IOP Conf. Series: Materials Science and Engineering **263** (2017) 042135 doi:10.1088/1757-899X/263/4/042135

Goodness of fit and model selection criteria for time series models

G Mokesh Rayalu, J Ravisankar and G Y Mythili
Department of Mathematics, School of Advanced Sciences,VIT University,Vellore-632014, India

E-mail : mokesh.g@vit.ac.in

Abstract. Time series is an important part of decision making and several results are constructed on estimates of forthcoming events. In the Time series process, since, for the coming actions include probability; the predictions are frequently not perfect. The objectives of the Time series are to decrease the prediction error; to create predictions that are rarely improper and that have minor prediction mistakes. In selecting a suitable Time series model, the researcher wants to be responsive that numerous altered models may have comparable properties. A good model will fit the data well. The goodness of fit recovers as additional limits are involved in the model. This arises a problem with the ARMA (p ,q) models, because p and q take low values. It should be noted that the best model fit need not imply to provide best forecasts. There exists several model selection criteria that trade off a decrease in the sum of squares of the errors for a more sparing model. A goodness of fit criterion for ARMA (p, q) model and modified selection criteria for Time series models have been suggested in the present study.

1. Introduction

Formation for the forthcoming is the principle of any business. Business requisite estimates of for the coming values of corporate variables. Industry needs predictions of quantity, transactions and request for production planning and financial decisions [1, 2]. There are suitably of Time series models available and “choosing the right one” is not an informal task.

The business nature, the data nature, forecast limit, projection life of the model and the estimated precision of the predictions. Predictions that are established on statistical Time series models are known as quantitative forecasts. Once the Prediction model has been specified, the corresponding forecasts are unbendable mechanically.

2. Selection of time series models

At the time series phase, the last time series model is recycled to obtain the predictions. These predictions are contingent on the quantified time series model, one makes certain that the Time series modelling and its parameters break continuous during the prediction period. The constancy of the projection modelling can be measured by checking the predictions beside the new explanations [3]. Prediction errors can be computed and probable deviations in the Time series modelling can be noticed.

The prediction distance is too vital, since these procedures that create long-term and short-term predictions differ. Generally, a forecaster prefers to build models [4], which are informal to recognize, practice and clarify. An intricate advanced model may principal to additional precise predictions but

IOP Conf. Series: Materials Science and Engineering **263** (2017) 042135 doi:10.1088/1757-899X/263/4/042135

may be additional expensive and problematic to instrument [5]. In practice, in a excellent between completing models [6]. Also, an significant thought in the special of an suitable prediction model is the obtain ability of appropriate data; one cannot imagine to build precise experiential prediction models from a incomplete and inadequate data base [10].

3. Time series from regression model

The regression models can be written as[15]

$$y_i = f(x_i;\beta) + \in_i \tag{3.1}$$

here $f(x_i;\beta)$ is a statistical function of the k independent variables

$x_i = (x_{1i}, x_{2i}, ..., x_{ki})'$ and unknown parameters $\beta = (\beta_1, \beta_2, ..., \beta_m)'$,

$\in_i$ is a random error variable.

Assumptions on $\in_i$:

(i) $E(\in_i) = 0$ and $Var(\in_i) = \sigma^2$, a constant, $\forall i = 1, 2, ..., n.$

(ii) $Cov(\in_i, \in_j) = E(\varepsilon_i\, \varepsilon_j) = 0, \forall i \neq j$ is

(iii) $\in_i \overset{*.}{\sim} N(0, \sigma^2)$

When the y dependent variable then the assumption as

(i) $E(y_i / x_i) = f(x_i;\beta)$ and $Var(y_i / x_i) = \sigma^2$ is independent of x_i.

(ii) $Cov(y_i, x_j) = E[y_i - f(x_i;\beta)][y_j - f(x_j;\beta)] = 0, \forall_i \; i \neq j = 1, 2, ..., n$

(iii) Conditional on x_i, the dependent variable;

$$y_i \overset{i.i.d.}{\sim} N(f(x_i;\beta), \sigma^2)$$

(iv) The independent variables are fixed and non-stochastic.

The linear model is

$$y_i = \beta_o + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} ....\beta_{k-1} x_{(k-1)i} + \beta_k x_{ki} + \in_i \tag{3.2}$$

Here K explanatory variables and If β's and σ are known parameters, then the conditional expectation can be written as $E[y_o / x_o] = f(x_o;\beta)$

The prediction of y_o by $\hat{y}_0$ and then the predicted error by $e_o = [y_o - \hat{y}_0]$. The expected value of this error is given by

$$E(\in_0^2) = E([f(x_o;\beta)] + \in_o - \hat{y}_o)^2 = \sigma^2 + E[f(x_o;\beta) - \hat{y}_o]^2 \tag{3.3}$$

It is minimized is $\hat{y}_o = f(x_o;\beta)$ (3.4)

Since, the parameters β's and σ are known parameters in regression model. $(y_1, x_1), (y_2, x_2)..........., (y_n, x_n)$. If $\hat{\beta}$ be an estimate of β, then $\hat{y}_o = f(x_o;\hat{\beta})$ can be taken as a forecast for y_o. The parameter estimates that minimize the sum of squared deviations

$R(\beta) = \sum_{i=1}^{n} [y_i - f(x_i;\beta)]^2$ are known as the least squares estimators and are denoted by $\hat{\beta}$.

4. ARMA(p, q) processes:

An ARMA (p,q) process contains both autoregressive and moving average terms as follows.

$$y_t = \phi_0 + \sum_{i=1}^{p}\phi_i\, y_{t-i} + \sum_{i=0}^{q}\theta_i \in_{t-i} \tag{4.1}$$

By the convention of normalizing units, θ_0 is always equal to unity. Thus, ARMA (p, q) can be rewritten as

$$\begin{aligned} y_t &= \phi_0 + \sum \phi_i\, y_{t-i} + ...+ \in_t \\ &+\theta_1 \in_{t-1} + \theta_2 \in_{t-2} + ... + \theta_q \in_{t-q} \end{aligned} \tag{4.2}$$

If q=0, then ARMA (p,q) reduces to a pure AR (p) process and if P=0, then ARMA (p,q) reduces to a pure MA (q) process.

By using the lag operators, (4.2) may be written as

$$\left(\phi_1\, L - \sum \phi_i\, y_{t-i}\right) y_t = \phi_0 + \left(1 + \theta_1 L + \theta_2 L^2 + \theta_3 L^3 + ... + \theta_q L^q\right) \in_t \tag{4.3}$$

Provided that the roots of

$1 - \sum \phi_i\, z^i = 0$ lie outer the unit circle and both sides of (4.3) can be divided by $\left(1 - \sum \phi_i\, L^i\right)$ to get

$$y_t = \mu + \psi(L)\,(\in_t) \tag{4.4}$$

where $\mu = \dfrac{\phi_0}{1 - \phi_1 - \phi_2 - .... - \phi_p}$ (4.5)

and $\psi(L) = \dfrac{1 + \theta_1 L + \theta_2 L^2 + ...\theta_q L^q}{1 - \phi_1 L - \phi_2 L^2 - .... - \phi_p L^p}$ (4.6)

$$\sum_{j=0}^{\alpha} \left|\psi_j\right| < \infty \tag{4.7}$$

The L is defined as linear operator such that for any value y_i

$$L^i y_t \equiv y_{t-i} \tag{4.8}$$

ARMA process be influenced by entirely on these parameters $\left(\phi_1, \phi_2 .... \phi_p\right)$ and not on the morning average parameters $\left(\theta_1, \theta_2 ....\theta_q\right)$ (4.9)

Thus, after q lags, the autocovariance function γ_1 follows the p^{th} order difference equation governed by the autoregressive parameters. (4.10)

Also, the autocorrelation coefficients ρ_j is satisfy

$$\rho_j = \sum \phi_i \rho_{j-i},\ \forall\, j = q+1, q+2, ... \tag{4.11}$$

The stationary ARMA (p,q) process satisfying (4.3) is identical to the stationary
With p-1 and q-1 ARMA Process
The ARMA (1,1) process model without constant term as

$$y_t = \phi_1 y_{t-1} + \in_t\ \theta_1 \in_{t-1} \tag{4.12}$$

The autocovariances for ARMA (1,1) are given by

$$\gamma_0 = \frac{1 + \theta_1^2 + 2\,\phi_1\,\theta_1}{(1 - \phi_1^2)} \tag{4.13}$$

$$\gamma_1 = \frac{(1 + \phi_1\,\theta_1)\,(\phi_1 + \theta_1)}{(1 - \phi_1^2)} \tag{4.14}$$

Also, $\gamma_s = \phi_i\ \gamma_{s-1}$, $s = 2, 3, ...$

The ARMA (1,1) process is given by $\rho_1 = \dfrac{(1+\phi_1\,\theta_1)\,(\phi_1+\theta_1)}{(1+\phi_1^2+2\phi_1\,\theta_1)}$ (4.15)

And $\rho_s = \phi_1\rho_{s-1}, \forall s \geq 2$ (4.16)

5. Goodness of fit criterion for ARMA model:

The standard practice for 'diagnostic checking' is to plot the errors to check outliers and for indication of periods in has not a goodness of fit in the model. If ARMA models reveal confirmation of a low fit during a practically long part in the sample, it is better practice to reflect models other ARMA models such as models for intervention analysis. Models for transfer function analysis etc. If the residuals variance increases then a logarithmic transformation may be appropriate or an ARCH model may be considered in practice.

If there are sufficient observations, fitting the ARMA (p, q) model to each of the subsamples can provide the useful information concerning the assumption that the data generating process is unchanging.

Suppose an ARMA (p, q) model has been estimated by using a sample of n observations. Divide the n observations into two subsamples with t_m observation in first subsample and $(n - t_m)$ observations in the second subsample, using each subsample to estimate the following two ARMA (p, q) models.

sub Sample I: $y_t=\alpha_0^1 + \alpha_1^1\, y_{t-1} + \ldots\ldots + \alpha_p^1 y_{t-p} + \in_t + \beta_1^1 \in_{t-1} + \ldots\ldots\ldots + \beta_q^1 \in_{t-q}$ (5.1)

Using $t_1, t_2 \ldots\ldots, t_m$

Sub Sample II : $y_t=\alpha_0^2+\alpha_1^2\, y_{t-1}+\ldots\ldots+\alpha_p^2\, y_{t-p} + \in_t + \beta_1^2 \in_{t-1} \ldots\ldots\ldots + \beta_q^2 \in_{t-q}$ (5.2)

Using $t_{m+1}, t_{m+2}, \ldots..t_n$

First compute the OLS residuals and then Internally studentized residuals and hence obtain the sum of squared internally studentized residuals from each model as $(RSS)_I$ and $(RSS)_{II}$ respectively. Let RSS be the sum of squares due to internally studentized residuals for the model based on n observations.

The corresponding coefficients in the two Time series models

$H_o : \alpha_0^1 = \alpha_0^2, \alpha_1^1 = \alpha_1^2, \ldots\ldots.., \alpha_p^1 = \alpha_p^2; \beta_1^1 = \beta_1^2, \beta_2^1 = \beta_2^2, \ldots\ldots, \beta_q^1 = \beta_q^2$

the F - test statistic is given by

$$F = \frac{RSS - (RSS)_I - (RSS)_{II} / k}{\left[(RSS)_I + (RSS)_{II}\right] / n-2k} \sim F_{k,\, n-2k} \tag{5.3}$$

Here, K = No. of Parameters estimated (k = p + q + 1, if intercept is included) for larger calculated value of F the H_o many not be accepted

6. Modified model selection criteria for time series models:

Generally, there will be several plausible models that one can select to use for forecasts. It should be noted that best model fit need not imply to provide best forecasts in model selection criteria [19].

If one wishes to choose between an AR (1) and an MA (2) one may estimate an ARMA (1, 2) and then restrict the MA (2) coefficients to equal zero. On the other hand one may restrict the AR (1) coefficients to equal zero. But, this method is unsatisfactory because it necessitates estimating the over parameterized ARMA (1, 2) model. Instead, one can use model selection criteria to choose between alternative models. Such model selection criteria can be noticed as goodness of fit measures. In parameter estimation, forecast error variance increases as a result of errors arising and it is not desirable to consider it for Time series in practice

The Akaike Information Criteria (AIC) and the Schwartz Bayesian Criterion (SBC) based on the concept of Mean Squared Prediction or Forecast Error (MSPE).

Mean Squared Prediction Error:

Consider the true data – generating process AR(1) model as

$$y_t = \alpha y_{t-1} + \in_t \tag{6.1}$$

If α is known, the forecast with y_{t+1} is

$$E_t\left(y_{t+1}\right) = \alpha y_t \tag{6.2}$$

The mean squared prediction error is given by

$$E_t\left[y_{t+1} - \alpha y_t\right] = E_t\left(\in_{t+1}^2\right) = \sigma^2 \tag{6.3}$$

when α is estimated from the data, the one-stop-ahead forecast of y_{t+1} is given by

$$E\left[y_{t+1}\right] = \hat{\alpha}\, y_t \tag{6.4}$$

Where $\hat{\alpha}$ is an estimate of α Now, the MSPE is given by,

$$\text{MSPE} = E_t\left(y_{t+1} - 2y_t\right)^2 = E_t\left[\left(\alpha y_t - \hat{\alpha} y_t\right) + \in_{t+1}\right]^2 \tag{6.5}$$

Since, $\in_{t+1}$ is independent of $\hat{\alpha}$ and y_t, one may obtain,

$$E_t\left(y_{t+1} - \hat{\alpha} y_t\right)^2 = E_t\left[\left(\alpha - \hat{\alpha}\right) y_t\right]^2 + \sigma^2 \simeq E_t\left[\left(\alpha - \hat{\alpha}\right)\right]^2 \left(y_t\right)^2 + \sigma^2 \tag{6.6}$$

Since, Var $\left(y_t\right) = \dfrac{\sigma^2}{1-\alpha^2}$ and in large samples,

$\text{Var}\left(\hat{\alpha}\right) = E_t\left[\left(\alpha - \hat{\alpha}\right)\right]^2 \simeq \dfrac{1-\alpha^2}{n}$, one may obtain,

$$E_t\left[\left(\alpha - \hat{\alpha}\right)\right]^2 \left(y_t\right)^2 + \sigma^2 \simeq \left[\frac{\left(1-\alpha^2\right)}{n}\right]\left[\frac{\sigma^2}{1-\alpha^2}\right] + \sigma^2 = \left[1 + \left(\frac{1}{n}\right)\right]\sigma^2 \tag{6.7}$$

It shows that as n increases the MSPE approaches σ^2. Here 'n' denotes the number of usable observations under the Finite Prediction Error (FPE) criterion, one may minimize the one-step-ahead MSPE. Then AR (P) process as

$$y_t = \sum \alpha_i y_{t-i} + \alpha_p y_{t-p} + \in_t \tag{6.8}$$

One may obtain the MSPE for this AR(p) process as

$$\text{MSPE} = \left[1 + \left(\frac{p}{n}\right)\right]\sigma^2. \tag{6.9}$$

Generally, σ^2 is unknown.

However σ^2 can be replaced with an estimate,

$$\tilde{\sigma}^2 = \frac{\sum_{r=1}^{n} e_t^{*2}}{n-p} \tag{6.10}$$

Where $\sum_{r=1}^{n} e_t^{*2}$ is the Internally studentized residuals sum of squares obtained by using AR (P) model.

Now, the modified FPE is given by

$$\text{FPE} = \left[1 + \left(\frac{p}{n}\right)\right]\left[\sum_{t=1}^{n} e_t^{*2} / n - p\right] \tag{6.11}$$

IOP Conf. Series: Materials Science and Engineering **263** (2017) 042135 doi:10.1088/1757-899X/263/4/042135

One can select p by minimizing FPE. By using logarithms one can approximate $\text{Ln}\left(1+\frac{p}{n}\right)$ by $\left(\frac{p}{n}\right)$. Thus, it is possible to select p to minimize the modified

FPE as $\frac{p}{n}+\ln\left[\sum_{i=1}^{n} e_t^{*2}\right]-\ln(n-p)$

Further, it is the same to minimize

$$p+n\ln\left[\sum_{t=1}^{n} e_t^{*2}\right]-(n)(\ln^*(n-(p)))$$

Since, $\ln(n-p)\simeq\ln[n-\frac{p}{n}]$, p may be selected by minimizing

$$p+(n)\ln\left(\sum_{t=1}^{n} e_t^{*2}\right)-\ln(n)+p$$

Generally, the AIC selects the K(=p+q+1) parameters of an ARMA (p, q) model so as to maximize the log – likelihood function including a penalty for each parameter estimated.

AIC = -2 ln maximized value of log likelihood + $\frac{K}{n}$ (6.12)

For a given value of n, selecting the values of p and q, so as to minimize ln (AIC) is equivalent to selecting p and q so as to minimize the sum

$$n\ \ln\left[\sum_{t=1}^{n} e_t^{*2}\right]+2K.$$

Now, the modified AIC is given by

$$[\text{AIC}]_* = n\ln\left[\sum_{t=1}^{n} e_t^{*2}\right]+2k \qquad (6.13)$$

Where k = number of parameters estimated

The SBC incorporates larger penalty $(p+q+1)l_n\ n$. under SBC the values of p and q may be selected so as to minimize

$$n\ l_n\left[\sum_{t=1}^{n} e_t^{*2}\right]+k\ l_n(n)$$

The modified SBC is given by

$$[\text{SBC}]_* = n\ l_n\left[\sum_{t=1}^{n} e_t^{*2}\right]+2k$$

Here Model I is better than Model II and modified AIC < Model II.

One may be quite confident in results, of both the $[\text{AIC}]_*$ and $[\text{SBC}]_*$ select the same Time series model. If they select different models, then one need to proceed cautiously.

7. Conclusions

Time series is becoming increasingly important both for the regulation of developed economics as well as for the planning of the economic development of underdeveloped countries. In framing policy decisions, It is important to be able to predict the value of the economic magnitudes. Such forecasts will be enable the policy maker to judge whether it is necessary to take any measure in order to influence the relevant economic variables. In the present study, a goodness of fit criterion for ARMA (p,q) model and modified selection criteria for Time series models have been suggested.

IOP Conf. Series: Materials Science and Engineering **263** (2017) 042135 doi:10.1088/1757-899X/263/4/042135

References

[1] Ahmad M H and Ismail Z 1999, Dept. *of Mathematics Technical Report* **399**
[2] Baki Md Billah Rob J Hyndman and Anne B.Kochler 2003 *Business and Economics Forecasting 2*1-19
[3] Billah M B Hyndman R J and Koehler A B 2005 ,*Journal of Statistical Computation and Simulation* **75**831-840
[4] Bowerman B L and O'Connell R T 1979. *DuxbUry Press* Belmont California
[5] Enders W 2005 *Uplied econometric time series*seventh edition John Willey and sons
[6] Etuk E H Natamba B 2015 *International Journal of Life Science and Engineering* **1** 165–170
[7] Hens N Aerts M and Molenberghs G 2006*Statistics in Medicine* **25** 2502-2520
[8] Huizou and Yuhong Yang 2004*International Journal of Forecastimg***5**69-84
[9] Inoue A and Kilian L 2006 *Journal of Econometrics* 130 273-306
[10] John Geweke and Charles whiteman 2005*Bayesian Forecasting* Wiley New York
[11] Kedem B and Fokianos K 2002*Regression Models for Time Series Analysis*John Wiley & Sons New York.
[12] Makridakis S G Wheelwrigh S C McGee V E 983*Forecasting: Methods and Applications* **2nd ed** New York Wiley
[13] Meade N 2000 *Journal of Forecasting* **19** 125-136
[14] Meade N 2000*Journal of Forecasting* **19**515-535
[15] Popale P G and Gorantiwar S D 2014*International Journal of Innovative Research in Science, Engineering and Technology* **3**85–196
[16] Ronchetti E 1985*Statistics and Probability* **3** 21-23