International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2018; 3(1): 505-510

www.mathsjournal.com
Received: 17-11-2017
Accepted: 18-12-2017

K Sreenivasulu
Lecturer in Statistics, Sri Sarvodaya College, Nellore, Andhra Pradesh, India

G Mokesh Rayalu
Assistant Professor, Department of Mathematics, School of Advanced sciences, VIT, Vellore, Tamilnadu, India

Correspondence
G Mokesh Rayalu
Assistant Professor, Department of Mathematics, School of Advanced sciences, VIT, Vellore, Tamilnadu, India

Statistical domain approaches in forecasting techniques

K Sreenivasulu and G Mokesh Rayalu

Abstract
Time series analysis has been created by four types of experts namely Engineers and Physical Scientists; Economists; Applied Statisticians and Econometricians; and Mathematical Statisticians and Probabilists Time series analysis was divided into two approaches namely Frequency Domain and Time Domain approaches. For Engineers and Physical Scientists, it is natural to regard a time series as made up of oscillations at different frequencies. This gives time domain analysis. There is duality between the frequency domain and time domain approaches to the time series analysis.

Keywords: Statistical domain approaches, forecasting techniques

1. Introduction

In univariate time series regression models, the regressor variable is time. A linear trend relationship can be written as

$$Y = \alpha + \beta T + \in \tag{1.1}$$

Where T indicates time. The T variable may be specified in many ways. When T has zero mean, the normal equations for fitting (1.1) will become

$$a = \bar{Y} \text{and } b = \frac{\sum Y}{\sum T^2}$$

One way of modeling such behavior is by means of autoregression. The simplest autoregressive model is

$$Y_t = \alpha + \beta Y_{t-1} + \in_t \tag{1.2}$$

This is called a first order autoregressive model and is denoted by AR(1). The order indicates the maximum lag in the equation.From the equation (1.2), we make the following assumptions about $\in$ variable

$$\begin{aligned} &E(\in_i) = 0 \quad \text{for all i} \\ &E(\in_i^2) = \sigma^2 \quad \text{for all I} \\ &E(\in_i \in_j) = 0 \quad \text{for all } i \neq j \end{aligned} \tag{1.3}$$

These assumptions define a white noise series. Here the crucial question is, how does the Y series behave over time. Assuming that process stated a very long time ago, we take

$$\begin{aligned} &Y_t = \alpha\left(1+\beta+\beta^2+......\right)+\left(\in_t + \beta \in_{t-1} + \beta_t^2 \in_{t-2} + ...\right) \\ &E(Y_t) = \alpha\left(1+\beta+\beta^2+.......\right) \end{aligned} \tag{1.4}$$

This expectation exists only when the infinite geometric series on the right hand side has limit. The necessary and sufficient conditions is $|\beta| < 1$.

The expectation is then $E(Y_t) = \mu = \dfrac{\alpha}{1-\beta}$ (1.5)

Here variance Y will be $Var(Y) = \sigma_y^2 - \dfrac{\sigma^2}{1-\beta^2}$ (1.6)

The Y series has a constant unconditional variance, independent of time.
The covariance of Y with a lagged value itself is known as autocovariance.
The lag auto covariance is defined as

$$\gamma_s = \beta^s \sigma_y^2 \quad s = 0,1,2, \tag{1.7}$$

So that first lag autocovariance is $\gamma_1 = \beta\sigma_y^z$

The autocovariances depend only on the lag length and are independent of t Dividing the covariances by the variance gives the autocorrelation coefficients, also known as series correlation coefficients. These will be defined as

$$\rho_s = \gamma_s / \gamma_o, \quad s = 0, 1, 2\ldots\ldots. \tag{1.8}$$

Where $\gamma_o = \sigma_u^z$

Plotting the autocorrelation coefficients against the lag lengths gives the correlogram of series.

When $|\beta|<1$ the mean, variance, and covariances of Y series are constants independent of time. The Y series is then said to be weekly or covariance stationary.

When $\beta = 1$, the AR(1) process is said to have unit root. The equation becomes.

$$Y_t = \alpha + Y_{t-1} + \in_1 \tag{1.9}$$

Which is called a random walk with drift. The conditional expectation and conditional variance are $E(Y_t / Y_0) = at + Y_o$

Which increases or decreases without limit as t increases $Var(Y_t / Y_0) = t\sigma^2$

Which increases without limit.In this case the unconditional mean and variance do not exist. The Y series is then said to be nonstationary.

2. Autoregressive Model with Order $p : AR(p)$

A common approach for modeling univariate time series is the autoregressive (AR) model. The AR(p) model is defined as

$$X_t = \delta + \phi X_{t-1} + \phi_2 X_{t-2} + \ldots\ldots\ldots\phi_p X_{t-p} + \in_t \tag{2.1}$$

where X_t is the time series, $\delta = \left(1 - \sum_{i-1}^{p} \phi_i\right)\mu,$ where μ is the process mean

$\in_t$ is the white noise

$\phi_1, \phi_2 \ldots\ldots\phi_p$ are the parameters of the model

p is the order of the AR model

Some constraints are necessary on the values of the parameters of this model so that model remains wide-sense stationary. For an AR(P) model to be wide-sense stationary, the roots of the polynomial $Z^p - \sum_{i=1}^{p} \phi_i Z^{p-i}$ must lie within the unit circle, i.e., each root Z_i must satisfy $|Z_i|>|$.

3. Moving Average Model With Order q : MA (q)

In time series analysis, the moving average (MA) model is common approach for modeling univariate time series models. The moving average model with order qMA(q) is

$$X_t = \mu + \in_t + \theta_1 \in_{t-1} + \ldots\ldots\theta_q \in_{t-q} \tag{3.1}$$

Where, μ is the mean of the series,

$\theta_{1,}\ \theta_2......,\theta_q$ are the parameters of the model,

$\in_t,\in_{t-1},..........$ are white noise error terms.

q is the order of the moving average model.

4. Autoregressive moving average (arma) model

It is convenient to use the notation ARMA (p, q), where p is the order of the autoregressive part and q is the order of the moving average part.

The general AR (p) model was represented as

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + \in_t \qquad (4.1)$$

Multiplying both sides by X_{t-k} yields

$$X_{t-k}X_t = \phi_1 X_{t-k}X_{t-1} + \phi_2 X_{t-k}X_{t-2} + ... + \phi_p X_{t-k}X_{t-p} + X_{t-k}\in_t$$

Taking expected value both sides and assuming stationarity gives

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + ... + \phi_p\gamma_{k-p}$$

Where γ_k is the covariance between X_t and X_{t-k}

The MA (q) model is written as

$$X_k = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}................\theta_q e_{t-q}$$

Multiplying bothsides by X_{t-k} yields

$$X_{t-k}X_t = \left(e_t - \theta_1\, e_{t-1} - \theta_2\, e_{t-2}..........\theta_q e_{t-q}\right) x \left(e_{t-k} - \theta_1\, e_{t-k-1} - \theta_2\, e_{t-k-2}..........\theta_q e_{t-k-q}\right) \quad (4.2)$$

The expected value of the above equation will depend upon the value of k.

If k=0, and all other terms of the equation on (4.2) dropout because

$$E\left(e_t e_{t+i}\right) = 0 \text{ for } i \neq 0 \text{ and } E\left(e_t\, e_{t+i}\right) = \sigma_e^2 \text{ for } i = 0.$$

Thus (4.1) becomes, $\gamma_0 = \sigma_e^2 + \theta_1^2\sigma_e^2 + \theta_2^2\sigma_e^2 + ... + \theta_q^2\,\sigma_e^2$

To obtain the initial estimates for ARMA models, combine AR and MA models:

$$\gamma_k = \phi_1 E\left(x_t x_{t-k}\right) + ... + \phi_p E\left(X_{t-p}X_{t-k}\right) + E\left(e_t X_{t-k}\right) - \theta_1 E\left(e_{t-1}X_{t-k}\right) - ..... - \theta_q E\left(e_{t-q}X_{t-k}\right) \qquad (4.3)$$

If k>q, the terms $E\left(e_t X_{t-k}\right) = 0$ which leaves

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + ... + \phi_p\gamma_{k-p}$$

When k<q, the past errors and the X_{t-k} will be correlated and the autocovariances will be affected by the moving average part of the process, requiring that it will be included. The variance and autocovariances of an ARMA (1, 1) process are therefore obtained as follows:

$$X_t = \phi_1 X_{t-1} + e_t - \theta_1\, e_{t-1}$$

Multiplying bothsides by X_{t-k} gives

$$X_{t-k}X_t = \phi_1 X_{t-k}X_{t-1} + X_{t-k}e_t - \theta_1 X_{t-k}e_{t-1}$$

Taking expected values both sides results in

$$E\left[X_{t-k}X_t\right]=\phi E\left[X_{t-k}X_{t-1}\right]+E\left[X_{t-k}e_t\right]-\theta_1 E\left[X_{t-k}e_{t-1}\right]$$

If k=0, $\gamma_0=\theta_1\gamma_1+E\left[\left(\phi_1 X_{t-1}+e_t-\theta_1 e_{t-1}\right)e_t\right]-\theta_1 E\left[\left(\phi_1 X_{t-1}+e_t-\theta_1 e_{t-1}\right)e_{t-1}\right]$ (4.4)

Since $x_t=\phi_1 k_{t-1}+e_t-\theta_1 e_{t-1}$

$$\gamma_0=\phi_1\gamma_1+\sigma_e^2-\theta_1\left(\phi_1-\theta_1\right)\sigma_e^2$$

Similarly if k=1, $\gamma_1=\phi_1\gamma_0-\theta_1\sigma_e^2$ (4.5)

Solving the equations (4.2) ad (4.3) for γ_0 and γ_1 we get

$$\gamma_0=\frac{1+\theta_1^2-2\phi_1\theta_1}{1-\theta_1^2} \quad (4.6)$$

$$\gamma_1=\frac{\left(1-\phi_1\theta_1\right)\left(\phi_1-\theta_1\right)}{1-\theta_1^2} \quad (4.7)$$

Dividing (4.5) by (4.4) gives

$$\rho_1=\frac{\left(1-\phi_1\theta_1\right)\left(\phi_1-\theta_1\right)}{1+\theta_1^2-2\phi_1\theta_1} \quad (4.8)$$

5. Autoregressive integrated moving average (aima) model

To identity the appropriate ARIMA model for a time series, one has to begin by identifying the order (r) of differencing needing to stationarize the series and remove the gross features of seasonality, in conjunction with a variance–stabilizing transformation such as lagging or deflating.

The equation for the simplest case ARIMA (1, 1, 1) is as follows

$$(1-B).(1-\phi_1 B)X_t=\mu^1+(1-\phi_1 B)e_t$$

The terms can be multiplied out and rearranged as follows.

$$\left[1-B\,(1+\phi_1)+\phi_1 B^2\right]x_t=\mu^1+e_t-\theta_1 e_{t-1}$$

$$X_t=(1+\phi_1)X_{t-1}-\phi_2 X_{t-2}+\mu^1+e_t-\theta_1 e_{t-1} \quad (5.1)$$

In this form, the ARIMA model looks more like a conventional regression equation, except that there is more than one error term on the right hand side. ARIMA models are quite flexible and can represent a wide range of characteristics of time series that occur in practice.

5.1 ARIMA Model : p^th order autoregressive model

$$Y_t=\phi_0+\phi_1 Y_{t-1}+\phi_2 Y_{t-2}+.......\phi_p Y_{t-p}+\in_t \quad (5.2)$$

5.2 ARIMA Model: q^th order moving average model

$$Y_t=\mu+\in_t-W_1\in_{t-1}-W_2\in_{t-2}...........W_q t_{t-q} \quad (5.3)$$

5.3 ARIMA Model: ARMA (p, q) model

$$Y_t=\phi_0+\phi_1 Y_{t-1}+\phi_2 Y_{t-2}+...........+\phi_p Y_{t-p}+\in_t$$
$$-W_1\in_{t-1}-W_2\in_{t-2}..........W_q\in_{t-q} \quad (5.4)$$

5.4 ARIMA (0, 1, 1) Model

$$Y_t - Y_{t-1} = \in_t - w_1 \in_{t-1} \quad (5.6)$$

6. Autoregresive distributed lag models (ARDL)

In case where the variables in the long-run relation of interest are trend stationary, the general practice has been to de-trend the series and to model the detrended series as stationary distributed lag or autoregressive distributed lag (ARDL) models. Estimation and inference concerning the long-run properties of the model are then carried out using standard asymptotic normal theory. The analysis becomes more complicated when the variables are difference–stationary, or integrated of order 1. The auto-regressive distributed lag model with p lags of dependent variable Y_t and q lags of additional regressor X_t, ADL (p, q) is defined as:

$$ADL(p,q) = Y_t = \beta_0 + \beta_1 Y_{t-1} + ... + \beta_p Y_{t-p} + \delta_1 X_{t-1} + ... + \delta_q X_{t-q} + \mu_t \quad (6.1)$$

$\beta(L)Y_t = \beta_0 + \delta(L)X_{t-1}$ with lag-polynomials defined by

$$\beta(L)Y_t = 1 - \beta_1 L - .... - \beta_p L_p$$

$$\delta(L) = \delta_1 + \delta_2 L + ... + \delta_q L_{q-1} \quad (6.2)$$

k additional predictors = ADL $(p, q_1, ..., q_k)$ (6.3)

$$\beta(L)Y_t = \beta_0 + \delta_1(L)X_{1,t-1} + \delta_2(L)X_{2,t-1} + ... + \delta_k(L)X_{k,t-1}$$

Model Assumptions

1. $E\left[u_t / Y_{t-1,t-2,} ......, X_{1t-1}, X_{1t-2,} ...... X_{k\,t-1},\ X_{kt-2} .....\right] = 0$ (6.4)
2. (a). $\left(Y_{t,}\ X_{1t,} .... X_{kt}\right)$ are (strictly) stationary

   (b). $\left(Y_{t,}\ X_{1t,} ...., \ X_{kt}\right)$ are ergodic (6.5)

   $\left(Y_{t,} X_{1t,} ...., X_{kt}\right)$ and $\left(Y_{t-j,}\ X_{1t-j}, ..., X_{kt-j}\right)$ becomes independent for $j \to \infty$
3. Y_t and $X_{1t,}\ ......, X_{kt}$ have nonzero, finite fourth moments
4. no perfect multicollinearity

7. Non stationary time series regression models

A time series X_t is said to be stationary if its expected value and population variance are independent of time and if the population covariance between its values at time t and t+s depends on s but not on time.

An example of a stationary time series is an AR (1) process

$$X_t = \beta_2 X_{t-1} + \in_t \quad (7.1)$$

Here $-1 < \beta_2 < 1$ and $\in_t$ is white noise

If equation (7.1) is valid for time period t, it is also valid for time period t-1 i.e. the series is stationary.

$$X_{t-1} = \beta_2 X_{t-2} + \in_{t-1}$$

Substituting for X_{t-1} in equation (7.1), we get

$$X_t = \beta_2^2 X_{t-2} + \beta_2 \in_{t-1} + \in_t$$

Continuing this process of lagging and substituting, we get

$$X_t = \beta_2^t X_0 + \beta_2^{t-1} \in_1 + ... + \beta_2 \in_{t-1} + \in_t \quad (7.2)$$

In the previous examples, if β_2 is equal to 1, the original series becomes

$$X_t = X_{t-1} + \in_t \quad (7.3)$$

This is an example of a nonstationary process which is known as a random walk
If it starts at X_0 at time 0, its value at time t is given by

$$X_t = X_0 + \in_t + \ldots.. \in_t \tag{7.4}$$

The key difference between this process and the corresponding A.R(1) process is that the contribution of each innovation is permanently built into the time series. In the more general version of the autogressive process with the constant β_1, the process becomes a random walk with drift, if β_2 equals1.

$$X_t = \beta_1 + X_{t-1} + \in_t \tag{7.5}$$

If the series starts at X_0 at time 0, X_t is given by

$$X_t = X_0 + \beta_{1t} + \in_1 + \ldots\ldots.. + \in_t$$

The expectation of X_t at time '0', $\left(X_0 + \beta_{1t}\right)$ is a function of 't'. Another common example of a nonstationary time series is one possessing a time trend:

$$X_t = \beta_1 + \beta_2 t + \in_t \tag{7.6}$$

This type of trend is sometimes described as a deterministic trend. The expected value of X_t at time 't' $\left(\beta_1 + \beta_2 t\right)$ is not independent of t and so X_t is nonstationary.

8. Conclusions

Time series analysis has advance from univariate modeling based on a single variable to multivariate models that employ the interrelationships between several such variables. Constructing such models requires the performing of tests to determine and to discover the interactions that exist between a given time series variables and one or more other variables. The given variable can be influenced not only by certain exogenous events occurring at particular points in time but also by contemporaneous, lagged and leading values of another variable or additional variables.

9. References

1. Amemiya T. Selection of Regressors. International Economic Review. 1980; 21:331-354.
2. Antoniadis A, Gijbels I, Gregoire G. Model Selection Using Wavelet Decomposition and Applications, Biometrika. 1997; 84(4):751-763.
3. Aoki M, Havenner A. State-Space Modelling of Multiple Time Series, Econometric Reviews. 1991; 10:1-59.
4. Atikinson AC. A Method for Discriminating Between Models. Journal of the Royal Statistical Society B. 1970; 32:323-353.
5. Box GEP, Jenkins GM. Time Series Analysis, Forecasting and Control, Rev. ed. San Francisco: Holden-Day, 1976.
6. Davidson R, MacKinnon JG. Estimation and Inference in Econometrics, New York: Oxford University Press, 1993.
7. Eubank RL, Hart JD. Testing Goodness-of-Fit in Regression Via Order Selection Criteria, Ann. Statist. 1992; 20:1412-1425.
8. Fan J. Local Linear Regression smoothers and Their Minimax Efficiencies, Annals of Statistics. 1993; 21:196-216.
9. Guttman I, Pereyra V, Scolnik HD. Least Squares Estimation for a Class of Non-Linear Model, Technometrics, 1973; 15:2.
10. Hall A. The Information Matrix Test for the Linear Model, Review of Economic Studies. 1987; 54:257-263.
11. Loh W. A New Method for Testing Separate Families of Hypotheses. Journal of the American Statistical Association. 1985; 80:362-368.
12. MacKinnon JG. Model Specification Tests Against Non-Nested Alternatives, Econometric Reviews, 1983; 2:85-110.
13. Rao CR. Linear Statistical Inference and its Applications, 2nd ed. (Wiley, New York), 1973.
14. White H, Olson L. Determinants of Wage Change on the Job: A Symmetric Test of Non-Nested Hypotheses, Mimeo. (University of Rochester, Rochester, NY), 1979.
15. Wu P. Variable Addition Test, Econometric Theory, Problem. 1993; 93(1-2)9:145-146.
16. Zellner A. An Introduction to Bayesian Influence in Econometrics, Wiley, New York, 1971.