Least Squares and Discounted Least Squares in Autoregressive Process

Phongsaphen Jantana and
Prapaisri Sudasna-na-Ayudthya

Abstract

This paper reports on a comparative study between the least squares (LS) method using rank-one update QR factorization and the discounted least squares with direct smoothing. Both approaches are applied to solve the problem of updating estimated parameters in long rolling periods of time-series forecasting using the autoregressive (AR) process. The corresponding model used in the experiment is undamped sinusoidal data with various autoregressive orders. The results of the study indicate that both methods improved effectively as the model order grow. However, the discounted least squares with direct smoothing had a more increasing rate of improvement.

Keywords : Autoregressive process; Estimating parameter; Least squares; Discounted Least squares

Introduction

In many industries, forecasting has become an important tool to reduce risk from uncertainty. Techniques can be applied to forecast demand in order to plan the production schedule or plan for reserving inventory or ordering material from suppliers. Such data are time-related and highly correlated so Time Series Analysis is a particularly significant tool. This study focused on one specific type of Time Series Model: the Autoregressive (AR) model. The AR model was developed by Box and Jenkins in 1970 (Box, 1994) to analyze historical data that had relations within itself.

In this study, the parameters were obtained from univariate stationary autoregressive data (Kendall, 1976) when the population mean of the process was not zero. The AR process utilizes the least squares (LS) method, and it is an analogous way to fit a model by minimizing the sum of square errors for estimating parameters. The LS method uses the normal equations to implement the linear system. The parameters can be solved by QR factorization.

The earlier observations in LS method receive the same weight as recent observations. However, the recent observations may be more important for the true behavior of the process so that, in discounted least squares method, the older observations receive proportionally less weight than the recent ones. Another aspect is that when periods roll for a long time, the parameters of the time series model must be changed for updating the model. The disadvantage in the use of LS as a forecasting method is that the estimated parameters need to be updated at the end of each period. This study attempted to reduce the computation time by using the discounted least squares method with direct smoothing to fit an AR model. Brown (1962) developed the discounted least squares method with direct smoothing for estimating regression model parameters.

The purpose of this study was to implement the discounted least squares method with direct smoothing for estimating autoregressive model parameters which were not found in previous research. As the observation period shifts, the direct smoothing can update the old estimates of the model parameters by smoothing them with the forecast error for the current period to obtain the revised estimates. In addition, a comparison between the LS method using rank-one update QR factorization and the discounted least squares with direct smoothing was performed. Cost and accuracy of forecasting were two aspects to be considered in comparing the methods.

Methods

In this study, the linear stochastic stationary time series data were fitted to an AR model by using two methods : the LS method and the discounted least squares method. In addition, we assumed that the population mean was known and not to be zero. We also considered the process by using two systems of equation :

$$x_t = \phi_0 + \sum_{t=1}^{T} \phi_i x_{t-i} + \varepsilon_t, \qquad i = 1,2,\ldots,p, \qquad (1)$$

$$y_t = \sum_{t=1}^{T} \phi_i y_{t-i} + \varepsilon_t, \qquad i = 1,2,\ldots,p, \qquad (2)$$

where x_t is an observation at time t , ϕ_0 is the scalar, and ϕ_i are the unknown parameters. When $\hat{x}_t$ is a forecasting value at time t, it has a random error ε_t that $E(\varepsilon_t) = 0$ and $V(\varepsilon_t) = \sigma_\varepsilon^2$. Equation (2) was obtained by defining $y = x - \mu_x$. An AR model is simply a linear regression of the current value of the series against one or more prior values of the series. The value of p is called order of AR model.

The least squares method : In LS method, the AR model parameters in equation (1) are estimated by minimizing the error sum of squares :

$$\text{Min } \boldsymbol{l}(\phi) = \sum_{t=1}^{T} \varepsilon_t^2 .$$

It gives the linear systems equation $Ax = b$ from least squares normal equation as follows :

$$\begin{bmatrix} \sum_{t=p+1}^{T} 1 & \sum_{t=p+1}^{T} x_{t-1} & \cdots & \sum_{t=p+1}^{T} x_{t-p} \\ \sum_{t=p+1}^{T} x_{t-1} & \sum_{t=p+1}^{T} x_{t-1}^2 & \cdots & \sum_{t=p+1}^{T} x_{t-1}x_{t-p} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{t=p+1}^{T} x_{t-p} & \sum_{t=p+1}^{T} x_{t-1}x_{t-p} & \cdots & \sum_{t=p+1}^{T} x_{t-p}^2 \end{bmatrix} \begin{bmatrix} \hat{\phi}_0 \\ \hat{\phi}_1 \\ \vdots \\ \hat{\phi}_p \end{bmatrix} = \begin{bmatrix} \sum_{t=p+1}^{T} x_t \\ \sum_{t=p+1}^{T} x_t x_{t-1} \\ \vdots \\ \sum_{t=p+1}^{T} x_t x_{t-p} \end{bmatrix}, \quad (3)$$

where "A" is a square matrix formed by using the sum of backward elements at a time, while "x" is a parameter vector. "b" is a right-hand side vector, and T is the end of observation time.

When $y = x-\mu_x$, the system equation gives the linear systems equation $Ax = b$ from least squares normal equation as follows :

$$\begin{bmatrix} \sum_{t=p+1}^{T} y_{t-1}^2 & \sum_{t=p+1}^{T} y_{t-1}y_{t-2} & \cdots & \sum_{t=p+1}^{T} y_{t-1}y_{t-p} \\ \sum_{t=p+1}^{T} y_{t-1}y_{t-2} & \sum_{t=p+1}^{T} y_{t-2}^2 & \cdots & \sum_{t=p+1}^{T} y_{t-2}y_{t-p} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{t=p+1}^{T} y_{t-1}y_{t-p} & \sum_{t=p+1}^{T} y_{t-2}y_{t-p} & \cdots & \sum_{t=p+1}^{T} y_{t-p}^2 \end{bmatrix} \begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \end{bmatrix} = \begin{bmatrix} \sum_{t=p+1}^{T} y_t y_{t-1} \\ \sum_{t=p+1}^{T} y_t y_{t-2} \\ \vdots \\ \sum_{t=p+1}^{T} y_t y_{t-p} \end{bmatrix}. \quad (4)$$

The first and second linear systems equations are solved using the QR factorization (Golub and Van Loan, 1996). It is defined as follows :

$$A_T x_T = b_T \,.$$

Since $A_T = Q_T R_T$ then

$$Q_T R_T x_T = b_T \,,$$

$$Q_T^{-1} Q_T R_T x_T = Q_T^{-1} b_T \,,$$

and

$$Q_T' = Q_T^{-1}$$

$$R_T x_T = Q_T' b_T$$

The results from this method were used as the model parameters. The AR models were obtained and the forecast errors computed to validate the result.

In time series analysis, data are always shifted to the next period because the parameters must have new estimated values to update the model. The parameters are updated by calculating the next estimated values that employ rank one update QR factorization as defined by Householder (Cavallo et al., 1996). Suppose that the $Q_T R_T = A_T \in \mathbb{R}^{m \times n}$ and $A_{T+1} = A_T + u_T v'_T = Q_{T+1} R_{T+1}$ where u , $v \in \mathbb{R}^n$ are given, such that

$$A_{T+1} = A_T + u_T v'_T = Q_T (R_T + w_T v') ,$$

where $w_T = Q'_T u_T$. Compute the rotation $J'_1 \dots J'_{n-1} w_T = \|w_T\|_2 e_1$. Each J_k is the rotation in plane k and $k+1$. The given rotations applied to R are shown as (Golub and Van Loan, 1996) :

$$H = J'_1 \dots J'_{n-1} R_T . \tag{5}$$

Consequently,

$$J'_1 \dots J'_{n-1} (R_T + w_T v'_T) = H \pm \| w_T \|_2 e_1 v'_T = H_1 . \tag{6}$$

Given rotations, G_k, $k = 1 : n - 1$ such that

$$G'_{n-1} \dots G'_1 H_1 = R_{T+1} \quad . \tag{7}$$

Then obtained is the QR factorization $A_{T+1} = A_T + u_T v'_T = Q_{T+1} R_{T+1}$, where

$$Q_{T+1} = Q_T J_{n-1} \dots J_1 G_1 \dots G_{n-1} . \tag{8}$$

Now, Q_{T+1} and R_{T+1} are obtained from updated Q_T and R_T.

The discounted least squares: The recent observations may be more important than the older observations. This procedure is the weighted least squares that gives a lower weight of older observations than of the recent ones. From equation (2) $y_t = \sum_{t=1}^{T} \phi_i y_{t-i} + \varepsilon_t$ where W_{tt} is the square root of the weight given to the t^{th} error (Montgomery, 1990). To minimize discounted sum of square of errors and obtain the estimated parameters, the objective function can be stated as

$$\text{Min } l(\phi) = \sum_{t=1}^{T} W_{tt}^2 \varepsilon_t^2,$$

W is the $T \times T$ diagonal matrix of square roots of the weights.

$$W = \begin{bmatrix} W_{11} & 0 & \cdots & 0 \\ 0 & W_{22} & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & \cdots & W_{TT} \end{bmatrix}.$$

The weighted sum of square error is

$$l(\phi) = (W\,\varepsilon)'(W\,\varepsilon)$$
$$l(\phi) = [\,y - Y(T)\,\boldsymbol{\phi}]' W^2 [y - Y(T)\,\boldsymbol{\phi}].$$

Then, the least squares normal equations can be written as :

$$Y'(T)\,\mathbf{W}^2\,Y(T)\,\hat{\boldsymbol{\phi}}_T = Y(T)\,\mathbf{W}^2\,y\,.$$

Since $\mathbf{G}(T) = [W\,Y(T)]'\,[W\,Y(T)]\,,$

and $g_i(T) = Y'(T)\,W^2\,y\,,$

then $\mathbf{G}(T)\,\hat{\boldsymbol{\phi}}_T = g(T)\,,$ (9)

and the solution is $\hat{\boldsymbol{\phi}}_i(T) = \boldsymbol{G}^{-1}(T)\,g(T)\,.$

L is the transition matrix (Montgomery, 1990) of the system equation (2), that is

$$L = \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix} .$$

The transition property for system equation (2) can be described as

$$y(t+1) = L\, y(t) .$$

Let the weights W_{tt}^2 be defined as $W_{T-j,T-j}^2 = \beta^j$ when $j = 0,1,\ldots, T-1$ and $0 < \beta < 1$

then $$G(T) = \sum_{j=0}^{T-1} \beta^j y(-j) y'(-j) .$$

The matrix $G(T)$ approaches a limit G, where

$$G \equiv \lim_{T\to\infty} G(T) = \sum_{j=0}^{\infty} \beta^j y(-j) y'(-j) .$$

Therefore G^{-1} needs to be computed only once. The right-hand side of the normal equation may be written as :

$$g(T) = \sum_{j=0}^{T-1} \beta^j y_{T-j}\, y(j) ,$$

or

$$g(T) = y_T y(0) + \beta L^{-1} g(T-1) .$$

Substituting for $g(T)$ in equation (9), then we obtain

$$\hat{\phi}_i(T) = y_T G^{-1} y(0) + \beta G^{-1} L^{-1} G \phi(T-1) . \qquad (10)$$

Since $h = G^{-1}y(0)$ and $H = \beta G^{-1}L^{-1}G$, then equation (10) may be written as

$$\hat{\phi}_i(T) = hy_T + H\phi(T-1)\ .$$

To simplify $\hat{\phi}_i(T)$

$$L^{-1}G = L^{-1}G(L')^{-1}G\ ,$$

$$L^{-1}G = \sum_{j=0}^{\infty} \beta^j L^{-1}y(-j)y'(-j)(L')^{-1}(L')\ ,$$

$$= \beta^{-1}(G - y(0)y'(0))L'\ .$$

Substituting $L^{-1}G$ in H, we now have

$$H = (I - G^{-1}y(0)y'(0))L'.$$

Substituting $h = G^{-1}y(0)$ in H, we obtain

$$H = L' - hy'(0)L'\ ,$$

and the single period error $e_1(T)$ is given by

$$e_1(T) = y_T - \hat{y}_T(T-1)\ ,$$

so $\hat{\phi}_i(T)$ now becomes

$$\hat{\phi}_i(T) = L'\hat{\phi}_i(T-1) + he_1(T)\ . \qquad (11)$$

The discounted least squares estimate the model parameters shown in equation (11). This method used to estimate the parameters at the end of T^{th} period is a linear combination of estimates made at the end of the previous period and the single-period forecast error.

The transition matrix (L) of equation (1) is of the form

$$L = \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & \cdots & \phi_{p-1} & \phi_p & \phi_0 \\ 1 & 0 & 0 & \cdots & 0 & 0 & 1 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix}.$$

The steps of discounted least squares method to obtain the model parameters can be described as follows :

Step 1: Calculate $G(T)$ when T=15,000 (assume period 15,000th $\rightarrow \infty$) and the optimal weight is defined.
Step 2: Calculate $G^{-1}(T)$.
Step 3: Obtain ϕ_T.
Step 4: Forecast one step ahead.
Step 5: Calculate $e_{T+1} = x_{T+1} - \hat{x}_{T+1}$.
Step 6: Obtain h_T.
Step 7: Obtain L_T.
Step 8: Obtain ϕ_{T+1} .

In this study, the undamped sinusoids autoregressive data (Martinez, 2002) were generated with four orders of 5, 12, 20 and 30. Therefore, AR(5), AR(12), AR(20), and AR(30) were investigated and the forecasts of those models were started at 15,000th observation by forecasting one step ahead. Then, repeated the step of forecasting one step ahead for 1000, 5000, 10000, 15000, 20000 and 25000 times. MATLAB 7.0 was used to generate the data while Minitab used to check the AR property. In addition, the QR factorization and the QR rank one update were applied by the algorithm in equations (6) - (9) using MATLAB function. The LS method and the discounted least squares method with direct smoothing were written as MATLAB M-file. The results were evaluated by measuring the computation time and prediction mean square error to compare computing cost and accuracy of the two forecast methods, respectively.

Results and discussion

The results for long rolling period forecasting of the least squares method and the discounted least squares method indicate the preference to the least squares. Table 1 illustrates the comparison via the operating time, while Table 2 illustrates the comparison via the prediction mean square error. The least squares method with QR factorization had both less operation time and less prediction mean square errors. However, as the order of the model grows, the variances of data decrease. Thus, for both methods, the prediction mean square errors were smaller. The discounted least squares with direct smoothing method gave higher decreasing rate in prediction mean square errors as the order of the model grew.

The update QR function in MATLAB has an accelerated tool but the direct smoothing uses only MATLAB language in M-file, and the discounted least squares with direct smoothing adjusts the transition matrix in each period. For these reasons, the discounted least squares method consumed more time than the other method. Besides, the selected weight may not be appropriated for the discounted least squares method and results in the higher prediction mean square error.

Table 1 Comparison of time of calculations to estimate parameters obtained for long rolling periods when using two methods.

AR(p)	No. of forecasts	Time (sec.) Least squares with QR factorization Eq. (1)	Least squares with QR factorization Eq. (2)	Discounted least squares with Direct smoothing Eq. (1)	Discounted least squares with Direct smoothing Eq. (2)
AR(5)	1000	0.621	1.052	1.131	1.783
	5000	0.961	2.392	2.223	3.134
	10000	1.181	5.248	4.356	5.718
	15000	1.422	8.272	7.691	11.537
	20000	1.733	12.097	11.456	17.435
	25000	2.163	16.393	16.374	23.954
AR(12)	1000	0.691	1.132	1.532	1.732
	5000	1.221	2.454	2.353	3.314
	10000	1.513	5.097	4.947	6.429
	15000	1.853	9.063	7.030	12.248
	20000	2.273	12.888	12.307	18.016
	25000	2.704	17.555	16.464	24.916
AR(20)	1000	0.862	1.552	1.813	2.293
	5000	1.382	2.884	3.104	3.545
	10000	2.043	6.179	5.697	6.349
	15000	2.564	9.113	8.352	13.039
	20000	3.134	14.541	11.136	19.368
	25000	3.896	17.746	17.946	25.147
AR(30)	1000	1.222	1.763	2.804	3.004
	5000	1.873	3.715	4.877	6.269
	10000	3.025	7.411	8.122	9.914
	15000	3.495	11.697	12.868	16.413
	20000	4.337	16.323	17.953	23.584
	25000	5.127	21.761	23.754	31.916

Table 2 Comparison of prediction mean square error of calculations to estimate parameters obtained for long rolling periods when using two methods.

AR(p)	No. of forecasts	Prediction Mean Square Error			
		Least squares with QR factorization		Discounted least squares with Direct smoothing	
		Eq. (1)	Eq. (2)	Eq. (1)	Eq. (2)
AR(5)	1000	15.80000	15.80000	30.80000	42.10000
	5000	14.70000	14.70000	29.50000	55.40000
	10000	14.60000	14.60000	31.10000	60.50000
	15000	14.60000	14.60000	31.00000	60.50000
	20000	14.60000	14.60000	30.70000	60.30000
	25000	14.60000	14.60000	30.80000	61.80000
AR(12)	1000	10.90000	10.90000	14.40000	20.00000
	5000	10.20000	10.20000	18.10000	109.90000
	10000	10.10000	10.10000	32.00000	98.30000
	15000	10.10000	10.10000	27.60000	89.10000
	20000	10.10000	10.10000	37.90000	98.40000
	25000	10.10000	10.10000	47.70000	118.00000
AR(20)	1000	0.68733	0.68747	0.75163	0.80624
	5000	0.64041	0.64034	0.83467	1.80000
	10000	0.63398	0.63394	1.00000	1.80000
	15000	0.63456	0.63452	0.97123	1.70000
	20000	0.63402	0.63400	1.10000	1.90000
	25000	0.63398	0.63403	1.20000	2.00000
AR(30)	1000	0.68706	0.68730	0.70959	0.71029
	5000	0.64085	0.64075	0.73606	1.10000
	10000	0.63442	0.63415	0.80181	1.10000
	15000	0.63483	0.63484	0.7871	1.00000
	20000	0.63497	0.63435	0.80071	1.10000
	25000	0.63430	0.63434	0.82564	1.10000

Conclusion

When forecasting a stationary autoregressive process for long rolling periods with $\mu \neq 0$, the least square with QR factorization is appropriated, leading to higher efficiency than the discounted least squares with direct smoothing. The autoregressive model as in equation (1) had more efficiency for both methods than the model in equation (2). When the order of the model grew, both methods had more accuracy in forecasting since the prediction mean square errors were decreased. The discounted least squares with direct smoothing used more operation time since the transition matrix was adjusted in every period.

For further study, the transition matrix shoud be tried without adjustment by time, and another weight approach should be tested.

References

Box, G.P. et al. (1994) *Time Series Analysis Forecasting and Control*, Prentice Hall, New Jersey.

Brown, R.G. (1962) *Smoothing, Forecasting and Prediction of Discrete Time Series*, Prentice-Hall Inc., New Jersey.

Cavallo, A. et al. (1996) *Using MATLAB, SIMULINK and Control System Toolbox*, Prentice-Hall Inc., London.

Golub, G.H. and Van Loan, C.F. (1996) *Matrix Computations*, The Johns Hopkins Press, London.

Hannan, E.J. et al. (1995) *Time Series in Time Domain*, Elsevier Science B.V., Netherlands.

Kendall, M. (1976) *Time Series*, Charles Griffin and Company Ltd., London.

Martinez,W.L. and Martinez, A.R. (2002) *Computational Statistics Handbook with MATLAB*, Chapman & Hall/CRC, New York.

Montgomery, D.C. et al. (1990) *Forecasting and Time Series Analysis*, McGraw-Hill Inc., Singapore.