

Lecture 16: ARIMA / GARCH Models

Steven Skiena

Department of Computer Science
State University of New York
Stony Brook, NY 11794–4400

<http://www.cs.sunysb.edu/~skiena>

Moving Average Models

A time series is said to be a *moving average* process of order q if it is a weighted linear sum of the last q random shocks / errors.

In general, an $MA(q)$ model has form

$$r_t = c_0 + \sum_{i=0}^{i=q-1} \theta_i a_{t-i}$$

The history of the model dictates how long the effects of the random shocks last. For an $MA(2)$ model,

$$r_t = c_0 + \theta_0 a_t + \theta_1 a_{t-1}$$

The order of such a model can be determined by analysis of the autocorrelation function, which is zero for all lags greater than q .

Fitting Moving Average Models

The parameters of this model cannot be set using least squares, because (1) the random shocks themselves are not given, and (2) solving for the error terms a_t from r_t depends upon the parameters being known.

A numerical fitting method tries all coefficient sets on a finite grid/space of parameter values, and selects the one which minimizes the sum of the random shocks, i.e. $\sum a_t$.

Transformations

Linear time series methods do not work well on non-linear functions.

Defining the transform $y_t = \ln x_t$ converts an exponential function to a linear one.

Any *trend* f_n observed in the time series (say, through curve fitting, regression analysis or first principles) can be subtracted out to leave what remains to be modeled:

$$y_t = x_t - f_n$$

Differencing Models

Differencing transforms a time series X into another series Y where $y_t = x_t - x_{t-1}$, trying to find a better fitting model.

Differencing does not require estimating a parameter, although it costs us one series point per difference.

Differencing is a better way to remove locally varying trends to make it stationary than explicitly subtracting a fitted trend.

The first difference accounts for a trend that impacts the change of the mean of the time series, the second for a change in the slope.

$$z_t = y_t - y_{t-1} = x_t - 2x_{t-1} + x_{t-2}$$

Autocorrelation Analysis of GDP

```
**** Demonstration of Analysis of U.S. Quarterly Real GDP: 1967-2002.
**** "<====" indicates my explanation of the command.
--
input xl,x. file 'q-rgdpf6702.dat' <==== Load data into SCA.

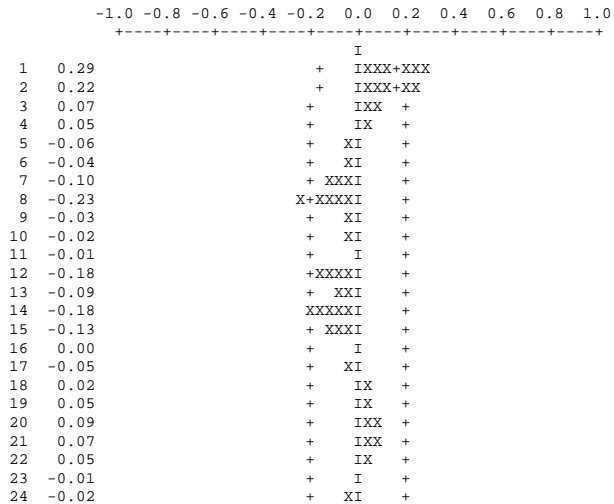
X1      , A 144 BY 1 VARIABLE, IS STORED IN THE WORKSPACE
X       , A 144 BY 1 VARIABLE, IS STORED IN THE WORKSPACE
--
y=ln(x) <==== Take natural log transformation
--
diff old y. new dy. comp. <==== Take first difference of the log series.
--
                1
DIFFERENCE ORDERS ARE (1-B )
SERIES Y IS DIFFERENCED, THE RESULT IS STORED IN VARIABLE DY
SERIES DY HAS 143 ENTRIES
--
iden dy. <==== Compute ACF and PACF of the differenced series.

NAME OF THE SERIES . . . . . DY
TIME PERIOD ANALYZED . . . . . 1 TO 143
MEAN OF THE (DIFFERENCED) SERIES . . . . . 0.0071
STANDARD DEVIATION OF THE SERIES . . . . . 0.0079
T-VALUE OF MEAN (AGAINST ZERO) . . . . . 10.7080

AUTOCORRELATIONS

1- 12      .29 .22 .07 .05 -.06 -.04 -.10 -.23 -.03 -.02 -.01 -.18
ST.E.      .08 .09 .09 .09 .09 .10 .10 .10 .10 .10 .10 .10
Q          12.4 19.8 20.4 20.9 21.4 21.7 23.2 31.1 31.3 31.3 31.4 36.5

13- 24     -.09 -.18 -.13 -.00 -.05 .02 .05 .09 .07 .05 -.01 -.02
ST.E.      .10 .10 .10 .11 .11 .11 .11 .11 .11 .11 .11 .11
Q          37.9 43.1 45.7 45.7 46.1 46.2 46.5 47.9 48.7 49.2 49.2 49.3
```



Thus autocorrelation analysis enables us to identify periodic cycles in the time series.

Autoregressive Integrated Moving Average (ARIMA) Models

An $ARIMA(p, d, q)$ model has p autoregressive (history) terms, q moving average terms, and d differencing operations. Integrated refers to the differencing process.

This describes a rich set of models for which numerical methods exist to fit models for any given p , d , and q .

This suggests an automated approach to try all models in some subset of parameter space and select the best one.

Error in Time Series Models

Evaluating a forecasting model requires quantifying the prediction error $\Delta_t = (F_t - S_t)$.

The *mean squared error* criteria $\sum_{t=s}^{n+s} (\Delta_t)^2 / n$ is commonly used, but makes sense only if twice the Δ_t is four times as undesirable.

The *mean absolute prediction error* $\sum_{t=s}^{n+s} (\Delta_t / S_t) / n$ is akin to percentage error, and appropriate for comparing results on different time series.

In Sample Testing vs. Best Fit

A model may perfectly fit the data but be useless for prediction (e.g. a degree n -polynomial.)

In sample testing separates some evaluation data from the training data, so the results measure ability to forecast instead of overfitting.

Evaluating Models

The quality of a model is a function of both the *fit*, the sum of squared residuals (S), and the number of parameters p .

Akaike's information criteria states

$$AIC \approx n \ln(S/n) + 2p$$

The Bayesian information criteria punishes extra parameters more:

$$BIC \approx n \ln(S/n) + p + p \ln n$$

These measures provide a way to compare models and select the most meaningful one.

GARCH Models

GARCH (generalized autoregressive conditional heteroscedastic) models are often used to model volatility. *Heteroscedastic* means a set of statistical distributions having different variances.

Let V_l be the long run average variance rate.

The GARCH(1,1) model is defined by

$$\sigma_t^2 = \gamma V_l + \alpha a_{t-1}^2 + \beta \sigma_{t-1}^2$$

where $0 \leq \alpha, \beta, \gamma \leq 1$ and $(\alpha + \beta + \gamma) = 1$

This reduces to EWMA smoothing when $\gamma = 0$.

Mean Reversion

These conditions imply that the model is *mean reverting*, which implies that it returns to an average value after reaching extremes, i.e.

$$E[\sigma_{n+t}^2] = V_l + (\alpha + \beta)^t(\sigma_n^2 - V_l)$$

The speed with which it reverts depends upon γ .

Maximum Likelihood Methods

An alternate model fitting criteria than minimizing residual error is based on probability.

In **maximum likelihood methods**, we seek the parameters which maximize the probability that our observed data occurred.

As an example, suppose we want to estimate the probability p that a given stock went down on a given day.

Suppose we sample ten stocks and only one went down. What is p ?

The Maximum Likelihood Approach

If p is the probability of an individual decline, the probability of seeing one up and nine down is $p(1 - p)^9$.

To maximize this, take the derivative with respect to p and set to 0:

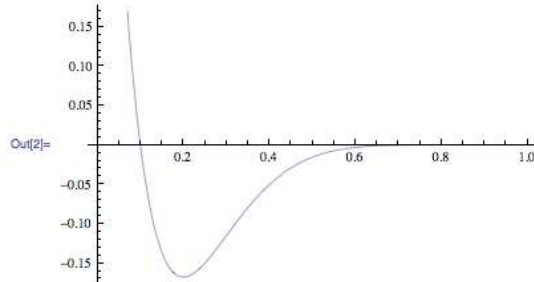
$$(1 - p)^9 - 9(1 - p)^8 p = 0$$

which implies that indeed $p = 1/10$.

```
In[1]:= D[p (1 - p)^9, p]
```

```
Out[1]:= (1 - p)^9 - 9 (1 - p)^8 p
```

```
In[2]:= Plot[(1 - p)^9 - 9 (1 - p)^8 p, {p, 0, 1}]
```



Maximum likelihood methods provide well-defined and defensible criteria for model selection.

Estimating GARCH(1,1) Parameters

Assume that daily returns/observations are normally distributed around a mean.

Then each day's observation u_i maps to a probability which is a function of σ_n^2 , namely

$$P(u_i) = \frac{1}{\sqrt{2\pi v_i}} \exp\left(-\frac{u_i^2}{2v_i}\right)$$

where $v_i = \sigma_i^2$.

The probability of the complete set of observations is

$$\prod_{i=1}^m P(u_i)$$

Taking Logarithms

Maximizing this probability is the same as maximizing its logarithm, so we seek to maximize

$$\sum_{i=1}^m \left(-\ln(v_i) - \frac{-u_i^2}{v_i} \right)$$

Given this criteria, the best values for α , β , and γ can be found numerically.

Asymmetric Volatility Models

Volatility modeling can get very sophisticated.

The volatility of a stock's price tends to be inversely related to its price.

If so, a negative move/observation u_{n-1} should have a bigger effect on σ_n than a positive move.

Accordingly, asymmetric GARCH models have been developed where σ_n depends upon the sign of u_{n-1} .