	0.91	1	-0.51	2	-0.04	4
		3		**1**		**2**		**4**

Table 13 - Frozen Vegetable (Test Set)

- Linear regression model is accurate than other econometric models
- Exponential smoothing performed better than ARIMA and random walk
- Linear correlation, *r*, value for Time series models low

Frozen Fish Gratin

	RW		LR		Expo		ARIMA	
	Val	Rk	Val	Rk	Val	Rk	Val	Rk
MAPE	24.32	4	8.11	1	19.38	3	10.03	2
WMAPE	0.28	4	0.08	1	0.23	3	0.10	2
RMSE	10.13	4	1.80	1	6.84	3	2.50	2
MAE	4.44	4	1.26	1	3.57	3	1.42	2
r	0.00	4	0.97	1	-0.19	3	0.96	2
		4		**1**		**3**		**2**

Table 14 - Frozen Fish Gratin (Training Set)

	RW		LR		Expo		ARIMA	
	Val	Rk	Val	Rk	Val	Rk	Val	Rk
MAPE	38.41	4	17.47	1	32.79	2	37.16	3
WMAPE	0.44	3	0.15	1	0.39	2	0.45	4
RMSE	15.12	4	3.49	1	10.34	2	12.92	3
MAE	6.94	4	2.38	1	6.02	2	6.64	3
r	0.09	2	0.94	1	0.07	3	-0.05	4
		3.4		**1**		**2.2**		**3.4**

Table 15 - Frozen Fish Gratin (Test Set)

- Linear Regression model is accurate than other econometric models
- Variation in the rank of models on different metrics observed
- Exponential smoothing performed better than ARIMA and Random walk
- Random walk and ARIMA differed in ranking on MAPE and WMAPE
- Linear correlation, *r*, value for Time series models low

Sandwich Ham

	RW		LR		Expo		ARIMA	
	Val	Rk	Val	Rk	Val	Rk	Val	Rk
MAPE	7.91	4	7.15	1	7.72	2	7.85	3
WMAPE	0.07	4	0.06	1	0.07	2	0.07	3
RMSE	16.01	3	14.23	1	15.80	2	20.71	4
MAE	11.65	4	10.51	2	11.35	3	9.69	1
r	0.11	4	0.79	2	0.76	3	0.81	1
		3.8		**1.4**		**2.4**		**2.4**

Table 16 - Sandwich Ham (Training Set)

	RW		LR		Expo		ARIMA	
	Val	Rk	Val	Rk	Val	Rk	Val	Rk
MAPE	11.99	2	10.13	1	21.90	3	24.10	4
WMAPE	0.10	2	0.08	1	0.18	3	0.20	4
RMSE	22.32	2	21.02	1	41.11	3	44.59	4
MAE	15.02	2	11.63	1	26.53	4	25.56	3
r	0.52	2	0.55	1	-0.44	4	-0.51	3
		2		**1**		**3.4**		**3.6**

Table 17 - Sandwich Ham (Test Set)

- Linear regression model is accurate than other econometric models
- Random walk performed better than ARIMA and exponential smoothing
- Linear correlation, *r*, value for Time series models not low in this data set

From the tables 9, 11, 13, 15 and 17 the test data sets of all products, it is seen that the linear regression model is more accurate than other considered econometric models. Exponential smoothing is seen more accurate than random walk and ARIMA model for most of test data set.

Average of average of average of all metrics for each model for each product is given in table 18.

	RW	LR	Expo	ARIMA
Frozen Chicken	3.6	1	2.6	2.8
Sausages	3.2	1.2	1.8	3.8
Frozen Vegetables	3	1	2	4
Frozen Fish Gratin	3.4	1	2.2	3.4
Sandwich Ham	2	1	3.4	3.6
Average	**3.6**	**1.04**	**2.4**	**2.8**

Table 18 – Average of all metrics of all products for each econometrics models

From table 18, it is seen that linear regression model is accurate in average of average of metrics of all products also. Hence from above table it is evident that linear regression is the best considered econometrics model.

6.2 EVALUATION OF PREDICTIVE TECHNIQUES

The values and ranks of error metrics for each considered **predictive models** are tabulated in following tables.

Frozen Chicken

	ANN		Boosting		Bagging	
	Val	Rk	Val	Rk	Val	Rk
MAPE	23.29	3	10.47	2	9.78	1
WMAPE	0.14	3	0.02	1	0.03	2
RMSE	52.76	3	16.70	1	26.68	2
MAE	33.88	3	11.81	1	15.38	2
r	0.98	3	1.00	1	1.00	2
		3		**1.2**		**1.8**

Table 19 - Frozen Chicken (Training Set)

	ANN		Boosting		Bagging	
	Val	Rk	Val	Rk	Val	Rk
MAPE	42.59	1	43.96	2	53.29	3
WMAPE	0.46	1	0.49	2	0.54	3
RMSE	139.83	1	151.11	2	154.53	3
MAE	89.83	1	90.43	2	97.99	3
r	0.95	3	0.97	1	0.96	2
		1.4		**1.8**		**2.8**

Table 20 - Frozen Chicken (Test Set)

- ANN is accurate than Ensemble with Boosting and ensemble with Bagging
- ANN, Boosting and Bagging method differed on MAE and *r*.

Sausages

	ANN		Boosting		Bagging	
	Val	Rk	Val	Rk	Val	Rk
MAPE	10.69	3	5.17	1	5.96	2
WMAPE	0.10	3	0.04	2	0.02	1
RMSE	7.24	3	2.54	1	3.10	2
MAE	4.65	3	1.94	1	2.33	2
R	0.94	3	0.99	1	0.99	2
		3		**1.2**		**1.8**

Table 21 - Sausages (Training Set)

	ANN		Boosting		Bagging	
	Val	**Rk**	**Val**	**Rk**	**Val**	**Rk**
MAPE	22.80	1	31.85	3	27.59	2
WMAPE	0.20	1	0.26	3	0.21	2
RMSE	11.11	1	18.15	3	15.37	2
MAE	9.05	2	10.04	3	8.27	1
R	0.77	2	0.76	3	0.79	1
Average		**1.4**		**3.0**		**1.6**

Table 22 - Sausages (Test Set)

- ANN performed better than both ensembles
- Variation in rank of same model on different metrics is observed.
- Ensemble with bagging is better than ensemble with boosting
- ANN and bagging method differed on RMSE and MAE

Frozen Vegetable

	ANN		Boosting		Bagging	
	Val	**Rk**	**Val**	**Rk**	**Val**	**Rk**
MAPE	10.43	3	6.97	2	6.95	1
WMAPE	0.14	3	0.03	1	0.03	2
RMSE	2.82	3	1.60	1	2.21	2
MAE	2.12	3	1.21	1	1.53	2
R	0.94	3	0.98	1	0.97	2
Average		**3**		**1.20**		**1.80**

Table 23 - Frozen Vegetable (Training Set)

	ANN		Boosting		Bagging	
	Val	**Rk**	**Val**	**Rk**	**Val**	**Rk**
MAPE	14.21	2	11.26	1	14.32	3
WMAPE	0.14	3	0.12	1	0.13	2
RMSE	4.06	3	3.65	1	3.76	2
MAE	3.04	3	2.45	1	2.80	2
R	0.87	3	0.91	1	0.89	2
Average		**2.80**		**1**		**2.20**

Table 24 - Frozen Vegetable (Test Set)

- Ensemble with boosting method is accurate than ANN and ensemble with Bagging method
- Rank of ANN and Ensemble with Bagging varied for MAPE and WMAPE

Frozen Fish Gratin

	ANN		Boosting		Bagging	
	Val	**Rk**	**Val**	**Rk**	**Val**	**Rk**
MAPE	8.04	3	4.53	2	4.19	1
WMAPE	0.08	3	0.02	2	0.02	1
RMSE	2.01	3	0.76	2	0.76	1
MAE	1.26	3	0.62	2	0.59	1
R	0.96	3	0.99	2	0.99	1
Average		**3**		**2**		**1**

Table 25 - Frozen Fish Gratin (Training Set)

	ANN		Boosting		Bagging	
	Val	**Rk**	**Val**	**Rk**	**Val**	**Rk**
MAPE	16.40	1	16.44	2	21.02	3
WMAPE	0.15	1	0.15	2	0.19	3
RMSE	3.24	2	3.11	1	4.25	3
MAE	2.65	2	2.15	1	3.04	3
R	0.97	2	0.97	1	0.92	3
Average		**1.60**		**1.40**		**3**

Table 26 - Frozen Fish Gratin (Test Set)

- Ensemble with boosting is accurate than ANN and ensemble with Bagging
- Change in rank of models varied on different metrics, i.e. rank of NN and ensemble with Boosting varied for WMAPE and RSME.

Sandwich Ham

	ANN		Boosting		Bagging	
	Val	**Rk**	**Val**	**Rk**	**Val**	**Rk**
MAPE	7.64	3	2.91	1	4.32	2
WMAPE	0.07	3	0.01	1	0.02	2
RMSE	15.45	3	6.63	1	9.80	2
MAE	11.15	3	4.84	1	6.71	2
R	0.72	3	0.96	1	0.90	2
Average		**3**		**1**		**2**

Table 27 - Sandwich Ham (Training Set)

	ANN		Boosting		Bagging	
	Val	**Rk**	**Val**	**Rk**	**Val**	**Rk**
MAPE	10.95	1	13.74	2	16.36	3
WMAPE	0.09	1	0.11	2	0.14	3
RMSE	21.50	1	24.89	2	27.54	3
MAE	12.98	1	16.60	2	20.14	3
R	0.53	1	0.36	2	0.34	3
Average		**1**		**2**		**3**

Table 28 - Sandwich Ham (Test Set)

- ANN performed better than both ensembles
- Ensemble with Boosting performed better than ensemble with bagging
- No change in rank of models on any metric

From tables 20, 22, 24, 26 and 28 following are evident:

Ensemble with boosting method is seen more accurate for frozen fish gratin, and frozen vegetables than ANN and ensemble with bagging method. ANN is more accurate for sausages, frozen chicken and sandwich ham. Further considering average of average of all metrics on all products:

	ANN	Boosting	Bagging
Frozen Chicken	1.4	1.8	2.6
Sausages	1.4	3	1.6
Frozen Vegetables	2.8	1	2.2
Frozen Fish Gratin	1.6	1.4	3
Sandwich Ham	1	2	3
Average	**1.64**	**1.84**	**2.48**

Table 29 – Average of average of all metrics of all products for each predictive model

From table 29, it is seen that in average of average of all metrics of all products, ANN is more accurate than ensemble with boosting and ensemble with bagging.

6.3 COMPARISON BETWEEN ANN AND LINEAR REGRESSION

From section 6.1 it is seen that linear regression is best econometric model and from section 6.2 it is seen ANN is best predictive model. Hence further comparison between linear regression and ANN using ranks on test data of each product is done.

Frozen Chicken

	LR		ANN	
	Val	**Rk**	**Val**	**Rk**
MAPE	45.61	2	42.59024	1
WMAPE	0.52	2	0.458169	1
RMSE	160.89	2	139.8346	1
MAE	94.29	2	89.8274	1
R	0.96	1	0.9498	2
Average		**1.8**		**1.2**

Table 30 – Frozen chicken

- ANN is better than linear regression for all metrics except linear correlation, *r*.

Sausages

	LR		ANN	
	Val	Rk	Val	Rk
MAPE	28.79	2	22.80	1
WMAPE	0.25	2	0.20	1
RMSE	12.22	2	11.11	1
MAE	9.70	2	9.05	1
r	0.88	1	0.77	2
Average		**1.8**		**1.2**

Table 31 – Sausages

- ANN is better than linear regression for all metrics expect for linear correlation, *r*.

Frozen Vegetables

	LR		ANN	
	Val	Rk	Val	Rk
MAPE	14.10	1	14.21	2
WMAPE	0.14	1	0.14	1
RMSE	3.92	1	4.06	2
MAE	3.02	1	3.04	2
r	0.91	1	0.87	2
Average		**1.0**		**1.8**

Table 32 – Frozen Vegetables

- Linear Regression is better than ANN on all metrics

Frozen Fish Gratin

	LR		ANN	
	Val	Rk	Val	Rk
MAPE	17.47	2	16.40	1
WMAPE	0.15	2	0.15	1
RMSE	3.49	2	3.24	1
MAE	2.38	1	2.65	2
r	0.94	2	0.97	1
Average		**1.8**		**1.2**

Table 33 – Frozen Fish Gratin

- ANN is better than linear regression on MAPE, WMAPE, RMSE and linear correlation.

- Linear regression is better than ANN on MAE.

Sandwich Ham

	LR		ANN	
	Val	Rk	Val	Rk
MAPE	10.13	1	10.95	2
WMAPE	0.08	1	0.09	2
RMSE	21.02	1	21.50	2
MAE	11.63	1	12.98	2
R	0.55	1	0.53	2
Average		**1.0**		**2.0**

Table 34 – Sandwich Ham

- Linear Regression is better than ANN on all metrics.

Average of Average of metrics for each product:

	Linear Regression	ANN
Frozen Chicken	1.8	1.2
Sausages	1.8	1.2
Frozen Vegetables	1	1.8
Frozen Fish Gratin	1.8	1.2
Sandwich Ham	1	2
Average	**1.48**	**1.48**

Table 35 – Average of Average of all metrics for each product of ANN and Linear Regression

Results

From tables 30, 31, 32, 33 and 34 it is seen that ANN is better than linear regression for three products: frozen chicken, sausages and frozen fish gratin, but not for frozen vegetables and sandwich ham. But average of average of all metrics for all product of both ANN and Linear regression are same, i.e. 1.48 for both, from table 35,. Hence in overall as ANN is more accurate for three product and linear regression for only two and averages of both are same it can be said that ANN is better than linear regression.

6.4 DIFFERENCES IN RANK OF ERROR METRICS

From the result tables in section 6.1 and 6.2 it is seen that the rank of a model differed on different metrics. So, to see how many times rank between any two metrics for a model differed, number of differences between any two metrics for all models is calculated and presented in the percentage form in table 36.

	MAPE	WMAPE	RMSE	MAE	R
MAPE	0				
WMAPE	36%	0			
RMSE	34%	30%	0		
MAE	39%	30%	26%	0	
r	59%	60%	51%	44%	0

Table 36 – Percentage difference between metrics

From table 36,

- Difference between MAPE and WMAPE is 36% in total
- Difference between MAPE and RMSE is 34%
- Difference between RMSE and MAE is 26% and so on.

7. ANALYSIS AND DISCUSSION

From the section 6, it is seen that the accuracy of ANN is better than ensemble and econometric models. ANN performed very well on the test data set. Ensemble's accuracy was very good on the training set, but it showed its drawback that of overfitting data on the test set. Its accuracy degraded on the test data set. Linear regression was accurate than other econometric models.

It is evident from the chapter "theoretical background" that the smoothing techniques and ARIMA do not consider the relationship between the explanatory and dependent variables. They consider only the relationship between the present value and the past values of variable. ARIMA tries to find optimal lag structured automatically. But the pattern of data considered for the experiment showed a strong relationship between commercials and sales. So, it is not surprising that they performed so badly. Hence it can be concluded that ARIMA and smoothing techniques are not suitable for the data where cause-and-effect relationship between variables are strongly evident and forecast should be based on that relationship. So, no further evaluation of random walk, exponential smoothing and ARIMA is done.

Referring to the section 3.8 it is evident that ANN is accurate than econometric model. Yoon and Swales (1991); Kudyba S. (1998) have concluded that regression model is less accurate than ANN but the regression model has capability of forecasting and explaining the results. Other similar studies conducted by Gruca T.S., Klemze B.R., Petersen E.A.F. (1999) using MAPE; Camargo M.E., et al (2009); Moshiri S., and Cameron N.(2000) using RMSE and MAE metrics have shown that: "ANNs are more accurate than econometric models". These findings correlates with the result of the experiment performed in this thesis.

From tables 18 and 29, it is evident that linear regression is the best considered econometric model and ANN is the best considered predictive model. Hence, further comparison between linear Regression and ANN is done. The value of each metric for both models for all products is considered and the difference between them is calculated. The following table 37 (for the test data set) shows value of each metric and the differences between them for both models for each product:

Test Set															
	Frozen Chicken			**Sausage**			**Frozen Vegetable**			**Frozen Fish Gratin**			**Sandwich Ham**			
Model Metric	**NN**	**LR**	**Diff**	**NN**	**LR**	**Diff**	**NN**	**LR**	**Diff**	**NN**	**LR**	**Diff**	**NN**	**LR**	**Diff**	**Avg**
MAPE	42.59	45.61	**3.02**	22.78	28.79	**6.01**	14.20	14.10	**0.10**	16.40	17.47	**1.08**	10.95	10.13	**0.82**	**2.20**
WMAPE	0.46	0.52	**0.06**	0.21	0.25	**0.04**	0.15	0.14	**0.01**	0.15	0.15	**0.00**	0.09	0.08	**0.01**	**0.02**
RMSE	139.83	160.89	**21.06**	11.43	12.22	**0.79**	4.21	3.92	**0.29**	3.24	3.49	**0.24**	21.50	21.02	**0.48**	**4.57**
MAE	89.83	94.29	**4.47**	9.05	9.70	**0.65**	3.04	3.02	**0.02**	2.65	2.38	**0.27**	12.98	11.63	**1.34**	**1.35**
RMSE	0.95	0.96	**0.01**	0.77	0.88	**0.10**	0.87	0.91	**0.04**	0.97	0.94	**0.02**	0.53	0.55	**0.02**	**0.04**

Table 37 - ANN versus Linear Regression (Test Set)

From table 37 it is seen that the differences between ANN and the linear regression models are quite small. A noticeable difference seen is only for the product Frozen chicken on RMSE. In most of other observations, differences were very small. Also the averages of differences were not large. This analysis suggests that the performance of linear regression is not bad compare to ANN. So, linear regression can be used as alternative to ANN. Several authors also have mentioned that linear regression has advantage of both forecasting and expressiveness over ANN. This further supports the use of linear regression as alternative to ANN. The interpretation of linear regression model used in experiment mentioned in appendix B further supports the expressiveness of regression model.

The following figures 9 and 10 shows the forecasted values produced by linear regression and ANN for test set for frozen chicken. The effect of commercial is also shown in these diagrams:

Linear Regression (Test Set)

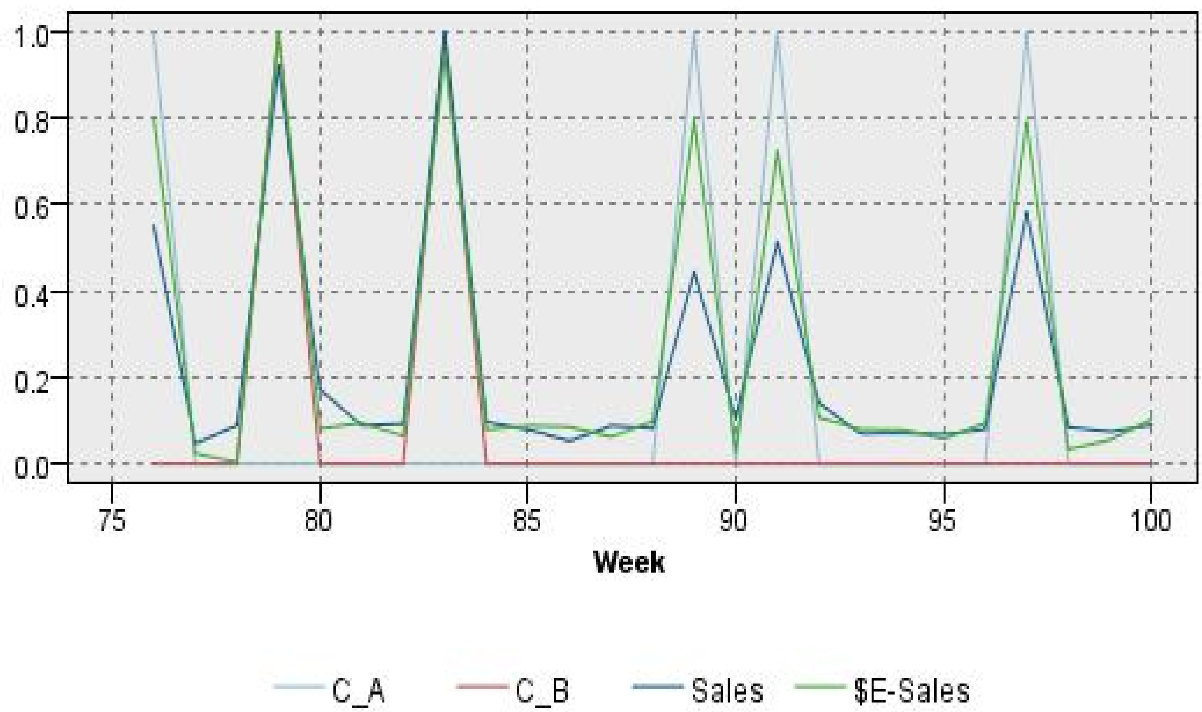


Figure 9 – Commercials, Sales, Estimated sales, on Test Data

ANN (Test Set)

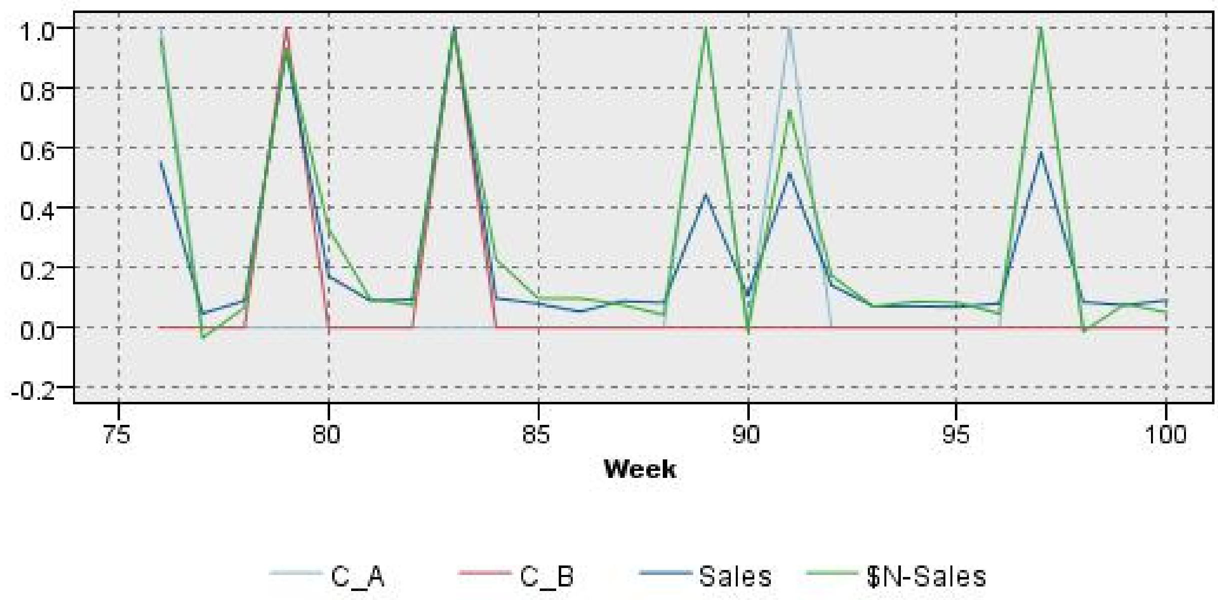


Figure 10 - Commercials, Sales, Estimated sales, on Test Data

The ranking of models done in this thesis helped to observe if the rank of same model on same data set differs according of different metrics or not. From the section 6.1, it is seen that the rank of same model on same data set differed on different metrics. It implies that one metric may show small error but another metric may show big error for the same model on the same data set. From the section 3.7 and 3.8 it is known that the accuracy of error metrics is highly influenced by nature of data. This suggests that the use of one metric for the evaluation of forecasting techniques is not suitable and it correlates with the findings of Armstrong (1985).

Further from table 35, follow can be derived:

- Number of differences between the ranks of any two metrics was high which suggests that the evaluation of models based on several metrics give more information about the model and thus facilitates more information on the choice of model.
- *r* stands out from other metrics and differs twice as much in rank. This may be due to that it is the only metric which doesn't consider the error size.
- Related metrics like MAPE and WMAPE or RMSE and MAE are not seen to be too much similar with each other. This may be due to the properties of datasets.

From theoretical background section it is known that WMAPE can be used to get a unskewed comprehensible measure of the error and linear correlation can be used to evaluate how much of the fluctuations of the sales that actually can be explained by the forecast model. Different metrics often ranks the models very differently, so the use of WMAPE and linear correlation can provide more information for choosing which forecast model to use for a decision maker.

8. CONCLUSION

Experiment**s** performed in this thesis showed that ANN is superior compare to all other considered models when evaluated on MAPE, WMAPE, RMSE, MAE, and linear correlation. But the accuracy of the linear regression model (the best econometric model among considered models) should not be neglected since the difference in accuracy of ANN and it was very small. It could be argued that the difference in performance is not large enough to motive the use of an opaque model such as the ANN model. If comprehensibility is considered linear regression is a much more appropriate choice for sales forecasting. ARIMA and other considered smoothing techniques are not suitable for the data where cause-and-effect relationships between variables are strongly evident.

The models sometimes were ranked differently by different error metrics and the number of differences in percentage of ranks between any two metrics was found high. Obviously, the different metrics measure different properties of the forecast and thus contribute to a more thorough evaluation of the model.

9. SUGGESTION AND FUTURE WORK

The pattern of data used in this study may have differed from data used in other studies. This may be reason for having different result in this study than others conducted where ANN was shown to outperform econometric models substantially. So, understanding pattern of data is necessary.

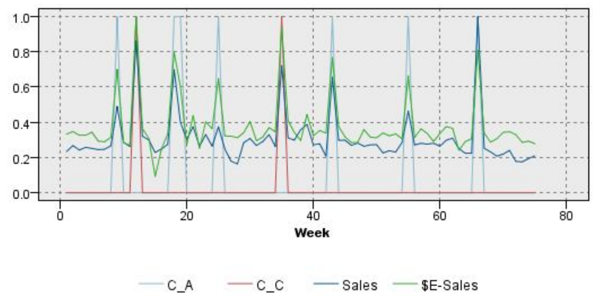
The major difference between this research and researches mentioned in the related work is the number of evaluation metrics used. In this thesis five different metrics are used to compare forecasting models. From the result of experiment, it is evident that the values of different metrics for same model on same data set differ. This suggests that more metrics should be used to evaluate performance of forecasting model and this correlates with Armstrong's (1985) findings.

In this research only one linear regression model was used. The result of that model was compared with results of neural network, ensemble, exponential smoothing, random walk and ARIMA (1, 1, 1) models. Actually several different regression models can be built using economic theories and experience of econometrician. The performance of each model may vary. It is possible to have better regression model than one considered in this experiment. The use of more than one regression model and their comparison with predictive model can be area of further study. Also there are many predictive models that could have been used. Another interesting comparison could have been towards a regression tree since these also are comprehensible.

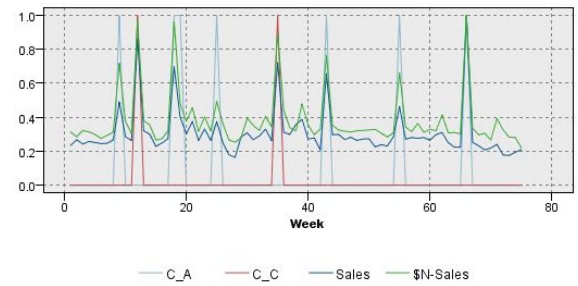
Appendix A

Sausage

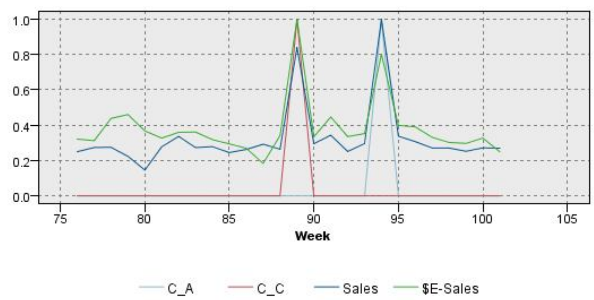
Linear Regression Training Set



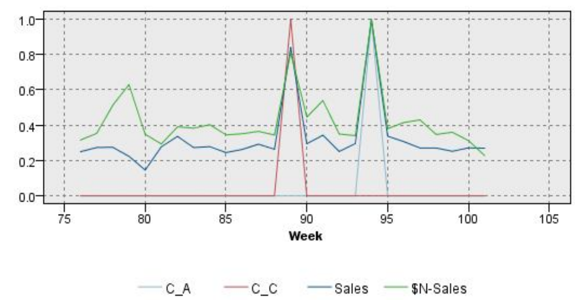
ANN Training Set



Linear Regression Test Set

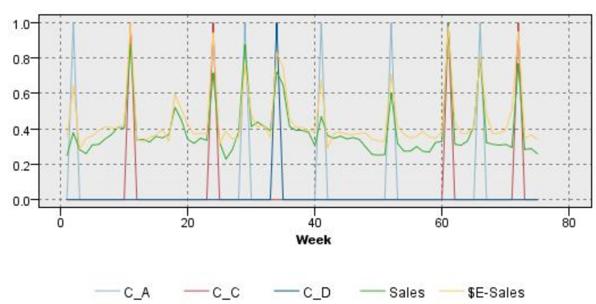


ANN Test set

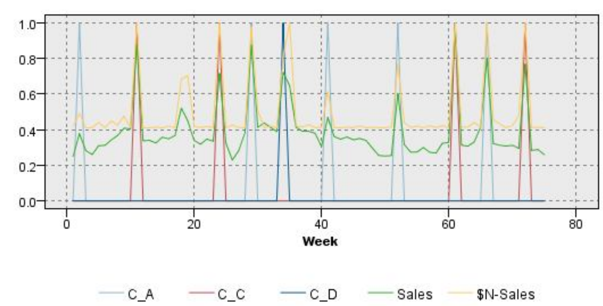


Frozen Vegetables

Linear Regression Training Set

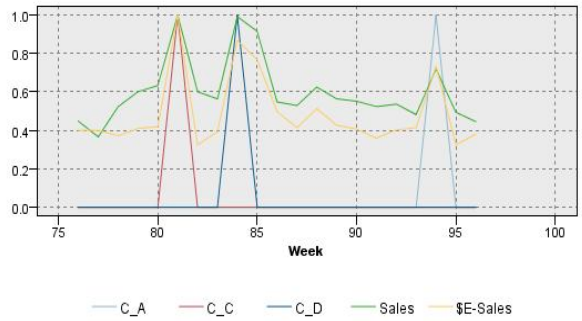


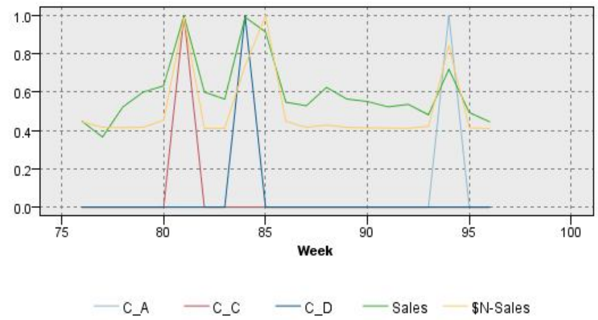
ANN Training Set



Linear Regression Test set

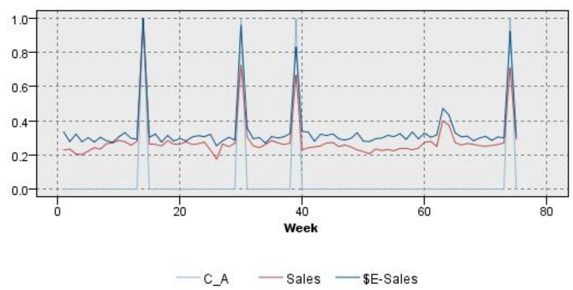
ANN Test Set



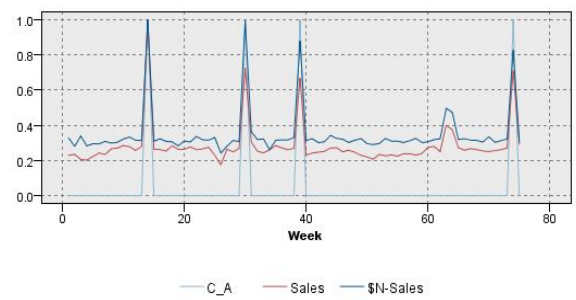


Froze fish Gratin

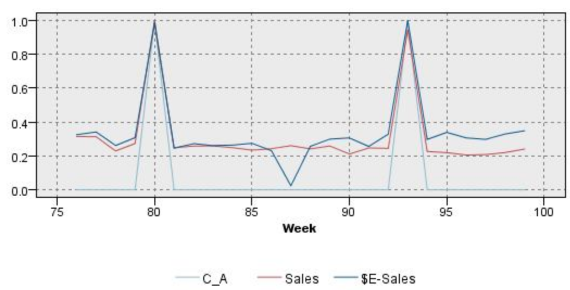
Linear Regression Training Set



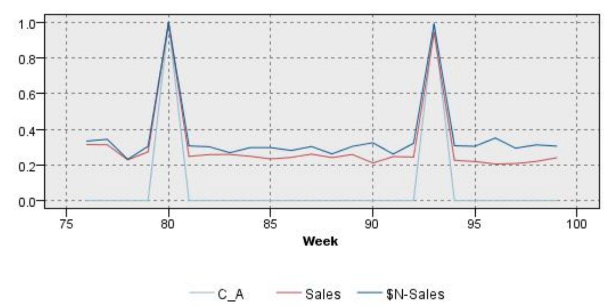
ANN Training Set



Linear Regression Test Set



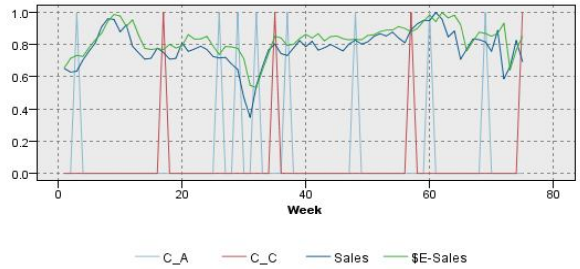
ANN Test Set

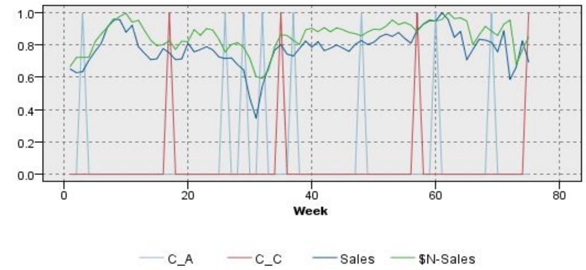


Sandwich Ham

Linear Regression Training Set

ANN Training Set





Linear Regression Test Set

ANN Test Set

Appendix B

Interpretation of Regression Model

The linear regression models have expressive ability that is the models can be described using the estimated parameters of the model. Hence to support this statement hypothetical interpretation (as no economic theory is used to build the model) of the model built by the PASW modeler is presented. The regression model for **Frozen fish gratin** built by PASW modeler 14 is considered here and interpreted as following:

$$Sales = \beta_0 + \beta_1 * CA + \beta_2 * CA\_1 + \beta_3 * Salary + \beta_4 * Salary\_1 + \beta_5 * CS + \beta_6 * CS\_1 + \beta_7 * PI + \beta_0 * Sales\_1$$

Where,

CA = Commercial A
CA_1 = One period lag value of CA
Salary_1 = One period lag value of Salary
CS = ChildSupport
CS_1 = One period lag value of ChildSupport
PI = PriceIndex
Sales_1 = One period lag value of Sales
$\beta_0, \ldots, \beta_8$ are parameters of model to be estimated

The estimated values for the parameters and estimated equation were:

$$sales = 183.6 + CA * 19.64 + CA_1 * -2.405 + Salary * 1.425 + salary\_1 * \ -0.8389 + CS * -1.373 + CS\_1 * -1.587 + PI * -158.3 + Sales_1 * 0.07146$$

The above regression model can be interpreted with the help of economical theory. It must be noted that "the model parameters should be interpreted only within the sampled range of the independent variable" (McClave, Benson, Sincich, 1998). Here hypothetical interpretation is made about the model to show expressiveness of regression model, as no economic theory is used to develop regression model.

- The intercept, β_0 = 183.6, this implies that the average sales of product frozen fish gratin will be 183.6 when all of the explanatory variables are zero.
- The relationship between sales and Commercial A is shown by CA = 19.64. This number shows that for each commercial the sales of product frozen fish gratin increase by 19.6 units. (all other explanatory variables held constant)
- The parameter CA_1 = -2.405 shows that the commercial in previous week does not have affect on the current week's sale. (all other explanatory variables held constant)
- The parameter $Salary$ = 1.425 implies that the salary distribution in current week increases the sales of product frozen fish gratin by 1.425 units. (all other explanatory variables held constant)
- The parameter $salary\_1$= -0.8389 shows that the salary distributed in previous week does not have affect on the current week's sale. There is decrease in sales by 0.8389 units. (all other explanatory variables held constant)
- The parameter CS= -1.373 implies that the childsupport provided in current week does not increase sales. It shows that the sales of product frozen fish gratin decrease by 1.373 units. (all other explanatory variables held constant)
- The parameter CS_1= -1.587 shows that the distribution of childsupport in previos week does not have affect on sales of current week. It shows decrease in the sales of product frozen fish gratin by 1.587 unit. (all other explanatory variables held constant)
- The parameter PI = -158.3 shows that the given value of priceindex decreases sales of product frozen fish gratin by 158.35 unit. (all other explanatory variables held constant)
- The parameter $Sales_1$ = 0.07146 implies that previous sales figure has affect on the current week's sales. Previous sales increases in sales of product frozen fish gratin by 0.07146 units. (all other explanatory variables held constant)

Appendix C

1. Creating stream in PASW modeler for each modeling technique. In new stream of PASW modeler:
    a. Data source is specified using Source Node.
        - Source: Excel file
        - Excel worksheet of first product (frozen chicken) is specified
        - Role of variables are specified. Sales as target variable and commercials, salary, PriceIndex, Childindex are input variables (explanatory)
        - Values for all variables are read
    b. Data selection criterion using Select node is specified. Select node is connected with the source node. Data selection criterion is mentioned as: **Type=”TRAIN”** and this node is named as **Training**
    c. Unnecessary variable are filtered out using **Filter** Node. Here Type variable is unnecessary
    d. Modeling technique is chosen. Chosen modeling technique is connected with the filter node.
    e. The modeling technique is executed. This generated **model** and is depicted by model node in the stream
    f. **Graph** node (**Multiplot**) is connected to the model node, X-axis and Y-axis are specified for the graph.
    g. **Table** and **analysis** nodes are connected to the model node.
    h. New Data Selection criterion is specified again using another Select node for TEST data set. This new select node is connected with the source node. Data selection criterion is mentioned as: **Type=”TEST”** and this node is named as **TEST**.
    i. Unnecessary variable are filtered out using **Filter** Node. Here Type variable is unnecessary.
    j. The model built in the step (f) is copied and connected with new filter node, i.e., filter node for the test data set.
    k. **Graph** node (**Multiplot**) is connected to the model node of the test data set by copying previous graph node connected to model node of the training data set
    l. **Table** and **Analysis** nodes are connected to model for TEST data set.

REFERENCES

2T Time Series, 2010, *Box-Jenkins Time Series Analysis*, 83th annual meeting of the Transportation Research Board. CDROM. Transportation

Apte C. et al, 2002, *Data Intensive Analytics for Predictive Modeling*, T.J. Watson Research Center, IBM Research Division

Armstrong J.S, 1985, *Principles of forecasting* A Handbook for Researchers and Practitioners,

Arsham H., 2010, *Forecasting by Smoothing Techniques*, Available at: http://home.ubalt.edu/ntsbarsh/Business-stat/otherapplets/ForecaSmo.htm; [Accessed 1 Dec 2010]

Burger et al., 2001, *A practitioners guide to a time-series methods for tourism demand forecasting – a case study of Durban, South Africa*, Tourism Management 2, 402-409

Camargo M. E., et. al, 2009, *Performance of the two approaches forecasting model of sales*, IX Congreso Galego de Estatística e Investigación de Operacións

Cho, V., 2003. *A comparison of three different approaches to tourist arrival forecasting*. Tourism Manage, Volume 24, Issue 3, June 2003, Pages 323-330

Decision 411 forecasting, 2010, *What's the bottom line? How to compare models*, Available at: file:///D:/Informatics/thesis/in%20KTM/error%20metrics/How%20to%20compare%20models.htm , [Accessed: 23 jan 2011]

Dunham M.H., 2003, *Data Mining Introductory and Advanced Topics*, Pearson Education

Friedman J.H., and Popescu B.E. (2005), *Predictive Learning via Rule Ensembles*, Available at: http://www-stat.stanford.edu/~jhf/ftp/RuleFit.pdf, accessed: 13 Dec 2010.

Gruca, Thomas S, Bruce R Kelmz and E. Ann Furr Petersen. *Mining Sales Data using a Neural Network Model of Market Response*, ACM, SIGKDD, Explorations Newsletter Vol.1, issue.1, June 1999, 39-43.

Han J., Kamber M., 2005, *Data Mining: Concepts and Techniques*, 2nd ed. Morgan Kauffman

Hirchey M., 2009, *Fundamentals of Managerial Economics*, 9th ed., South-Western

Hopkins W.G., (2000), *Quantitative Research Design*, Sportscience, Available at :http://www.sportsci.org/jour/0001/wghdesign.html [Accessed 20 dec 2010]

Horvath Z., Johnston R., 2006, *AR(1) Time Series Process Econometrics 7590*,

Howrey E.P., 1980, *The role of time series Analysis in Econometric Model evaluation*, Evaluation of Econometric Models, Academic Press

Idri A., Khoshgoftaar T. M., Abran A., 2002, *Can Neural Networks be easily Interpreted in Software Cost Estimation?,* World Congress on Computational Intelligence, Honolulu, Hawaii, May 12-17.Investigate the Frequency and Severity of Road Traffic Accidents. In Proceedings of the

Kahforoushan E., Zarif M., Mashahir E. B., 2010, *Prediction of added value of agricultural subsections using artificial neural networks: Box-Jenkins and Holt-Winters methods*, Journal of Development and Agricultural Economics Vol. 2(4), pp 1150121, April 2010.

Kalekar P.S., 2004, *Time series Forecasting using Holt-Winters Exponential Smoothing*, Kanwal Rehik School of Information Technology

Keller G., Warrack B., 1999, *Statistics for management and Economics*, 5th ed, Duxbury

Kenny J.J., 2002, *Statistics for Science and Engineering*, Pearson Education Inc

Kline D. K., & Kohers G., (2000), *Exploring the Need for Node Biases in Time Series Forecasting with Artificial Neural Networks*, Working Paper

Kmenta J., Ramsey J.B., (1980), *Evaluation of Econometric Models*, Academic Press, Available at: http://www.nber.org/books/kmen80-1

Kon S. C., and Turner W. L. (2005). *Neural network forecasting of tourism demand,* Tourism Economics, 11, 301-328

Kourentzes N, Crone S.F., 2008, *Forecasting High-and Low-Frequency Data with Neural Networks Challenges in Modeling the Input Vector*, DMIN'08, the 4th International conference on Data Mining

Kudyba S., 1998, *Are neural networks a better forecaster?, Technology & Trading*, Available at: http://www.marketdoctor.com/Docs/FutartNN.pdf, [accessed: 17 dec 2010].

Liu C., Wu K., Tsa, M, (2005), *Energy Efficient Information collection With THE ARIMA MODEL IN WIRELESS SENSOR*, *GLOBECOM '05* 2005, 2470-2474

MathsBit.com, 2010, *correlation coefficient*, Available at: http://mathbits.com/mathbits/tisection/statistics2/correlation.htm [Accessed 20 dec 2010]

Maier H.R., Dandy G.C., 1999, *Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications*, Department of Civil and Environment Engineering, The University of Adelaide, Adelaide, S.A., 5005 Australia

McClave, Benson, Sincich, 1998, *Statistics for Business and Economics*, seventh edition, Prentice-Hall, Inc,

Meyler A., Kenny G., Quinn T., 1998, *Forecasting Irish Inflation Using ARIMA models*, Technical paper

Moshiri S., Cameron N., 1999, *Neural Network Versus Econometric Models in Forecasting Inflation*, Journal of Forecasting, J. Forecast. 19, 201-207 (2000)

Opitz D., Maclin R., (1999), *Popular Ensemble Methods: An Empirical Study*, Journal of Artificial Intelligence Research 11 (1999) 169-198

Ostafe D., n.d., *Neural Network Hidden layer Number Determination Using Pattern Recognition Techniques*.

Palmer A., Jose Montano, J. J., and Sese, A. (2006). *Designing an artificial neural network for forecasting tourism time series*. Tourism Management, 27, 781-790

Pindyck R.S., Rubinfeld D.L., 1991, *Econometrics Models & Economic Forecasts*, 3rd ed., McGraw-Hill, Inc

Pope R.D., et. al., 1979, *An Evaluation of Econometric Models of U.S. Farmland Prices*, Western Journal of Agricultural Economics

Rao K.V. G, Chand P.P, Murthy M.V.R (2007), *Soft Computing-Neural Networks Ensembles*, Journal of Theoretical and Applied Information Technology

Real Options Valuation, 2007, *Advanced Forecasting Techniques and Models: ARIMA*, Available at: www.realoptionsvaluatition.com

Rosin P.L, Fierens F., 1995, *Improving Neural Network Generalization*. In Proceedings of

IGARSS’95, Firenze, Italy, July.

Schutz W., Kolassa S., 2010, Foresight: Advantages of MAD/Mean over Ratio over the MAPE, Available at:www.saf-ag.com, [Accessed: 23 Jan 2011]

Shachmurove Y., (2005) *Business Applications of Emulative Neural Networks*, Internal Journal of Business, 10(4), 2005

Shu C., Burn D.H., 2003, Artificial Neural Network ensembles and their application in pooled flood frequence analysis, Department of Civil engineering, University of Waterloo.

Song H. and Li, G. (2008), *Tourism demand modeling and forecasting – a review of recent research*, Tourism Management, Vol 29, No 2, pp 203-220.

STL Warehousing, (2010), *Forecasting*, Available at: http://www.stlwarehousing.com.au/info/ST_Forecasting.pdf, [Accessed 7 nov 2010]

Sun MicroSystem, 2010, *PASW modeler 14*,

The Decision Makers’ Direct, 2001, *Choosing the Right forecasting Technique*, Available at: http://www.decisioncraft.com/dmdirect/forecastingtechnique.htm, [Accessed: 2 dec 2010]

Tjung L. C., Kwon O., Tseng K.C., Bradley-Geist J., 2010, *Forecasting Financial Stocks using Data Mining*, Global Economy and Finance Journal, Volume 3. Number 2. September 2010. Pp. 13-26

Verbeek M, 2008, *A Guide to Modern Econometrics*, 3rd ed., John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England

Watson P.K., and Teelucksing S.S. (2002), *A Practical Introduction to Econometric Methods: Classical and Modern*, University of the West Indies Press,

Wooldridge, 2006, *Introductory Econometrics A Modern Approach*, Thomson South-western

Yoon, Swales G., 1991, *Predicting Stock Price Performance: A Neural Network Appoarch*, Southwest Missouri State University, Springfield, Missouri 65804-0095

Zellner A., Palm F.C., 2004, *The structural econometric time series analysis approach*, Cambridge University Press, page 44-48

Zeng T., Swanson N.R., 1998, *Predictive Evaluation of Econometric Forecasting Models in Commodity Futures Markets*, Aeltus Investment management, Inc., 242 Trumbull street, ALT6, Penn State University, 521 Kern Graduate Bldg, Department of Economics, University Part,

ZHU Y., 2005, *Ensemble Forecast: A New Approach to Uncertainty and Predictability*, Advances in atmospheric sciences, Vol. 22, No. 6, 2005, 781-788

Žilinskas and Žilinskas, 2010, *On Global Optimization in Least Squares NonLinear Regression by Means of Interal Methods*, 24th Mini Euro conference

2T Time Sereis, 2010, *Box-Jenkins Time Series Analysis*, 83th annual meeting of the Transportation Research Board. CDROM. Transportation

University of Borås is a modern university in the city center. We give courses in business administration and informatics, library and information science, fashion and textiles, behavioral sciences and teacher education, engineering and health sciences.

In the **School of Business and Informatics (IDA),** we have focused on the students' future needs. Therefore we have created programs in which employability is a key word. Subject integration and contextualization are other important concepts. The department has a closeness, both between students and teachers as well as between industry and education.

Our **courses in business administration** give students the opportunity to learn more about different businesses and governments and how governance and organization of these activities take place. They may also learn about society development and organizations' adaptation to the outside world. They have the opportunity to improve their ability to analyze, develop and control activities, whether they want to engage in auditing, management or marketing.

Among our **IT courses**, there's always something for those who want to design the future of IT-based communications, analyze the needs and demands on organizations' information to design their content structures, integrating IT and business development, developing their ability to analyze and design business processes or focus on programming and development of good use of IT in enterprises and organizations.

The **research** in the school is well recognized and oriented towards professionalism as well as design and development. The overall research profile is Business-IT-Services which combine knowledge and skills in informatics as well as in business administration. The research is profession-oriented, which is reflected in the research, in many cases conducted on action research-based grounds, with businesses and government organizations at local, national and international arenas. The research design and professional orientation is manifested also in InnovationLab, which is the department's and university's unit for research-supporting system development.

UNIVERSITY OF BORÅS

SCHOOL OF BUSINESS AND INFORMATICS

VISITING ADDRESS: JÄRNVÄGSGATAN 5 · POSTAL ADDRESS: ALLÉGATAN 1, SE-501 90 BORÅS

PHONE: + 46 33 435 40 00 · E-MAIL: INST.IDA@HB.SE · WEB: WWW.HB.SE/IDA