Autoregressive–moving-average model

From Wikipedia, the free encyclopedia

In the statistical analysis of time series, **autoregressive–moving-average** (**ARMA**) **models** provide a parsimonious description of a (weakly) stationary stochastic process in terms of two polynomials, one for the autoregression and the second for the moving average. The general ARMA model was described in the 1951 thesis of Peter Whittle, *Hypothesis testing in time series analysis*, and it was popularized in the 1970 book by George E. P. Box and Gwilym Jenkins.

Given a time series of data X_t , the ARMA model is a tool for understanding and, perhaps, predicting future values in this series. The model consists of two parts, an autoregressive (AR) part and a moving average (MA) part. The AR part involves regressing the variable on its own lagged (i.e., past) values. The MA part involves modeling the error term as a linear combination of error terms occurring contemporaneously and at various times in the past.

The model is usually referred to as the ARMA(p,q) model where p is the order of the autoregressive part and q is the order of the moving average part (as defined below).

ARMA models can be estimated following the Box–Jenkins approach.

Contents

Autoregressive model

The notation AR(p) refers to the autoregressive model of order p. The AR(p) model is written

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t.$$

where $\varphi_1, \ldots, \varphi_p$ are parameters, c is a constant, and the random variable ε_t is white noise.

Some constraints are necessary on the values of the parameters so that the model remains stationary. For example, processes in the AR(1) model with $|\varphi_1| \geq 1$ are not stationary.

Moving-average model

The notation MA(q) refers to the moving average model of order q:

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

where the $\theta_1, \ldots, \theta_q$ are the parameters of the model, μ is the expectation of X_t (often assumed to equal 0), and the ε_t, ε_{t-1},... are again, white noise error terms.

ARMA model

The notation ARMA(p, q) refers to the model with p autoregressive terms and q moving-average terms. This model contains the AR(p) and MA(q) models,

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}.$$

The general ARMA model was described in the 1951 thesis of Peter Whittle, who used mathematical analysis (Laurent series and Fourier analysis) and statistical inference.[1][2] ARMA models were popularized by a 1971 book by George E. P. Box and Jenkins, who expounded an iterative (Box–Jenkins) method for choosing and estimating them. This method was useful for low-order polynomials (of degree three or less).[3]

Note about the error terms

The error terms ε_t are generally assumed to be independent identically distributed random variables (i.i.d.) sampled from a normal distribution with zero mean: ε_t ~ N(0,σ^2) where σ^2 is the variance. These assumptions may be weakened but doing so will change the properties of the model. In particular, a change to the i.i.d. assumption would make a rather fundamental difference.

Specification in terms of lag operator

In some texts the models will be specified in terms of the lag operator *L*. In these terms then the AR(*p*) model is given by

$$\varepsilon_t = \left(1 - \sum_{i=1}^{p} \varphi_i L^i\right) X_t = \varphi(L) X_t$$

where φ represents the polynomial

$$\varphi(L) = 1 - \sum_{i=1}^{p} \varphi_i L^i.$$

The MA(*q*) model is given by

$$X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t = \theta(L) \varepsilon_t,$$

where θ represents the polynomial

$$\theta(L) = 1 + \sum_{i=1}^{q} \theta_i L^i.$$

Finally, the combined ARMA(*p*, *q*) model is given by

$$\left(1 - \sum_{i=1}^{p} \varphi_i L^i\right) X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t \, ,$$

or more concisely,

$$\varphi(L) X_t = \theta(L) \varepsilon_t$$

or

$$\frac{\varphi(L)}{\theta(L)} X_t = \varepsilon_t \, .$$

Alternative notation

Some authors, including Box, Jenkins & Reinsel use a different convention for the autoregression coefficients.[4] This allows all the polynomials involving the lag operator to appear in a similar form throughout. Thus the ARMA model would be written as

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right) X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t \, .$$

Moreover, if we set $\phi_0 = \theta_0 = 1$, then we get an even more elegant formulation: $\sum_{i=0}^{p} \phi_i L^i \; X_t = \sum_{i=0}^{q} \theta_i L^i \; \varepsilon_t \, .$

Fitting models

Choosing p and q

Finding appropriate values of *p* and *q* in the ARMA(*p*,*q*) model can be facilitated by plotting the partial autocorrelation functions for an estimate of *p*, and likewise using the autocorrelation functions for an estimate of *q*. Further information can be gleaned by considering the same functions for the residuals of a model fitted with an initial selection of *p* and *q*.

Brockwell & Davis recommend using AICc for finding *p* and *q*.[5]

Estimating coefficients

ARMA models in general cannot be, after choosing *p* and *q*, fitted by least squares regression to find the values of the parameters which minimize the error term. It is generally considered good practice to find the smallest values of *p* and *q* which provide an acceptable fit to the data. For a pure AR model the Yule-Walker equations may be used to provide a fit.

Implementations in statistics packages

- In R, the *arima* function (in standard package *stats*) is documented in ARIMA Modelling of Time Series (http://search.r-project.org/R/library/stats/html/arima.html). Extension packages contain related and extended functionality, e.g., the *tseries* package includes an *arma* function, documented in "Fit ARMA Models to Time Series" (http://finzi.psych.upenn.edu/R/library/tseries/html/arma.html); the *fracdiff* package (https://cran.r-project.org/web/packages/fracdiff) contains *fracdiff()* for fractionally integrated ARMA processes, etc. The CRAN task view on Time Series (https://cran.r-project.org/web/views/TimeSeries.html) contains links to most of these.
- Mathematica has a complete library of time series functions including ARMA.[6]
- MATLAB includes functions such as *arma* (http://www.mathworks.com/help/econ/arma-model.html) and *ar* (http://www.mathworks.com/help/ident/ref/ar.html) to estimate AR, ARX (autoregressive exogenous), and ARMAX models. See System Identification Toolbox (http://www.mathworks.com/help/ident/ug/estimating-ar-and-arma-models.html) and Econometrics Toolbox (http://www.mathworks.com/help/econ/arima.estimate.html) for more information.
- Statsmodels Python module includes many models and functions for time series analysis, including ARMA. Formerly part of Scikit-learn it is now stand-alone and integrates well with Pandas. See here for more details (http://statsmodels.sourceforge.net/).
- PyFlux has a Python-based implementation of ARIMAX models, including Bayesian ARIMAX models. See here for details (http://www.pyflux.com/).
- IMSL Numerical Libraries are libraries of numerical analysis functionality including ARMA and ARIMA procedures implemented in standard programming languages like C, Java, C# .NET, and Fortran.
- gretl can also estimate ARMA model, see here where it's mentioned (http://constantdream.wordpress.com/2008/03/16/gnu-regression-econometrics-and-time-series-library-gretl/).
- GNU Octave can estimate AR models using functions from the extra package octave-forge (http://octave.sourceforge.net/).
- Stata includes the function *arima* which can estimate ARMA and ARIMA models. See here for more details (https://www.stata.com/help.cgi?arima.).
- SuanShu is a Java library of numerical methods, including comprehensive statistics packages, in which univariate/multivariate ARMA, ARIMA, ARMAX, etc. models are implemented in an object-oriented approach. These implementations are documented in "SuanShu, a Java numerical and statistical library" (http://www.numericalmethod.com/javadoc/suanshu/).
- SAS has an econometric package, ETS, that estimates ARIMA models. See here for more details (http://support.sas.com/rnd/app/ets/proc/ets_arima.html).

Applications

ARMA is appropriate when a system is a function of a series of unobserved shocks (the MA or moving average part) as well as its own behavior. For example, stock prices may be shocked by fundamental information as well as exhibiting technical trending and mean-reversion effects due to market participants.

Generalizations

The dependence of X_t on past values and the error terms ε_t is assumed to be linear unless specified otherwise. If the dependence is nonlinear, the model is specifically called a *nonlinear moving average* (NMA), *nonlinear autoregressive* (NAR), or *nonlinear autoregressive–moving-average* (NARMA) model.

Autoregressive–moving-average models can be generalized in other ways. See also autoregressive conditional heteroskedasticity (ARCH) models and autoregressive integrated moving average (ARIMA) models. If multiple time series are to be fitted then a vector ARIMA (or VARIMA) model may be fitted. If the time-series in question exhibits long memory then fractional ARIMA (FARIMA, sometimes called ARFIMA) modelling may be appropriate: see Autoregressive fractionally integrated moving average. If the data is thought to contain seasonal effects, it may be modeled by a SARIMA (seasonal ARIMA) or a periodic ARMA model.

Another generalization is the *multiscale autoregressive* (MAR) model. A MAR model is indexed by the nodes of a tree, whereas a standard (discrete time) autoregressive model is indexed by integers.

Note that the ARMA model is a **univariate** model. Extensions for the multivariate case are the Vector Autoregression (VAR) and Vector Autoregression Moving-Average (VARMA).

Autoregressive–moving-average model with exogenous inputs model (ARMAX model)

The notation ARMAX(*p*, *q*, *b*) refers to the model with *p* autoregressive terms, *q* moving average terms and *b* exogenous inputs terms. This model contains the AR(*p*) and MA(*q*) models and a linear combination of the last *b* terms of a known and external time series d_t. It is given by:

$$X_t = \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \sum_{i=1}^{b} \eta_i d_{t-i}.$$

where $\eta_1, \ldots, \eta_b$ are the ***parameters*** of the exogenous input d_t.

Some nonlinear variants of models with exogenous variables have been defined: see for example Nonlinear autoregressive exogenous model.

Statistical packages implement the ARMAX model through the use of "exogenous" or "independent" variables. Care must be taken when interpreting the output of those packages, because the estimated parameters usually (for example, in R[7] and gretl) refer to the regression:

$$X_t - m_t = \varepsilon_t + \sum_{i=1}^{p} \varphi_i (X_{t-i} - m_{t-i}) + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}.$$

where m_t incorporates all exogenous (or independent) variables:

$$m_t = c + \sum_{i=0}^{b} \eta_i d_{t-i}.$$

See also

- Exponential smoothing

- Linear predictive coding
- Predictive analytics
- Arima, autoregressive integrated moving average

References

1. Hannan, Edward James (1970). *Multiple time series*. Wiley series in probability and mathematical statistics. New York: John Wiley and Sons.
2. Whittle, P. (1951). *Hypothesis Testing in Time Series Analysis*. Almquist and Wicksell. Whittle, P. (1963). *Prediction and Regulation*. English Universities Press. ISBN 0-8166-1147-5.

   Republished as: Whittle, P. (1983). *Prediction and Regulation by Linear Least-Square Methods*. University of Minnesota Press. ISBN 0-8166-1148-3.

3. Hannan & Deistler (1988, p. 227): Hannan, E. J.; Deistler, Manfred (1988). *Statistical theory of linear systems*. Wiley series in probability and mathematical statistics. New York: John Wiley and Sons.
4. Box, George; Jenkins, Gwilym M.; Reinsel, Gregory C. (1994). *Time Series Analysis: Forecasting and Control* (Third ed.). Prentice-Hall. ISBN 0130607746.
5. Brockwell, P. J.; Davis, R. A. (2009). *Time Series: Theory and Methods* (2nd ed.). New York: Springer. p. 273. ISBN 9781441903198.
6. Time series features in Mathematica (http://www.wolfram.com/products/applications/timeseries/features.html) Archived (https://web.archive.org/web/20111124032002/http://www.wolfram.com/products/applications/timeseries/features.html) November 24, 2011, at the Wayback Machine.
7. ARIMA Modelling of Time Series (http://search.r-project.org/R/library/stats/html/arima.html), R documentation

Further reading

- Mills, Terence C. (1990). *Time Series Techniques for Economists*. New York: Cambridge University Press. ISBN 0521343399.
- Percival, Donald B.; Walden, Andrew T. (1993). *Spectral Analysis for Physical Applications*. New York: Cambridge University Press. ISBN 052135532X.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Autoregressive–moving-average_model&oldid=783947099"

Categories: Autocorrelation
