ARIMA Algorithms

The ARIMA procedure computes the parameter estimates for a given seasonal or non-seasonal univariate ARIMA model. It also computes the fitted values, forecasting values, and other related variables for the model.

Notation

The following notation is used throughout this chapter unless otherwise stated:

y_t (t=1, 2, ..., N)	Univariate time series under investigation.
N	Total number of observations.
a_t (t = 1, 2, ... , N)	White noise series normally distributed with mean zero and variance σ_a^2.
p	Order of the non-seasonal autoregressive part of the model
q	Order of the non-seasonal moving average part of the model
d	Order of the non-seasonal differencing
P	Order of the seasonal autoregressive part of the model
Q	Order of the seasonal moving-average part of the model
D	Order of the seasonal differencing
s	Seasonality or period of the model
$\phi_p(B)$	AR polynomial of B of order p, $\phi_p(B) = 1 - \varphi_1 B - \varphi_2 B^2 - ... - \varphi_p B^p$
$\theta_q(B)$	MA polynomial of B of order q, $\theta_q(B) = 1 - \vartheta_1 B - \vartheta_2 B^2 - ... - \vartheta_q B^q$
$\Phi_P(B^s)$	Seasonal AR polynomial of BS of order P, $\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{s2} - ... - \Phi_P B^{sP}$
$\Theta_Q(B^s)$	Seasonal MA polynomial of BS of order Q, $\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{s2} - ... - \Theta_Q B^{sQ}$
Δ	Differencing operator $\Delta = (1-B)^d (1-B^s)^D$
B	Backward shift operator with $BY_t = Y_{t-1}$ and $Ba_t = a_{t-1}$

Models

A seasonal univariate ARIMA$(p,d,q)(P,D,Q)_S$ model is given by

$$\Phi(B)[\Delta y_t - \mu] = \Theta(B)a_t \qquad t = 1, \ldots, N$$

where

$$\Phi(B) = \phi_p(B)\,\Phi_P(B)$$

$$\Theta(B) = \theta_q(B)\,\Theta_Q(B)$$

and μ is an optional model constant. It is also called the stationary series mean, assuming that, after differencing, the series is stationary. When NOCONSTANT is specified, μ is assumed to be zero.

An optional log scale transformation can be applied to y_t before the model is fitted. In this chapter, the same symbol, y_t, is used to denote the series either before or after log scale transformation.
Independent variables x_1, x_2, …, x_m can also be included in the model. The model with independent variables is given by

$$\Phi(B)\left[\Delta\left(y_t - \sum_{i=1}^{m} c_i x_{it}\right) - \mu\right] = \Theta(B)a_t$$

where

$c_i, i = 1, 2, \ldots, m$, are the regression coefficients for the independent variables.

Estimation

Basically, two different estimation algorithms are used to compute maximum likelihood (ML) estimates for the parameters in an ARIMA model:

- **Melard's algorithm** is used for the estimation when there is no missing data in the time series. The algorithm computes the maximum likelihood estimates of the model parameters. The details of the algorithm are described in (Melard, 1984), (Pearlman, 1980), and (Morf, Sidhu, and Kailath, 1974).
- A **Kalman filtering algorithm** is used for the estimation when some observations in the time series are missing. The algorithm efficiently computes the marginal likelihood of an ARIMA model with missing observations. The details of the algorithm are described in the following literature: (Kohn and Ansley, 1986) and (Kohn and Ansley, 1985).

Initialization of ARMA parameters

The ARMA parameters are initialized as follows:

Assume that the series Y_t follows an ARMA(p,q)(P,Q) model with mean 0; that is:

$$Y_t - \varphi_1 Y_{t-1} - \cdots - \varphi_p Y_{t-p} = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

In the following c_l and ρ_l represent the lth lag autocovariance and autocorrelation of Y_t respectively, and $\hat{c}_l$ and $\hat{\rho}_l$ represent their estimates.

Non-seasonal AR parameters

For AR parameter initial values, the estimated method is the same as that in appendix A6.2 of (Box, Jenkins, and Reinsel, 1994). Denote the estimates as $\hat{\varphi}_1^{'}, \cdots, \hat{\varphi}_{p+q}^{'}$.

Non-seasonal MA parameters

Let

$$w_t = Y_t - \varphi_1 Y_{t-1} - \cdots - \varphi_p Y_{t-p} = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

The cross covariance

$$\lambda_l = E(w_{t+l} a_t) = E((a_{t+l} - \theta_1 a_{t+l-1} - \cdots - \theta_q a_{t+l-q}) a_t) = \begin{cases} \sigma_a^2 & l = 0 \\ -\theta_1 \sigma_a^2 & l = 1 \\ \cdots & \cdots \\ -\theta_q \sigma_a^2 & l = q \\ 0 & l > q \end{cases}$$

Assuming that an AR(p+q) can approximate Y_t, it follows that:

$$Y_t - \varphi'_1 Y_{t-1} - \cdots - \varphi'_p Y_{t-p} - \varphi'_{p+1} Y_{t-p-1} - \cdots - \varphi'_{p+q} Y_{t-p-q} = a_t$$

The AR parameters of this model are estimated as above and are denoted as $\hat{\varphi}_1^{'}, \cdots, \hat{\varphi}_{p+q}^{'}$.

Thus λ_l can be estimated by

$$\begin{aligned} \lambda_l &\approx E\Big((Y_{t+l} - \varphi_1 Y_{t+l-1} - \cdots - \varphi_p Y_{t+l-p}) \left(Y_t - \varphi'_1 Y_{t-1} - \cdots - \varphi'_{p+q} Y_{t-p-q} \right) \Big) \\ &= \left(\rho_l - \sum_{j=1}^{p+q} \varphi_j \rho_{l+j} - \sum_{i=1}^{p} \varphi_i \rho_{l-i} + \sum_{i=1}^{p} \sum_{j=1}^{p+q} \varphi_i \varphi_j \rho_{l+j-i} \right) c_0 \end{aligned}$$

And the error variance σ_a^2 is approximated by

$$\hat{\sigma}_a^2 = Var\left(-\sum_{j=0}^{p+q} \varphi'_j Y_{t-j} \right) = \sum_{i=0}^{p+q} \sum_{j=0}^{p+q} \varphi'_i \varphi'_j c_{i-j} = c_0 \sum_{i=0}^{p+q} \sum_{j=0}^{p+q} \varphi'_i \varphi'_j \rho_{i-j}$$

with $\hat{\varphi}_0^{'} = -1$.

Then the initial MA parameters are approximated by $\theta_l = -\lambda_l / \sigma_a^2$ and estimated by

$$\hat{\theta}_l = -\hat{\lambda}_l / \hat{\sigma}_a^2 = \frac{\rho_l - \sum_{j=1}^{p+q} \hat{\varphi}_j \rho_{l+j} - \sum_{i=1}^{p} \hat{\varphi}_i \rho_{l-i} + \sum_{i=1}^{p} \sum_{j=1}^{p+q} \hat{\varphi}_i \hat{\varphi}_j \rho_{l+j-i}}{\sum_{i=0}^{p+q} \sum_{j=0}^{p+q} \hat{\varphi}_i^{'} \hat{\varphi}_j^{'} \rho_{i-j}}$$

So $\hat{\theta}_l$ can be calculated by $\hat{\varphi}_j^{'}$, $\hat{\varphi}_i$, and $\{\hat{\rho}_l\}_{l=1}^{p+2q}$. In this procedure, only $\{\hat{\rho}_l\}_{l=1}^{p+q}$ are used and all other parameters are set to 0.

Seasonal parameters

For seasonal AR and MA components, the autocorrelations at the seasonal lags in the above equations are used.

Diagnostic Statistics

The following definitions are used in the statistics below:

N_p	Number of parameters. $N_p = \begin{cases} p+q+P+Q+m & \text{without model constant} \\ p+q+P+Q+m+1 & \text{with model constant} \end{cases}$
SSQ	Residual sum of squares $SSQ = \mathbf{e}^{'}\mathbf{e}$, where **e** is the residual vector
$\hat{\sigma}_a^2$	Estimated residual variance. $\hat{\sigma}_a^2 = \frac{SSQ}{df}$, where $df = N - N_p$
SSQ'	Adjusted residual sum of squares. $SSQ' = (SSQ)\|\Omega\|^{1/N}$, where Ω is the theoretical covariance matrix of the observation vector computed at MLE

Log-Likelihood

$$L = -N\ln(\hat{\sigma}_a) - \frac{SSQ'}{2\hat{\sigma}_a^2} - \frac{N\ln(2\pi)}{2}$$

Akaike Information Criterion (AIC)

$$AIC = -2L + 2N_p$$

Schwartz Bayesian Criterion (SBC)

$$SBC = -2L + \ln(N)N_p$$

Generated Variables

Predicted Values

Forecasting Method: Conditional Least Squares (CLS or AUTOINT)

In general, the model used for fitting and forecasting (after estimation, if involved) can be written as

$$y_t - D(B)\,y_t = \Phi(B)\mu + \Theta(B)a_t + \sum_{i=1}^{m} c_i\Phi(B)\Delta x_{it}$$

where

$$D(B) = \Theta(B)\Delta - 1$$

$$\Phi(B)\mu = \Phi(1)\mu$$

Thus, the predicted values $(FIT)_t$ are computed as follows:

$$(FIT)_t = \hat{y}_t = D(B)\,\hat{y}_t + \Phi(B)\,\mu + \Theta(B)\,\hat{a}_t + \sum_{i=1}^{m} c_i \Phi(B)\,\Delta x_{it}$$

where

$$\hat{a}_t = y_t - \hat{y}_t \quad 1 \le t \le n$$

Starting Values for Computing Fitted Series. To start the computation for fitted values, all unavailable beginning residuals are set to zero and unavailable beginning values of the fitted series are set according to the selected method:

CLS. The computation starts at the (d+sD)-th period. After a specified log scale transformation, if any, the original series is differenced and/or seasonally differenced according to the model specification. Fitted values for the differenced series are computed first. All unavailable beginning fitted values in the computation are replaced by the stationary series mean, which is equal to the model constant in the model specification. The fitted values are then aggregated to the original series and properly transformed back to the original scale. The first d+sD fitted values are set to missing (SYSMIS).

AUTOINIT. The computation starts at the [d+p+s(D+P)]-th period. After any specified log scale transformation, the actual d+p+s(D+P) beginning observations in the series are used as beginning fitted values in the computation. The first d+p+s(D+P) fitted values are set to missing. The fitted values are then transformed back to the original scale, if a log transformation is specified.

Forecasting Method: Unconditional Least Squares (EXACT)

As with the CLS method, the computations start at the (d+sD)-th period. First, the original series (or the log-transformed series if a transformation is specified) is differenced and/or seasonally differenced according to the model specification. Then the fitted values for the differenced series are computed. The fitted values are one-step-ahead, least-squares predictors calculated using the theoretical autocorrelation function of the stationary autoregressive moving average (ARMA) process corresponding to the differenced series. The autocorrelation function is computed by treating the estimated parameters as the true parameters. The fitted values are then aggregated to the original series and properly transformed back to the original scale. The first d+sD fitted values are set to missing (SYSMIS). The details of the least-squares prediction algorithm for the ARMA models can be found in (Brockwell and Davis, 1991).

Residuals

Residual series are always computed in the transformed log scale, if a transformation is specified.

$$(ERR)_t = y_t - (FIT)_t \quad t = 1, 2, \ldots, N$$

Standard Errors of the Predicted Values

Standard errors of the predicted values are first computed in the transformed log scale, if a transformation is specified.

Forcasting Method: Conditional Least Squares (CLS or AUTOINIT)

$$(SEP)_t = \hat{\sigma}_a \quad t = 1, 2, \ldots, N$$

Forecasting Method: Unconditional Least Squares (EXACT)

In the EXACT method, unlike the CLS method, there is no simple expression for the standard errors of the predicted values. The standard errors of the predicted values will, however, be given by the least-squares prediction algorithm as a byproduct.

Standard errors of the predicted values are then transformed back to the original scale for each predicted value, if a transformation is specified.

Confidence Limits of the Predicted Values

Confidence limits of the predicted values are first computed in the transformed log scale, if a transformation is specified:

$$(LCL)_t = (FIT)_t - t_{1-\alpha/2,df}(SEP)_t \quad t = 1, 2, \ldots, N$$

$$(UCL)_t = (FIT)_t + t_{1-\alpha/2,df}(SEP)_t \quad t = 1, 2, \ldots, N$$

where $t_{1-\alpha/2,df}$ is the $(1-\alpha/2)$-th percentile of a *t* distribution with *df* degrees of freedom and α is the specified confidence level (by default α=0.05).

Confidence limits of the predicted values are then transformed back to the original scale for each predicted value, if a transformation is specified.

Forecasting

Forecasting Values

Forcasting Method: Conditional Least Squares (CLS or AUTOINIT)

$\hat{y}_t(l)$, the *l-step-ahead* forecast of y_{t+l} at the time *t*, can be represented as:

$$\hat{y}_t(l) = D(B)\hat{y}_{t+l} + \Phi(B)\mu + \Theta(B)\hat{a}_{t+l} + \sum_{i=1}^{m} c_i \Phi(B)\Delta x_{i,t+l}$$

Note that

$$\hat{y}_{t+l-i} = \begin{cases} y_{t+l-i} & \text{if } l \leq i \\ \hat{y}_t(l-i) & \text{if } l > i \end{cases}$$

$$\hat{a}_{t+l-j} = \begin{cases} y_{t+l-i} - \hat{y}_{t+l-i-1}(1) & \text{if } l \leq i \\ 0 & \text{if } l > i \end{cases}$$

Forecasting Method: Unconditional Least Squares (EXACT)

The forecasts with this option are finite memory, least-squares forecasts computed using the theoretical autocorrelation function of the series. The details of the least-squares forecasting algorithm for the ARIMA models can be found in (Brockwell et al., 1991).

Standard Errors of the Forecasting Values

Forcasting Method: Conditional Least Squares (CLS or AUTOINIT)

For the purpose of computing standard errors of the forecasting values, the model can be written in the format of weights (ignoring the model constant):

$$y_t = \frac{\vartheta_q(B)\Theta_Q(B)}{\phi_p(B)\Phi_P(B)} a_t = \psi(B)\, a_t = \sum_{i=0}^{\infty} \psi_i B^i a_{t-i}$$

where

$$\psi_0 = 1$$

Then

$$\text{se}\left[\hat{y}_t(l)\right] = \left\{1 + \psi_1^2 + \psi_2^2 + ... + \psi_{l-1}^2\right\}^{\frac{1}{2}} \hat{\sigma}_a$$

Note that, for the predicted value, $l = 1$. Hence, $(SEP)_t = \hat{\sigma}_a$ at any time *t*.

Computation of Ψ Weights. Ψ weights can be computed by expanding both sides of the following equation and solving the linear equation system established by equating the corresponding coefficients on both sides of the expansion:

$$\phi_p(B)\,\Phi_P(B)\,\Delta\psi(B) = \theta_q(B)\,\Theta_Q(B)$$

An explicit expression of Ψ weights can be found in (Box et al., 1994).

Forecasting Method: Unconditional Least Squares (EXACT)

As with the standard errors of the predicted values, the standard errors of the forecasting values are a byproduct during the least-squares forecasting computation. The details can be found in (Brockwell et al., 1991).

References

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. 1994. *Time series analysis: Forecasting and control*, 3rd ed. Englewood Cliffs, N.J.: Prentice Hall.

Brockwell, P. J., and R. A. Davis. 1991. *Time Series: Theory and Methods*, 2 ed. : Springer-Verlag.

Kohn, R., and C. Ansley. 1985. Efficient estimation and prediction in time series regression models. *Biometrika*, 72:3, 694–697.

Kohn, R., and C. Ansley. 1986. Estimation, prediction, and interpolation for ARIMA models with missing data. *Journal of the American Statistical Association*, 81, 751–761.

Makridakis, S. G., S. C. Wheelwright, and R. J. Hyndman. 1997. *Forecasting: Methods and applications*, 3rd ed. ed. New York: John Wiley and Sons.

Melard, G. 1984. A fast algorithm for the exact likelihood of autoregressive-moving average models. *Applied Statistics*, 33:1, 104–119.

Morf, M., G. S. Sidhu, and T. Kailath. 1974. Some new algorithms for recursive estimation in constant, linear, discrete-time systems. *IEEE Transactions on Automatic Control*, AC-19:4, 323–315.

Pearlman, J. G. 1980. An algorithm for the exact likelihood of a high-order autoregressive-moving average process. *Biometrika*, 67:1, 233–232.

Bibliography

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. 1994. *Time series analysis: Forecasting and control*, 3rd ed. Englewood Cliffs, N.J.: Prentice Hall.

Brockwell, P. J., and R. A. Davis. 1991. *Time Series: Theory and Methods*, 2 ed. : Springer-Verlag.

Kohn, R., and C. Ansley. 1986. Estimation, prediction, and interpolation for ARIMA models with missing data. *Journal of the American Statistical Association*, 81, 751–761.

Kohn, R., and C. Ansley. 1985. Efficient estimation and prediction in time series regression models. *Biometrika*, 72:3, 694–697.

Makridakis, S. G., S. C. Wheelwright, and R. J. Hyndman. 1997. *Forecasting: Methods and applications*, 3rd ed. ed. New York: John Wiley and Sons.

Melard, G. 1984. A fast algorithm for the exact likelihood of autoregressive-moving average models. *Applied Statistics*, 33:1, 104–119.

Morf, M., G. S. Sidhu, and T. Kailath. 1974. Some new algorithms for recursive estimation in constant, linear, discrete-time systems. *IEEE Transactions on Automatic Control*, AC-19:4, 323–315.

Pearlman, J. G. 1980. An algorithm for the exact likelihood of a high-order autoregressive-moving average process. *Biometrika*, 67:1, 233–232.