QUANTITATIVE APPROACH TO FORECASTING USING "R"

SAMBŌDHI

QUANTITATIVE APPROACH TO FORECASTING USING "R"

Dutt, V.[1], Mohan, V.,[2] and Raj, N.[3]

[1] Research Manager, Sambodhi Research & Communications, varun.dutt@sambodhi.co.in
[2] Research Manager, Sambodhi Research & Communications, varun@sambodhi.co.in
[3] Research Manager, Sambodhi Research & Communications, nikhil.raj@sambodhi.co.in

ABSTRACT

Forecasting has been playing a major role in various spheres of life and has assumed significant importance during recent years. It has found its applications in varied domains such as Economic Forecasting, Technology Forecasting, Weather Forecasting, Sales Forecasting, Demand Forecasting etc.

The paper attempts to understand various Quantitative Forecasting Techniques and explores reasons for using a particular method in a given study scenario. The paper also describes the applications of different methods and lists down the processes of applying those techniques using Objective Approach of Time Series Models through "R Statistics Software". Though the paper tries to answer such questions on theoretical grounds, it also attempts at providing practitioner's point of view so as to enable applications of Time Series Forecasting Techniques using Exponential Smoothing Methods in practice.

Keywords: Forecasting, Time series models, R

INTRODUCTION

Forecasting can be defined as a planning tool used to predict, project or estimate some future activity, event or occurrence. Forecasting is a useful tool to estimate the outcomes, plan for resource requirement and is also used as an integral part of the performance management process. (SocialSolutions,-). Forecasting has found application along wide range of domains. Economic forecasting, sales forecasting, weather forecasting, yield forecasting etc. are the different types of forecasting which are widely practiced.

Forecasting can be done either by using qualitative or quantitative approaches taking into account the forecast horizons-short, intermediate or long. Qualitative Forecasting Methods are based on judgements, opinions, intuitions, emotions, or personal experiences and are subjective in nature. They do not rely on any rigorous mathematical computations. There are primarily four key types of qualitative forecasting methods. First, executive opinion method in which a group of managers meet and collectively develop a forecast. Second, market survey approach which uses surveys and interviews to judge the preference of customers and assess the demand. Third, sales forecasting composite, mainly used in sales and operations, in which each sales person is asked to estimate sales in each of his or her region. And lastly, Delphi method in which a group of experts are asked to respond to a series of questionnaires. Based on the compilation of their responses, a forecast is made.

On the other hand, Quantitative Methods are based on mathematical models and are objective in nature. They heavily rely on mathematical computations. The two quantitative methods that are frequently used are classified as associative models and time series modelling. The first method i.e. Associative Models, assumes that there are cause-effect relationships between the variable forecasted and the other variables in the environment. The forecast projections are made in this model based on these associations. The second method i.e. Time Series Modelling looks at the past trends and patterns and based on the trends, cyclicity, seasonality, outliers, residuals and autocorrelation to make projections.

FORECASTING IN DEVELOPMENT INTERVENTIONS

Forecasting has a very important role to play, specifically in context to development interventions. It can buttress the project planning, reducing the uncertainty and setting of reasonable outcome targets for the projects. Estimating the uncertainty through forecasting is desirable in public welfare projects which are targeted towards disadvantaged groups who would be more vulnerable. (Anderson, 1989). Forecasting, if done accurately, can be a vital information source which can

contribute to more effective monitoring and also evaluations of the project. Forecasting of the expected outcome indicators associated with the project at various project stages as per the previous trends, can help the project management to fix accurate and reasonable targets that should be expected from the intervention.

Forecasting has been used in social welfare projects targeted at domains like public health, poverty reduction by mitigating agriculture production risk, enhancing livelihood opportunities and employment generation, education etc. There are various tools and techniques which are used for forecasting at different stages in social welfare projects. Marginal Budgeting of Bottlenecks (MBB), a tool developed by United Nations International Children Emergency Fund (UNICEF) and World Bank, is useful in estimating the costs of removing the constraints or bottlenecks of the existing system and that of scaling up of coverage of proven high impact interventions (Devinfolive, 2011) . It has been widely used in public health projects focused towards intervention areas like nutrition and diseases like malaria, tuberculosis, HIV/AIDS etc. It is used to plan and forecast the potential costs, henceforth budgets and also in assessing the impact of the scaling up investments which aim at reducing health system constrains.

The conceptual idea behind MBB is to identify the sector and system level bottlenecks faced in delivering the services at various levels like community, frontline health workers and health facility level. These bottlenecks have been identified and have further been categorized into three demand side and three supply side barriers. These six barriers or determinants can help in examining the level of system functionality at a given point of time. By assessing the bottleneck patterns across various geographical intervention areas, the decision makers can adjust the delivery modes i.e. the mix of the intervention while taking into account the context in which the services are delivered. Further, based on the targeted incremental increase in the outcome indicators, the cost of overcoming the bottlenecks of each of the above six factors would be calculated, which would essentially help to assess the required budget. The bottleneck analysis that is used by MBB has also been included in the Unified Health Model (UHM), another planning and forecasting tool, which has been developed by Inter Agency working group comprising institutions like World Health Organization (WHO), United Nations Population Fund (UNFPA), United Nations Programme on HIV and AIDS, World Bank, United States Agency for Indenational Development (USAID), and UNICEF.

To forecast the expected outcome level indicators from development programmes and policies, computer based models have also been developed. These models use

historical and existing data to project future trends and patterns of the chosen outcome indicators based on the inputs provided in the programme or the projects (USAID, 2009). Spectrum Policy Modelling system is such a windows based integrated package software. It can be used to project the need of services across various domains like family planning, reproductive health, and maternal health HIV/AIDS services. It has various components like DemProj, FamPlan, AIM, RAPID, Lives Saved Tool (LIST) etc. aimed towards the above mentioned domains (Futures Group, n.d.). LIST (Lives Saved Tool) has been widely used for forecasting the impact of projects, specifically aimed to improve maternal, child and neonatal health. It is an evidence-based tool which is used to estimate the amount of mortality reductions that can be achieved increasing the coverage level of specific interventions. This estimation is done while keeping in mind the current demographic characteristics and also the initial and planned coverage of the intervention. The impact calculations of LIST are derived by using the childhood and neonatal morbidity and mortality data generated by Child Health Epidemiology Reference Group (CHERG). CHERG was established by WHO with a mandate of reviewing and improving the existing data collection methods for better estimation of cause specific distribution of child deaths (Devinfolive, 2011).

Projection tools like LIST when used in combination with evidence based planning, forecasting and budgeting tools like MBB or UHM, can be used by decision makers like national programme managers, Ministries of Health and Finance and other development sector agencies to engage in better project or programme planning by 1) prioritizing resource investments 2) deciding the mix of project area specific interventions and 3) forecasting the expected impact of planned interventions based on the inputs provided.

Similarly, forecasting of weather has also been widely done and avidly used across all sections of society, specifically by the farmer community. Before time intimation of expected adverse weather helps to plan in advance to mitigate the effect of extreme weather conditions and can be helpful to farmers in safeguarding their crops. Climate based strategic agronomic planning can be very useful for the farmers by decreasing their vulnerability to adverse climatic situations and incidents of pests and diseases (Das et al., n.d.). Crop or yield forecasting is also a technique which can be useful to farmers by which the crop yield can be predicted even couple of months before harvest. Crop forecasting is done based on models which attempts to stimulate the plant- weather-soil interactions.

Forecasting can also play an important role in providing inputs for better planning and further buttressing environmental conservation projects. For example- forecasting the dynamics of forest biodiversity in relation to the global climate

change can be helpful to the forest policy makers to increase their decision making capability based on the multidimensional factors. This can be further helpful in facilitating collaboration and cooperation between and within countries. Forecasting has also been used innovatively to predict the potential impact of urbanization on our natural resources and environment. Land use change models and webs based environment impact models have are been used for such forecasting studies. The information available through this type of forecasting can have significant implications on urban planning and descions making in an effort to conserve the natural resources of the impact area (Tanga, 2005). There are also powerful computational forecasting tools, which use long term monitoring data to develop early warning systems for predicting catastrophic shifts in ecosystems (Muys et al., 2011).

APPLICATIONS OF FORECASTING IN MONITORING & EVALUATION

Change caused by an intervention in the intervention target region is detected by studying variance exhibited over time in variables of interest. Appropriate techniques of forecasting can assist an evaluator in determining where the variable may be in the future had the intervention not taken place. Forecasting in intervention area for endline using pre-baseline data will provide apposite estimate of trends that would have been witnessed in that area, without taking into account the intervention factor. This forecasted value can then be used as a reference for ascertaining change. The forecasted data thus acts as a pseudo-counterfactual, and serves as a tool for valid attribution.

This forecasted value can be used for attributing change in cases where a suitable counterfactual doesn't exist, moreover even in cases where an appropriate counterfactual does exist the forecasted data can be used for validating results obtained from comparing project and counterfactual. Thus, Forecasting Models may be built in order to carry out Monitoring & Evaluation of projects. The models might provide useful insights in determining when the evaluation activities should commence.

QUANTITATIVE TECHNIQUES OF FORECASTING

As discussed earlier, there are two commonly used quantitative techniques of forecasting. The *Associative or Casual Model* assumes that the dependent variable (variable to be forecasted) has a cause and effect relationship with one or more of the independent variables. They forecast using either Simple Linear Regression, Multiple Linear Regression, Non-linear Regression. *Time Series Model* looks at the past Trend, Cyclicity, Seasonality and Random Variations in the data and based on

these the forecasts are made using either Smoothing Models or Decomposition Models.

Associative/Casual Model	▪ Simple Linear Regression ▪ Multiple Linear Regression ▪ Non-linear Regression
Time Series Model	▪ Smoothing Models (Simple & Weighted Moving Average, Exponential Smoothing) ▪ Decomposition Models(Additive Models, Multiplicative Models)

The Decomposition Models may either be Additive or Multiplicative in nature. A Time Series Data is described as a function of four components:

$$Y= T*C*S*I \text{ (Multiplicative Model)}$$
$$Y=T+C+S+I \text{ (Additive Model)}$$

where:
Y= actual value of time series
T= Trend Component
C=Cyclical Component
S=Seasonal Component
I=Irregular/Random Component or Residual

The Additive Model can also be described pictorially as:

Given the scope of the paper, we would be discussing the Smoothing Model only.

If there is a Time Series Data that is fairly stable and has no significant trend, seasonal or cyclical effects, then one may resort to *Smoothing Methods*. These methods would help in averaging out the irregular components in the Time Series Data. Basically, there are three methods to smoothing:

Simple Moving Average: This method calculates the average of the most recent *n* data values for the Time Series Data. This average is then used in order to forecast the value of the time series for the next period.

Weighted Moving Average: This method calculates the weighted average of the most recent *n* data values for the Time Series Data. This weighted average is then used in order to forecast the value of the time series for the next period. One thing that is pertinent to mention is that unlike the Simple Moving Average, different weights are assigned/allocated to past data values in Weighted Moving Average. Specifically, more recent data values are given more weight as compared to the older values.

Exponential Smoothing: This method calculates the forecast for the next period by making use of forecast and a proportion of forecast error of the current period. The following mathematical model highlights the rationale of this method (Lightner, 2004).

$$F_{t+1} = F_t + \alpha \, (Y_t - F_t)$$

where:
α is the smoothing constant (between 0 and 1)
F_t is the forecast for period t
F_{t+1} is the forecast for period t+1
Y_t is the actual data value for period t

INTRODUCTION TO R

During the last decade, the momentum coming from both academia and industry has lifted the R programming language to become the single most important tool for computational statistics, visualization and data science. R is the most comprehensive statistical analysis package available. It incorporates all of the standard statistical tests, models, and analysis. Apart, it also provides a comprehensive language for managing and manipulating data. R is a programming language and environment developed for statistical analysis by practising statisticians and researchers. It reflects well on a very competent community of computational statisticians. R is free and open source software, allowing anyone to use and, importantly, to modify it. R is licensed under the GNU General Public License, with copyright held by The R Foundation for Statistical Computing. R also runs on many operating systems and different hardware. It is popularly used on GNU/Linux, Macintosh, and Microsoft Windows, running on both 32 and 64 bit processors. R plays well with many other tools, importing data, for example, from CSV les, SAS, and SPSS, or directly from Microsoft Excel, Microsoft Access, Oracle, MySQL, and SQLite. It can also produce graphics output in PDF, JPG, PNG, and SVG formats, and table output for LATEX and HTML.

R has a steep learning curve, it is not easy to use for a novice. There are several simple-to use graphical user interfaces (GUIs) for R that encompass point and-click interactions, but they generally do not have the polish of the commercial offerings. Documentation is sometimes patchy and terse, and impenetrable to the non-statistician. However, some very high-standard books are increasingly plugging the documentation gaps. The quality of some packages are less than perfect, although if a package is useful to many people, it will quickly evolve into a very robust product through collaborative efforts. Many R commands give little thought to memory management, and so R can very quickly consume all available memory. This can be a restriction when doing data mining.

TIME SERIES FORECASTING USING "R"

We would now focus on Time Series Forecasting Techniques by making use of R Statistics Software through Exponential Smoothing Method. The pre-condition for using this technique is quite obvious- there should be a Time Series Data. A data which has a set of observations which have been measured at successive points in time or over successive periods of time is considered to be a Time Series Data. *Also, the necessary condition for using Exponential Smoothing Method is that there should be a Time Series Data that can be described using an Additive Model.*

While there are different quantitative approaches to forecasting, we will be discussing about two of them namely "Simple Exponential Smoothing" and "Holt's Exponential Smoothing". We will be making use of real life Time Series Data to help understand those techniques in a better way.

SIMPLE EXPONENTIAL SMOOTHING

This method can be used for Time Series Data that can be described using an Additive Model and that has constant level and no seasonality. The method helps in making short term forecasts. We now use the Annual Average Temperature (in degrees) data of Delhi from 1973 to 2012 to describe the method.

Year	Annual Average Temperature	Year	Annual Average Temperature
1973	25.2	1993	24.8
1974	25.3	1994	24.7
1975	24.6	1995	25.0
1976	24.8	1996	24.3
1977	24.8	1997	23.6
1978	24.6	1998	24.7
1979	25.1	1999	25.0
1980	25.2	2000	24.8
1981	25.0	2001	24.5
1982	24.4	2002	24.8
1983	23.9	2003	24.4
1984	24.7	2004	25.2
1985	25.0	2005	24.8
1986	24.5	2006	25.4
1987	25.7	2007	24.8
1988	25.1	2008	24.6
1989	24.4	2009	25.4
1990	24.8	2010	25.6
1991	24.8	2011	24.7
1992	24.8	2012	24.9

We will read the data into R and plot it by the following command:

> atdata <- read.csv ("Delhi Avg Temp Data.csv")
> atdata
> atdataseries<-ts(atdata, start=c(1973))
> plot.ts (atdataseries)

The following graph will be displayed:
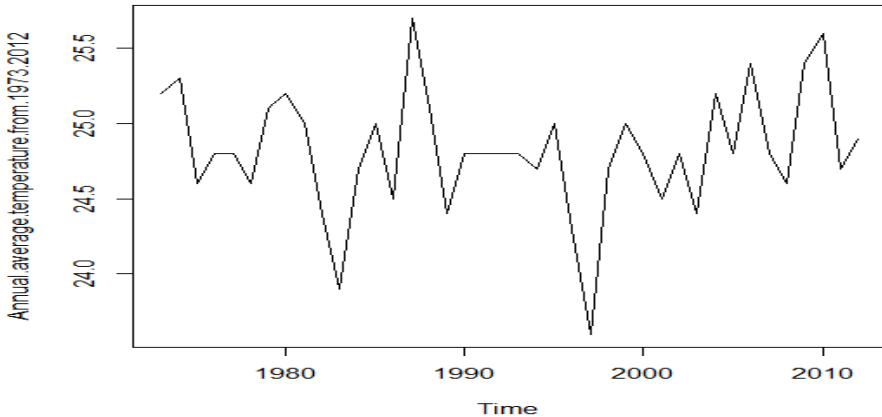
We can see from the above plot that there is more or less constant level (with mean close to 24.8 degrees). Also, it is evident that the random fluctuations seem to be more or less constant in size over time (1973 to 2012). Taking cues from these, we can probably describe the data set using an additive model. Consequently, we can make short-term forecasts using Simple Exponential Smoothing.

"R" comes with a "HoltWinters()" function in order to make forecasts using simple exponential smoothing by fitting a simple exponential smoothing predictive model. The pre-condition of using this function is that we have to set the parameters "beta=false" and "gamma=false".

For the above data set, we will type the following commands:

> atdataseriesforecasts <- HoltWinters(atdataseries, beta=FALSE, gamma=FALSE)
> atdataseriesforecasts

The following output will be generated in R:

Smoothing parameters:
 alpha: 0.174841
 beta : FALSE
 gamma: FALSE
Coefficients:
    [,1]

a 24.95756

The alpha parameter is about 0.174. This is quite close to zero giving us an indication that the forecasts are based on both recent and less recent observations. The mean (defined by the parameter"a") is 24.95. These forecasts are for the time period 1973-2012 (the data set is also for the same time period and HoltWinters() makes forecasts for the same period that is covered by the time series).

The forecasted values are stored in "fitted". In order to get the values, we can type the following:

> atdataseriesforecasts$fitted

We will get the following output:

Time Series:
Start = 1974
End = 2012
Frequency = 1
*   xhat   level*
1974 25.20000 25.20000
1975 25.21748 25.21748
1976 25.10952 25.10952
1977 25.05541 25.05541
1978 25.01075 25.01075
1979 24.93893 24.93893
1980 24.96710 24.96710
1981 25.00782 25.00782
1982 25.00645 25.00645
1983 24.90042 24.90042
…………………..
2006 24.73637 24.73637
2007 24.85240 24.85240
2008 24.84324 24.84324
2009 24.80071 24.80071
2010 24.90549 24.90549
2011 25.02692 25.02692
2012 24.96976 24.96976

Now, we will plot the original time series against the forecasts by the following command:

> plot(atdataseriesforecasts)

Following plot will be displayed:

Holt-Winters filtering
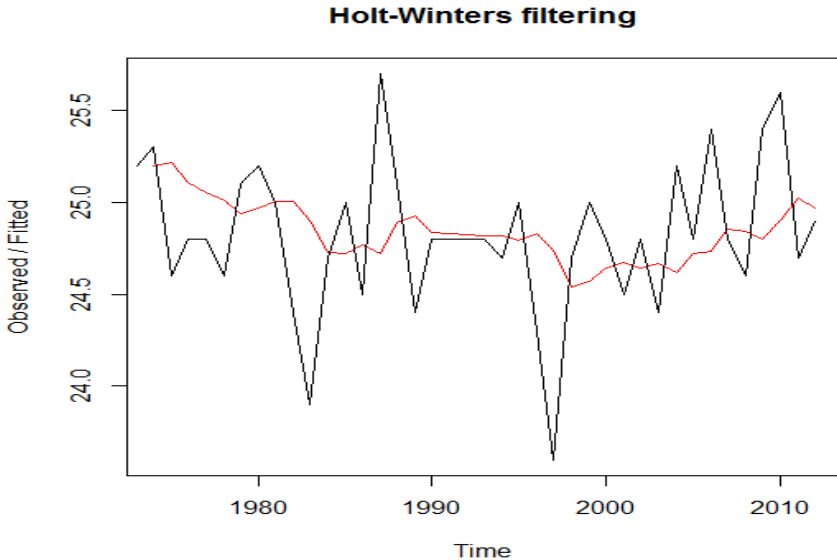
The above plot shows both the original time series as well as the forecasts. It is evident that the time series of forecasts is much smoother as compared to the time series of the original data.

In order to check the accuracy of the forecasts, we may calculate the Sum of Squared Errors (SSE) for the forecast errors. We will get its value by typing the following command:

> atdataseriesforecasts$SSE
[1] 7.398225

The SSE is 7.398225

We have forecasted the values for the time period of the original data set i.e. for 1973-2012. Now, if we want to make forecasts for further time period, we may make use of "forecast.HoltWinters()" function in the R "forecast" package. In order

to use this function, we first need to install the "forecast" R package and then load it by the following command:

> *library("forecast")*

We will make use of the predictive model (atdataseriesforecasts) that we already fitted using the HoltWinters() function. Suppose, we need to forecasts for the time period 2013-2026 i.e. for 14 more years, so we will put "h=14" in the "forecast.HoltWinters()" function. We will type the following command:

> *atdataseriesforecasts2 <- forecast.HoltWinters(atdataseriesforecasts, h=14)*
> *atdataseriesforecasts2*

Following output will be generated:

```
Point Forecast    Lo 80   Hi 80   Lo 95   Hi 95
2013      24.95756 24.39398 25.52115 24.09564 25.81949
2014      24.95756 24.38543 25.52969 24.08257 25.83256
2015      24.95756 24.37701 25.53812 24.06968 25.84544
2016      24.95756 24.36871 25.54642 24.05699 25.85814
2017      24.95756 24.36052 25.55461 24.04446 25.87066
2018      24.95756 24.35244 25.56269 24.03211 25.88302
2019      24.95756 24.34447 25.57066 24.01992 25.89521
2020      24.95756 24.33661 25.57852 24.00789 25.90724
2021      24.95756 24.32884 25.58629 23.99601 25.91912
2022      24.95756 24.32116 25.59397 23.98427 25.93086
2023      24.95756 24.31358 25.60155 23.97267 25.94246
2024      24.95756 24.30608 25.60905 23.96121 25.95392
2025      24.95756 24.29867 25.61646 23.94988 25.96525
2026      24.95756 24.29135 25.62378 23.93867 25.97646
```

The function gives us the forecast for a year and an 80% as well as 95% prediction interval for the forecast. For instance, the forecasted annual average temperature (in degrees) in Delhi for 2026 is about 24.95, with an 80% prediction interval of [24.29, 25.62] and an 95% prediction interval of [23.93, 25.97].

We may use the following function in order to plot the predictions made by forecast.HoltWinters():
> *plot.forecast (atdataseriesforecasts2)*

Following plot will be displayed:

Forecasts from HoltWinters
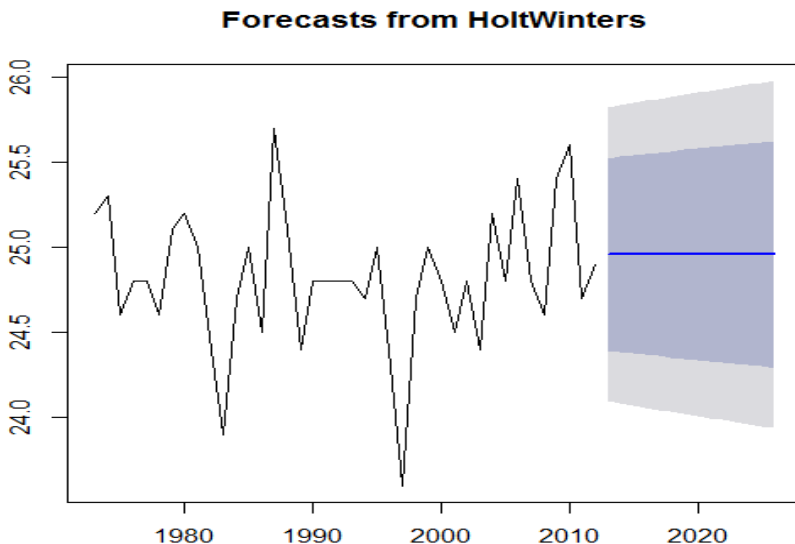
The forecasts for the time period 2013-2026 have been highlighted as a line in the shaded region, the 80% prediction interval in dark grey shaded area and the 95% prediction interval as a light grey shaded area.

Thus, we have forecasted the annual average temperature (in degrees) in Delhi for the time period 1973-2012 as well as for 2013-2026 using Simple Exponential Smoothing. Now, we need to measure the accuracy of the predictive model i.e. whether it is the best fit or whether it could be improved upon by another forecasting technique. For this, we will look at the following:
- Whether there are correlations between forecast errors for successive predictions?
- Whether the forecast errors are normally distributed with mean zero and constant variance?

In order to check whether there are correlations between forecast errors for successive predictions, we will calculate correlogram of the forecast errors using the "acf()" function. For our case, we will calculate a correlogram of the forecast errors for lags 1, 20 using the following command:

> *acf(atdataseriesforecasts2$residuals, lag.max=20)*

We will get the following output:

object$x
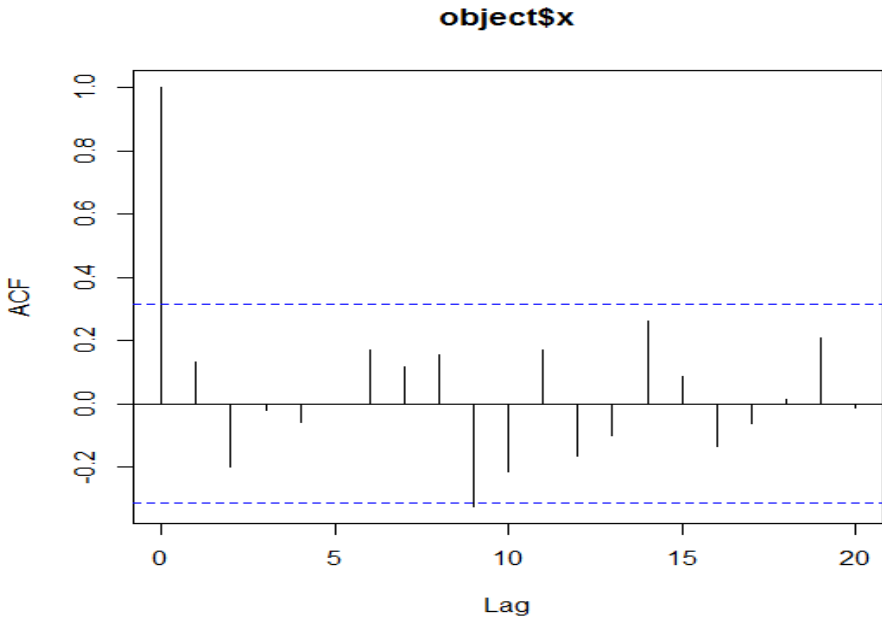
We can see from the correlogram that the sample autocorrelation for the in-sample forecast errors at lag 9 exceeds the significance bounds. However, we would expect one in 20 of the autocorrelations for the first twenty lags to exceed the 95% significance bounds by chance alone. We can carry out a Ljung-Box Test to ascertain whether there is significant evidence for non-zero correlations at lags 1-20. We will use the following command:

> Box.test (atdataseriesforecasts2$residuals, lag=20, type="Ljung-Box")

We will get the following output:
     Box-Ljung test
data: atdataseriesforecasts2$residuals
X-squared = 27.9271, df = 20, p-value = 0.1111

The Ljung-Box test statistic is 27.9 and the p-value is 0.1. It is pretty evident from these figures that there is no evidence of non-zero autocorrelations in the in-sample forecast errors at lags 1-20.

Now to check whether the forecast errors have constant variance, we will make a time plot of the in-sample forecast errors by the following command:

> *plot.ts(atdataseriesforecasts2$residuals)*

Following output will be displayed:
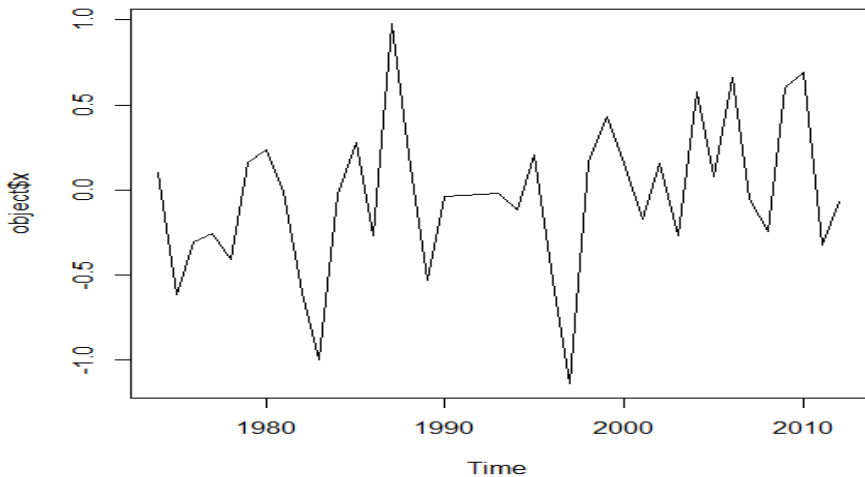
It is evident from the plot that the in-sample forecast errors seem to have roughly constant variance over time.

Now to check whether the forecast errors are normally distributed with mean zero, we will make use of the following commands:
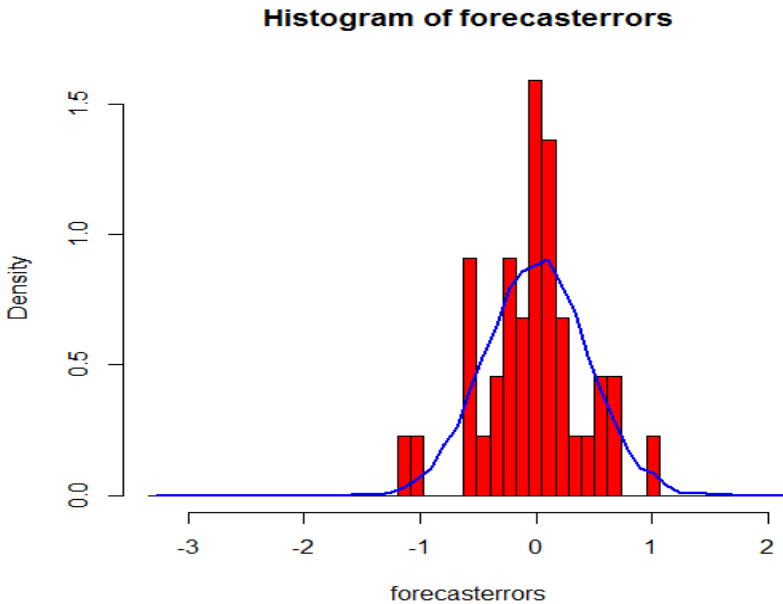
> *plotForecastErrors <- function(forecasterrors)*
+ {
+ # make a histogram of the forecast errors:
+ mybinsize <- IQR(forecasterrors)/4
+ mysd <- sd(forecasterrors)
*+ mymin <- min(forecasterrors) - mysd*5*
*+ mymax <- max(forecasterrors) + mysd*3*
+ # generate normally distributed data with mean 0 and standard deviation mysd
+ mynorm <- rnorm(10000, mean=0, sd=mysd)
+ mymin2 <- min(mynorm)

```
+ mymax2 <- max(mynorm)
+ if (mymin2 < mymin) { mymin <- mymin2 }
+ if (mymax2 > mymax) { mymax <- mymax2 }
+ # make a red histogram of the forecast errors, with the normally distributed data
overlaid:
+ mybins <- seq(mymin, mymax, mybinsize)
+ hist(forecasterrors, col="red", freq=FALSE, breaks=mybins)
+ # freq=FALSE ensures the area under the histogram = 1
+ # generate normally distributed data with mean 0 and standard deviation mysd
+ myhist <- hist(mynorm, plot=FALSE, breaks=mybins)
+ # plot the normal curve as a blue line on top of the histogram of forecast errors:
+ points(myhist$mids, myhist$density, type="l", col="blue", lwd=2)
+ }
> plotForecastErrors(atdataseriesforecasts2$residuals)
```

Following output will be displayed:

Histogram of forecasterrors

The above plot shows that the forecast errors have zero mean and the distribution is more or less normal, although it seems to be slightly negatively skewed. However, the skewness towards left is not that large and so it seems that the forecast errors are normally distributed with zero mean. Thus, we see that as per the Ljung-Box

test, there is no evidence of non-zero autocorrelations in the in-sample forecast errors and that the distribution of forecast errors is normally distributed with zero mean and constant variance. These findings amply conclude that the Simple Exponential Smoothing Method provides an adequate predictive model for "Annual Average Temperature (in degrees) in Delhi from 1973-2012" which probably cannot be improved upon.

HOLT'S EXPONENTIAL SMOOTHING

Holt's Exponential Smoothing method can be used for Time Series Data that can be described using an Additive Model with either an increasing or decreasing trend and which has no seasonality. This method will help in making short term forecasts. To illustrate, we use the Data Set of "Infant Mortality Rate (IMR) in India from 1991 to 2010" (SRS, 2011). We will read the data into R and plot it by the following command:

> *IMR <- read.csv ("IMR India.csv")*
> *IMR*

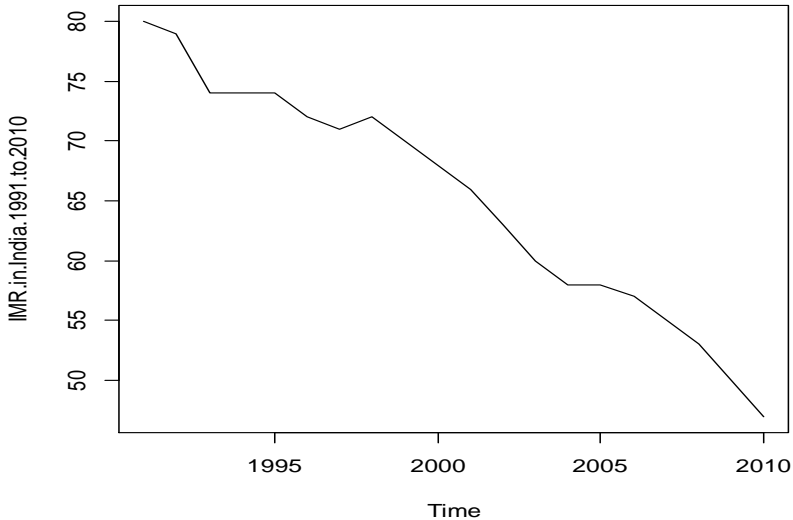Following output will be generated:

```
 IMR.in.India.1991.to.2010
1              80
2              79
3              74
4              74
5              74
6              72
7              71
8              72
9              70
10             68
11             66
12             63
13             60
14             58
15             58
16             57
17             55
18             53
19             50
```

20   47

> IMRseries <- ts(IMR, start=c(1991))
> plot.ts (IMRseries)

The following graph will be displayed:

We can see from the above plot that there is a decreasing trend in the data set (the IMR reduces from 80 in 1991 to 47 in 2010). Also, it seems that there is no seasonality. So, we can probably describe the data set using an additive model. Consequently, we can make short-term forecasts using Holt's Exponential Smoothing.

"R" comes with a "HoltWinters()" function in order to make forecasts using Holt's Exponential Smoothing by fitting a predictive model. The pre-condition of using this function is that we have to set the parameter "gamma=false".

For the above data set, we will type the following commands:

> IMRseriesforecasts <- HoltWinters(IMRseries, gamma=FALSE)
> IMRseriesforecasts

The following output will be generated in R:

Smoothing parameters:
 alpha: 1
 beta: 0.1029905
 gamma: FALSE
*Coefficients:*
    [,1]
a 47.000000
b -1.822245

The alpha parameter is 1, telling that the estimate of the current value of the level is based mostly upon very recent observations in the time series. The estimated value of beta is 0.10 which is close to zero giving us an indication that the estimate of the slope b of the trend component is based on both recent and less recent observations. These forecasts are for the time period 1991-2010 (the data set is also for the same time period and HoltWinters () makes forecasts for the same period that is covered by the time series).

Now, we will plot the original time series against the forecasts by the following command:

> plot (IMRseriesforecasts)

Following plot will be displayed:

Holt-Winters filtering
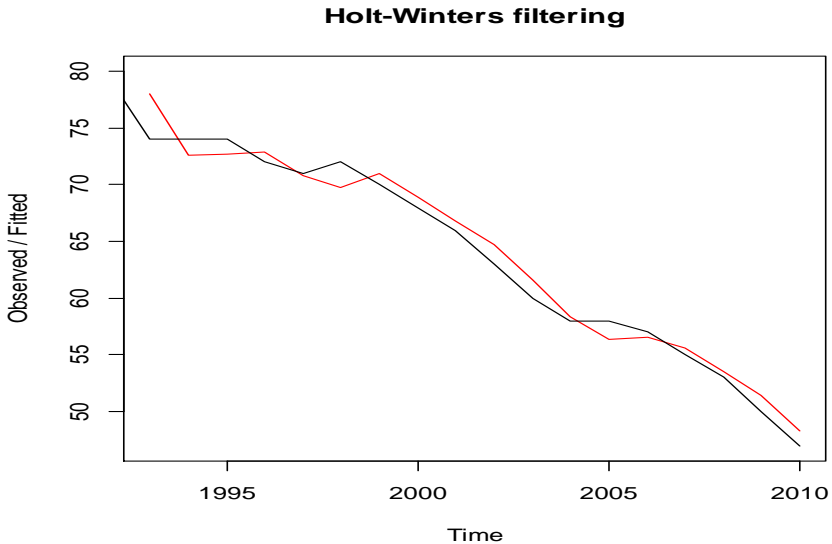
The above plot shows both the original time series (in black) as well as the forecasts (in red). It is evident that the in-sample forecasts tend to lag behind the observed values a bit. In order to check the accuracy of the forecasts, we may calculate the Sum of Squared Errors (SSE) for the forecast errors. We will get its value by typing the following command:

> *IMRseriesforecasts$SSE*
[1] 40.82339

The SSE is 40.82339

We have forecasted the values for the time period of the original data set i.e. for 1991-2010. Now, if we want to make forecasts for further time period, we may make use of "forecast.HoltWinters()" function in the R "forecast" package. In order to use this function, we first need to install the "forecast" R package and then load it by the following command:

> *library("forecast")*

We will make use of the predictive model (IMRseriesforecasts) that we already fitted using the HoltWinters() function. Suppose, we need to forecasts the IMR values for the time period 2011-2020 i.e. for 10 more years, so we will put "h=10" in the "forecast.HoltWinters()" function. We will type the following command:

> IMRseriesforecasts2 <- forecast.HoltWinters(IMRseriesforecasts, h=10)
> IMRseriesforecasts2

Following output will be generated:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2011	45.17776	43.27990	47.07561	42.27524	48.08027
2012	43.35551	40.52994	46.18108	39.03418	47.67684
2013	41.53327	37.89701	45.16952	35.97210	47.09443
2014	39.71102	35.30719	44.11486	32.97594	46.44611
2015	37.88878	32.73372	43.04383	30.00481	45.77275
2016	36.06653	30.16390	41.96917	27.03923	45.09383
2017	34.24429	27.59073	40.89785	24.06855	44.42003
2018	32.42204	25.01008	39.83400	21.08643	43.75765
2019	30.59980	22.41937	38.78023	18.08891	43.11068
2020	28.77755	19.81691	37.73820	15.07344	42.48167

The function gives us the forecast for a year and an 80% as well as 95% prediction interval for the forecast. For instance, the forecasted IMR in India for 2020 is about 29, with an 80% prediction interval of [19.81, 37.73] and an 95% prediction interval of [15.07, 42.48]. We may use the following function in order to plot the predictions made by forecast.HoltWinters():

> plot.forecast (IMRseriesforecasts2)

Following plot will be displayed:

Forecasts from HoltWinters
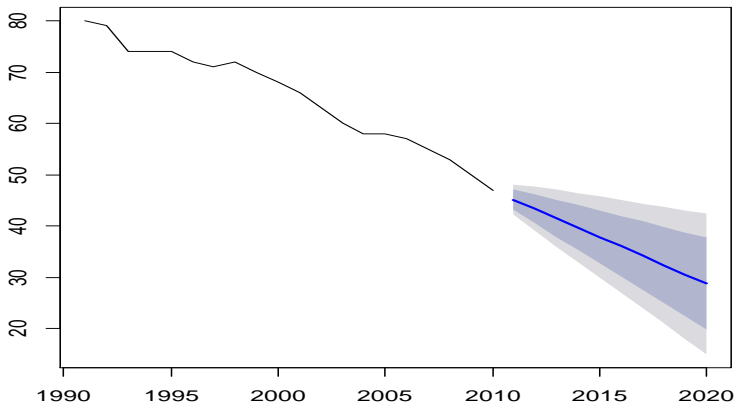
The forecasts for the time period 2011-2020 have been highlighted as a blue line, the 80% prediction interval as a dark grey shaded area and the 95% prediction interval as a light grey shaded area. Thus we have forecasted the IMR in India for the time period 1991-2010 as well as for 2011-2020 using Holt's Exponential Smoothing. Now, we need to measure the accuracy of the predictive model i.e. whether it is the best fit or whether it could be improved upon by another forecasting technique. For this, we will look at the following:

- Whether there are correlations between forecast errors for successive predictions?
- Whether the forecast errors are normally distributed with mean zero and constant variance?

In order to check whether there are correlations between forecast errors for successive predictions, we will calculate correlogram of the forecast errors using the "acf()" function. For our case, we will calculate a correlogram of the forecast errors for lags 1, 20 using the following command:

> *acf(IMRseriesforecasts2$residuals, lag.max=20)*

We will get the following output:

object$x
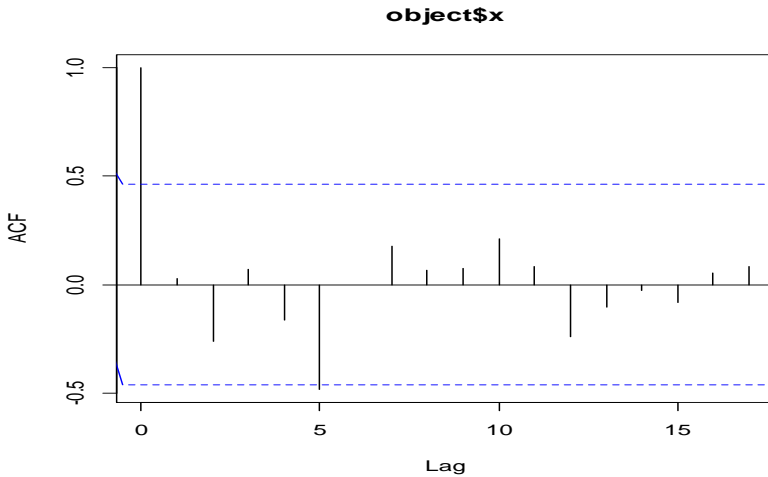
We can see from the correlogram that the autocorrelation at lag 5 exceeds the significant bound. However, there is only one such instance where the 95% significance bounds have been exceeded and that could have happened by chance alone. But, when we carry out a Ljung-Box Test at lags 1-20, we are still unable to comment whether there are correlations between forecast errors for successive predictions or not.

Now to check whether the forecast errors have constant variance, we will make a time plot of the in-sample forecast errors by the following command:

> plot.ts (IMRseriesforecasts2$residuals)

Following plot will be displayed:
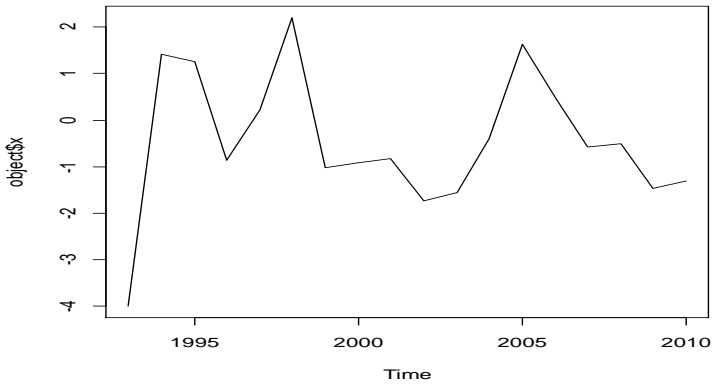
It is evident from the plot that the in-sample forecast errors do not seem to have constant variance over time. Now to check whether the forecast errors are normally distributed with mean zero, we will make use of the following command:

> *plotForecastErrors (IMRseriesforecasts2$residuals)*

Following Histogram will be displayed:

Histogram of forecasterrors
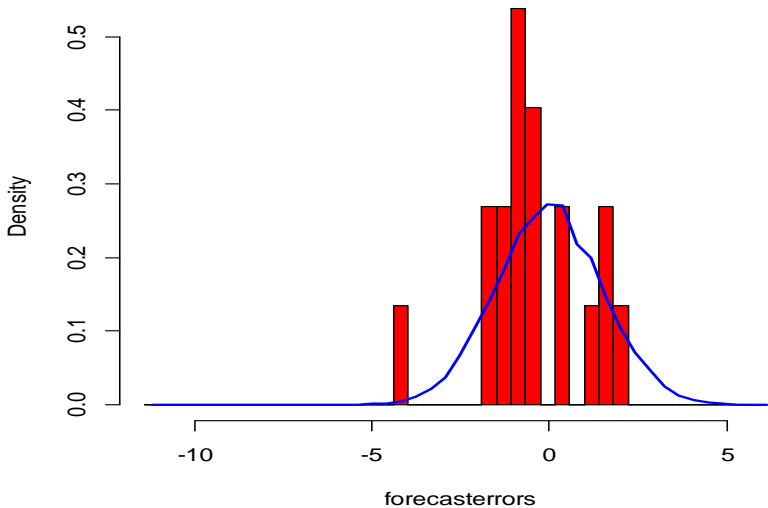
The above plot shows that the forecast errors do not seem to follow a normal distribution with zero mean and constant variance. We see that as per the Ljung-Box test, we can't comment whether there are correlations in the in-sample forecast errors for successive predictions or not. Also, the distribution of forecast errors do not seem to be normally distributed with zero mean and constant variance. These findings conclude that the Holt's Exponential Smoothing Method does not provide an adequate predictive model for "IMR in India from 1991-2010" and thus could probably be improved upon by some other forecasting technique may be by Autoregressive Integrated Moving Average (ARIMA) (Coghlan, 2014).

SUMMING-UP

The paper introduced the basic concepts of forecasting, its approaches & techniques, highlighted its importance in the Development Sector, listed down its applications in Monitoring & Evaluation and attempted to describe the Quantitative Approach to Forecasting. The paper also highlighted how forecasting has been historically used as a critical planning tool in the context of social welfare programmes and projects. Many forecasting tools like MBB and Spectrum Policy Modelling system have been widely used in the estimation of resources required and also for projecting the output indicators from a project based on the inputs provided in the project.

Taking into account the limitations of the purview of the paper, however, the applications of only some of those approaches have been listed. In particular, the paper attempted at providing practitioner's point of view so as to enable applications of Time Series Forecasting Techniques by making use of R Statistics Software through Exponential Smoothing Method. The paper also highlighted other forecasting tools such as LIST, MBB, and Spectrum and highlighted the advantages and disadvantages of using R. It was also amply evident from the paper that the type of method to be employed in order to forecast a Time Series Data solely depended upon the nature of the data. Some methods suited well for one particular data set while the other approaches fitted to the other data sets. The paper also highlighted that if a particular approach was unable to fully forecast a Time Series Data then the other methods should be resorted to. Though understanding and clarity on every method is important, it is also pertinent to note that the best method to be used, depends on the researchers' objective and experience which at the end will decide what is best for the study and the program.

REFERENCES

Anderson, J. (1989). *Forecasting, Uncertainty and Public Project Appraisal*. Policy Research Working Paper Series No 154. The World Bank

Coghlan, A. (2014). *A Little Book of R for Time Series Release 2.0*. Retrived from https://media.readthedocs.org/pdf/a-little-book-of-r-for-time-series/latest/a-little-book-of-r-for-time-series.pdf

Das, H. P., Doblas-Reyes, F. J., Garcia, A., Hansen, J., Mariani, L. Nain, A, Ramesh, K., Rathore, L.S. and Venkatraman, R (n.d.). *Weather and Climate Forecasts for Agriculture.*Retrived from http://www.agrometeorology.org/files-folder/repository/gamp_chapt4.pdf.

Devinfolive (2011). *Marginal Budgeting for Bottelnecks.* Retrieved from Devinfolive: http://www.devinfolive.info/mbb/mbbsupport/download.php?file=MBB_a96ba1 a0_MBB-LiST_short_overview_25APR11.doc. accessed on December 27, 2014

Futures Group (n.d.). *Spectrum.* Retrieved from Futures Group: http://futuresgroup.com/resources/software_models/spectrum accessed on January 4, 2015

Lightner, D. C. (2004). Forecasting techniques. Retrived from http://www.docstoc.com/docs/46411219/Sales-Forecasting---PowerPoint.

Muys, B., Achten ,W., Verbist, B., Aerts, R., Vincent, K., Martin, H., Poppe, J., Marx, A., Verheyen, K., De Tavernier, J., Fanta, J., Van Rompaey, A., Vranken, L., Dondeyne, S., Lenaerts, L., Nabuurs, G. (2011). *Conservation and Management of Forests for Sustainable development: where science meets policy.* Metaforum Working Group, University of Leuven, Netherland.

Roberts. (1992). *Time Series Data Library*. Retrieved from Data Market: http://datamarket.com/data/set/22kz/annual-precipitation-inches-lake-michigan-1900-to-1986#!ds=22kz&display=line

Social Solutions (n.d.). Retrieved from Social Solutions: http://www.socialsolutions.com/blog/turning-measurement-into-effectiveness-the-power-of-forecasting accessed on December 12, 2014

Social Solutions (n.d.). *Turning measurement into effectiveness: The power of Forecasting.* Retrieved from Social Solutions: http://www.socialsolutions.com/blog/turning-measurement-into-effectiveness-the-power-of-forecasting

SRS. (2011). Sample Registration System. Retrieved from Census of India: http://www.censusindia.gov.in/2011-Common/Sample_Registration_System.html, http://www.censusindia.gov.in/vital_statistics/SRS_Bulletins/MMR_release_070711.pdf accessed on December 6, 2014

Tanga, B. E. (2005). *Forecasting land use change and its environmental impact*. Journal of Environmental Management 76, 35-45.

USAID (2009). *AIM A Computer based programme for Making HIV/AIDS Projections and Examining the Demographic and Social Impact of AIDS*. Retrived from http://data.unaids.org/pub/Manual/2009/20090414_aim_manual_2009_en.pdf

Notes

SAMBODHI