MCMC analysis of classical time series algorithms.

Isambi Sailon
mbalawata@yahoo.com

Lappeenranta University of Technology

Lappeenranta, 19.03.2009

Outline

The aim of this work is to combine the modern MCMC approach with classical time series algorithms, to analyse predictive distributions of estimated parameters.

There are three main phases to be analysed: phase one will be about Time Series analysis, phase two will be MCMC series and phase three will be the MCMC analysis for classical time series.

Box-Jenkins models

Mathematical models used typically for accurate short-term forecasts of 'well-behaved' data (that shows predictable repetitive cycles and patterns) and find the best fit of a time series to past values of this time series, in order to make forecasts.

A time series is a sequence of observations based on a regular timely basis, *e.g.* hourly, daily, monthly, annually, *etc.* The classical time series analysis covers fitting autoregressive (AR) and moving average (MA) models.

AR(r)

- $x_t = C + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_n x_{t-r} + u_t$

Where C is a constant and $\phi_1 \cdots \phi_n$ are autoregressive parameters.

MA(m)

- $x_t = C + \psi_1 u_{t-1} + \psi_2 u_{t-2} + \ldots + \psi_n u_{t-m} + u_t$

Where C is a constant and $\psi_1 \cdots \psi_n$ are moving average parameters.

ARMA(r, m)

- $x_t = C + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_n x_{t-r} + \psi_1 u_{t-1} + \psi_2 u_{t-2} + \ldots + \psi_n u_{t-m} + u_t$

Where $u_t \sim N(0, \sigma^2)$ – white noise and x_t is the time series

Autocorrelation (AC)

A certain type of correlation concept that portrays the dependence of two consecutive observations of time series. Thus Autocorrelations are statistical measures that indicate how a time series is related to itself over time.

For example: the autocorrelation at lag 1 is the correlation between the original series and the same series moved forward one period.

- $r_k = \frac{E[(x_t - \mu)(x_{t+k} - \mu)]}{\sigma_x^2}$

where k is the specified lag number, x_t are series observations, μ is the series mean value and σ_x^2 is the variance.

The ACF will first test whether adjacent observations are autocorrelated; that is, whether there is correlation between observations 1 and 2, 2 and 3, 3 and 4, etc. This is known as lag one autocorrelation, since one of the pair of tested observations lags the other by one period or sample.

Similarly, it will test at other lags. For instance, the autocorrelation at lag four tests whether observations 1 and 5, 2 and 6, ...,19 and 23, etc. are correlated.

Estimates at longer lags have been shown to be statistically unreliable (Box and Jenkins, 1970). In some cases, the effect of Autocorrelation at smaller lags will influence the estimate of autocorrelation at longer lags. For instance, a strong lag one autocorrelation would cause observation 5 to influence observation 6, and observation 6 to influences 7.

This results in an apparent correlation between observations 5 and 7, even though no direct correlation exists.

Partial autocorrelation (PAC)

The Partial Autocorrelation Function (PACF) removes the effect of shorter lag autocorrelation from the correlation estimate at longer lags. PAC is Similar to AC, except that when calculating it, the ACs with all the elements within the lag are partialled out.

- $r_{kk} = \frac{r_k - r_{k-1}^2}{1 - r_{k-1}^2}$

PCA measures the correlation between an observation k periods ago and current observation, after controlling for the observations at intermediate lags (all lags $< k$), that is, the correlation between y_t and y_{t-k}, after removing the effects of $y_{t-k+1}, y_{t-k+2}, \cdots, y_{t-1}$

Stages in building a Box-Jenkins time series model

1. Model Identification

This involves determining the order of the model required to capture the dynamic features of the data. Graphical procedures are used to determine the most appropriate specification (Making sure that the variables are stationary, identifying seasonality, and the use of plots of the autocorrelation and partial autocorrelation functions)

2. Model Estimation

This involves estimation of the parameters of the model specified in Model Identification. The most common methods used are maximum likelihood estimation and non-linear least-squares estimation, depending on the model.

3. Model Validation

This involves model checking, that is, determining whether the model specified and estmated is adequate. Thus: Model validation deals with testing whether the estimated model conforms to the specifications of a stationary univariate process. It can be done by Overfitting and Residual diagnosis

MODEL IDENTIFICATION

In Box-Jenkins model one has to determine if the series is stationary and if there is any significant seasonality that needs to be modelled.

A stationary process has the property that the mean, variance and autocorrelation structure do not change over time. It can be assessed by run sequence plot (graph that displays observed data in a time sequence (A graphical data analysis technique for preliminary scanning of the data).

Also, apart from Run sequence plot, the condition for testing stationarity of AR model is that the roots of the characteristics equation all lie outside the unit circle.
For example $x_t = 3x_{t-1} - 2.75x_{t-2} + 0.75x_{t-3} + u_t$ is not stationary because its roots $1, \frac{2}{3}$ and 2 only one lies outside the unit.

$x_t = 3x_{t-1} - 2.75x_{t-2} + 0.75x_{t-3} + u_t$ is expressed using lag operator notation thus it will be as follows:

$x_t = 3Lx_t - 2.75L^2x_t + 0.75L^3x_3 + u_t$

$(1 - 3L + 2.75L^2 - 0.75L^3)x_t = u_t$

The characeristic equation is $1 - 3z + 2.75z^2 - 0.75z^3 = 0$. Solving for z; the solution set is $1, \frac{2}{3}, 2$

For seasonality, it means periodic fluctuations, that is, a certain basic pattern tends to be repeated at regular seasonal intervals. This can be assessed by a run sequence plot, a seasonal subseries plot, multiple box plots, or autocorrelation plot.

Once stationarity and seasonality have been addressed, one needs to identify the order (r and m) of the autoregressive and moving average terms. The primary tools for doing this are the autocorrelation plot and the partial autocorrelation plot.

The sample autocorrelation plot and the sample partial autocorrelation plot are compared to the theoretical behaviour of these plots when the order is known

The ARMA model identification is based on autocorrelation and partial autocorrelation function values. Naturally the model, whose values are closest to calculated ones, is chosen. Understanding the concept of autocorrelation can be tested by trying to conclude the mentioned theoretical values; here they are presented in a table:

Model	Theoretical r_k	Theoretical r_{kk}
AR(0)	All zero	All zero
AR(1)	Vanish toward zero	Zero after 1st lag
AR(2)	Vanish toward zero	Zero after 2nd lag
MA(1)	Zero after 1st lag	Vanish toward zero
MA(2)	Zero after 2nd lag	Vanish toward zero
ARMA(1,1)	Vanish toward zero	Vanish toward zero

MODEL ESTIMATION

The classical Box-Jenkins model estimation uses a recursive algorithm. The objective is to minimize the sum of squares of errors.

If we assume for simplicity the ARMA(1,1) model

$$x_t = \phi x_{t-1} + \psi u_{t-1} + u_t$$

then the algorithm considers all possible values of ϕ and ψ and for them minimizes the sum of squares $\sum_{i=1}^{t} u_i^2$.

In particular, assume we have $x_0, x_1, x_2, \ldots, x_t$ as the observed stationary series. Initially $u_0 = 0$. Then we have:

$$\begin{aligned}
x_1 &= \phi x_0 + \psi \cdot 0 + u_1 && \Rightarrow && u_1 = x_1 - \phi x_0 \\
x_2 &= \phi x_1 + \psi \cdot u_1 + u_2 && \Rightarrow && u_2 = x_2 - \phi x_1 - \psi u_1 \\
x_3 &= \phi x_2 + \psi \cdot u_2 + u_3 && \Rightarrow && u_3 = x_3 - \phi x_2 - \psi u_2 \\
&\vdots \\
x_n &= \phi x_{n-1} + \psi \cdot u_{n-1} + u_n && \Rightarrow && u_n = x_n - \phi x_{n-1} - \psi u_{n-1}
\end{aligned}$$

The parameters ϕ and ψ are varied between (-1,1) (since the series is required to be stationary), and the sums $\sum_{i=1}^{t} u_i^2$ are calculated each time. Then the proper estimates ϕ and ψ, and the u_t noise values are those that give the minimal value of the u_t sum of squares.

a) Generate simple series of at least 1000 observations with coefficients such that the series are stationary.

- ARMA(1,0) $\rightarrow x_t = \theta x_{t-1} + u_t$
- ARMA(1,1) $\rightarrow x_t = \theta x_{t-1} + \psi u_{t-1} + u_t$
- ARMA(2,0) $\rightarrow x_t = \theta_1 x_{t-1} + \theta_2 x_{t-2} + u_t$
- ARMA(2,1) $\rightarrow x_t = \theta_1 x_{t-1} + \theta_2 x_{t-2} + \psi_1 u_{t-1} + u_t$
- ARMA(2,2) $\rightarrow x_t = \theta_1 x_{t-1} + \theta_2 x_{t-2} + \psi_1 u_{t-1} + \psi_2 u_{t-2} + u_t$

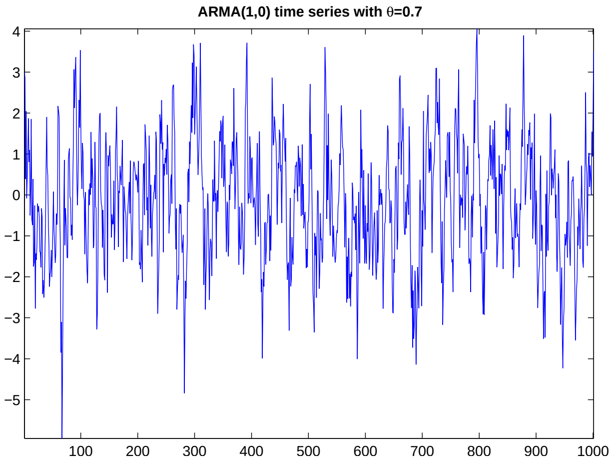

Figure: Example ARMA(1,0) time series with $x_0 = 3$, $\theta = 0.7$ and white noise $u_t \sim N(0, 1)$.

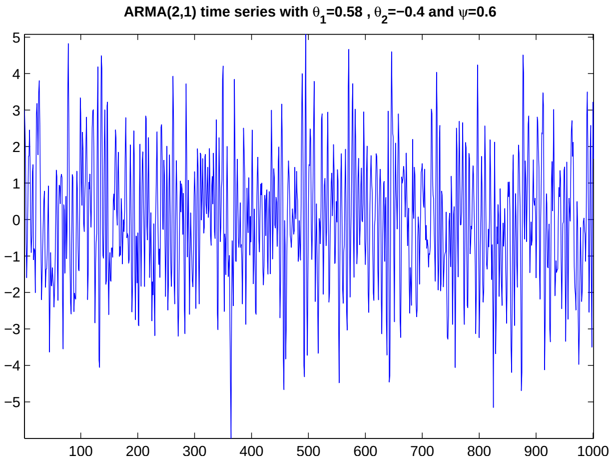

Figure: Example ARMA(2,1) time series with $x_0 = 3$, $x_1 = 2.5$, $\theta_1 = 0.58$, $\theta_2 = -0.4$, $\psi = 0.6$ and white noise $u_t \sim N(0, 1)$.

b) Perform the Box-Jenkins recursive calculation to estimate model parameters and compare results with the originally defined model parameters.

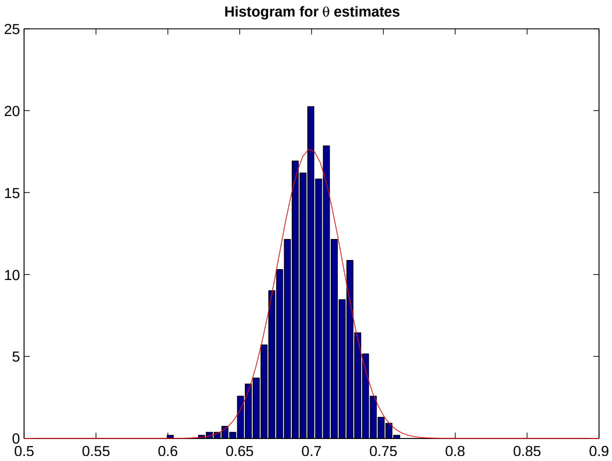

Figure: Normalized histogram of θ estimates obtained from Box-Jenkins recursive estimation (original $\theta = 0.7$).

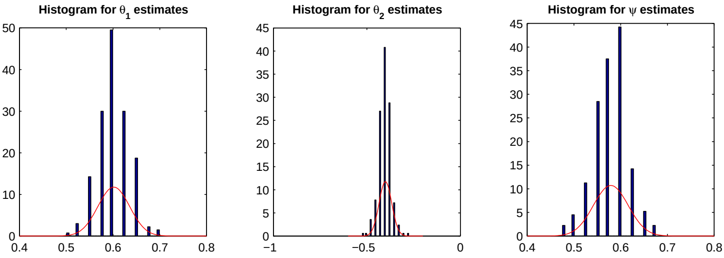

Figure: Normalized histogram of θ_1, θ_2, ψ estimates obtained from Box-Jenkins recursive estimation (original $\theta_1 = 0.58$, $\theta_2 = -0.4$, $\psi = 0.6$).

c) Generate new values for the noise in all models and estimate parameters using regression (backslash operation in Matlab).

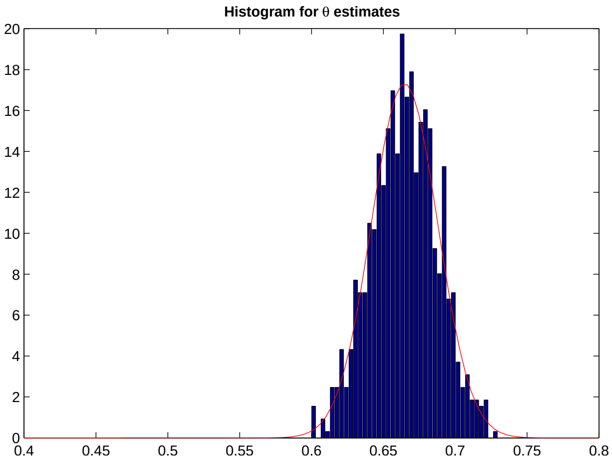

Figure: Normalized histogram of θ estimates obtained from regression estimation (original $\theta = 0.7$).

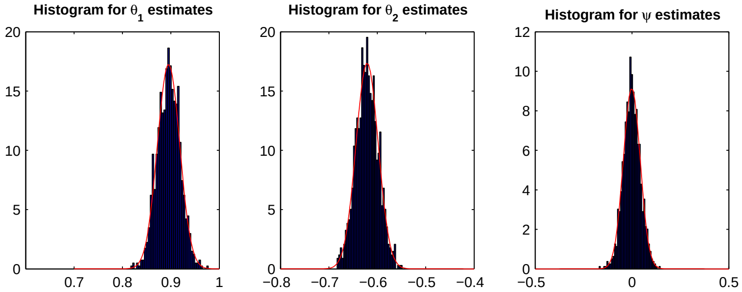

Figure: Normalized histogram of θ_1, θ_2, ψ estimates obtained from regression estimation (original $\theta_1 = 0.58$, $\theta_2 = -0.4$, $\psi = 0.6$).

d) Use Matlab GARCH toolbox (function `garchfit.m` in particular) and compare its estimates with the original model values.

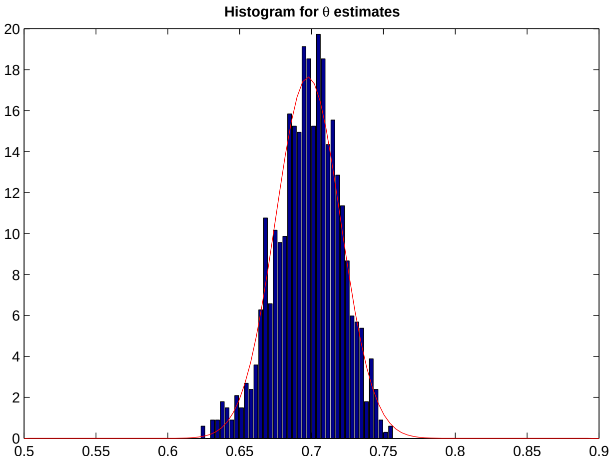

Figure: Normalized histogram of θ estimates obtained from Matlab garchfit.m estimation (original $\theta = 0.7$).

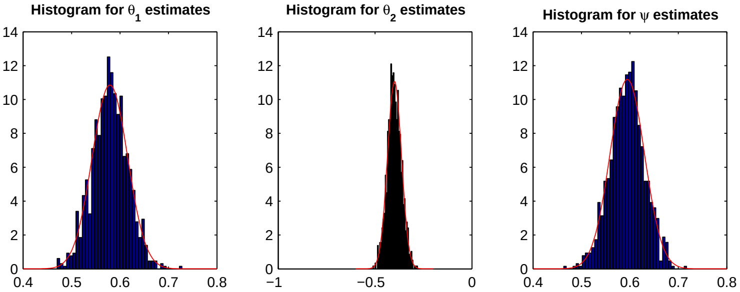

Figure: Normalized histogram of θ_1, θ_2, ψ estimates obtained from Matlab garchfit.m estimation (original $\theta_1 = 0.58$, $\theta_2 = -0.4$, $\psi = 0.6$).

Recursive Estimation

The recursive estimation gives better results. The multiple runs give results that are close to the original parameters values.

New Noise and Regression Estimation

With new noise values used for regression estimation, the results are significantly different from the original ones. For instance the results show zero value for ψ while the original ψ was 0.6.

GARCHFIT.M

Using the built in MATLAB function garchfit.m, the results are close to the original parameter values.

Future work

After this classical time series, the next step is MCMC algorithms which will be followed by combining the MCMC analysis with the classical time series. Here the MCMC will be used to study the time series algorithms for parameter prediction (that is the use of formal MCMC to analyse the classical Box-Jenkins parameter estimation).

ASANTENI