Estimation of AR, MA & ARMA Models

Aditya Pathak
Anirudh Jayaraman
Prashant Shukla
Rohit Arora

IGIDR

November 17, 2016

Model Building of ARIMA(p,d,q)

- Transformation
- Identification
- Parameter Estimation
- Diagnostic Checking

ACF and PACF

Basic idea

- ACF: Mathematical representation of the degree of similarity between a given time series and a lagged version of itself over successive time intervals
- PACF: Partial correlation of a time series with its own lagged values, controlling for the values of the time series at all shorter lags

Identification

- To determine the value of p and q, we use the graphical properties of the ACF and the PACF

	MA(q)	AR(p)	ARMA(p,q)
ACF	Cuts after q	Tails off	Tails off
PACF	Tails off	Cuts after p	Tails off

Parameter Estimation

- Method of Moments Estimator
- Conditional Estimator
  - NLLS
  - MLE
- Unconditional Estimator
  - NLLS
  - MLE

Method of Moments

Principle

- Work out the theoretical ACFs i.e. ρ_k in terms of the parameters of the model
- Estimate r_k and set $\rho_k = r_k$.

$$r_k = \frac{\frac{1}{T}\sum_{t=1}^{t=T-k}(X_t - \bar{X})(X_{t+k} - \bar{X})}{\frac{1}{T}\sum_{t=1}^{t=T-k}(X_t - \bar{X})^2} \tag{1}$$

- Solve for the parameters

AR(p) Process

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + ... + a_k X_{t-k} + ... + a_p X_{t-p} \tag{2}$$

$$a(B)\rho_k = 0 \tag{3}$$

where

$$a(B) = (1 - a_1 B - a_2 B^2 - .... - a_p B^p) \tag{4}$$

So,

$$\rho_1 = a_1\rho_0 + a_2\rho_1 + ... + a_p\rho_{p-1} \tag{5}$$

Similarly,

$$\rho_2 = a_1\rho_1 + a_2\rho_0 + ... + a_p\rho_{p-2} \tag{6}$$

.

.

$$\rho_p = a_1\rho_{p-1} + a_2\rho_{p-2} + ... + a_p\rho_0 \tag{7}$$

...contd

In Matrix form,

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ . \\ . \\ . \\ \rho_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & . & . & . & \rho_{p-1} \\ \rho_1 & 1 & & & & & \\ \rho_2 & \rho_1 & 1 & & & & \\ . & & & . & & & \\ . & & & & . & & \\ . & & & & & . & \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & . & . & . & 1 \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ . \\ . \\ . \\ a_p \end{bmatrix}$$

$$\rho = R\, a$$
$$\mathrm{R}^{-1}\rho = a$$
$$\Rightarrow \hat{a} = \hat{R}^{-1} r$$

MA(q) Process

$$X_t = \epsilon_t + b_1\epsilon_{t-1} + ... + b_q\epsilon_{t-q} \tag{8}$$

$$\rho_k = \frac{b_k + b_1 b_{k+1} + ...b_{q-k}b_q}{1 + b_1^2 + b_2^2 + ... + b_q^2} \qquad k \leq q \tag{9}$$

- We have q equations and q parameters
- Solve by replacing ρ_k by r_k
- Multiple solutions possible but there will be only one which is invertible

ARMA(p,q) Process

$$X_t = a_1 X_{t-1} + ... + a_p X_{t-p} + \epsilon_t + b_1 \epsilon_{t-1} + ... + b_q \epsilon_{t-q} \quad (10)$$

- We will use a recursive procedure
- We know ACF of ARMA(p,q) behaves like ACF of AR(p) after q lags

...contd

$\Rightarrow$ a(B)$\rho_{k'}$ where $k^{'}$ = q+1,......,q+p

$\Rightarrow$ Can get $a_1, ..., a_p$

- Then construct $a(B)X_t = X_t^{'}$
- $X_t^{'} \simeq \epsilon_t + b_1\epsilon_{t-1} + ... + b_q\epsilon_{t-q}$
- Use MA(q) method to solve for $b_1, ..., b_q$

Conditional Estimators

NLLS

- AR(p) process

$$(X_t - \mu) = a_1(X_{t-1} - \mu) + a_2(X_{t-2} - \mu) + \dots\dots\dots\dots + a_p(X_{p-1} - \mu) \quad (11)$$

$$\sum_{p+1}^{T} {\epsilon_t}^2 = \sum_{p+1}^{T} [(X_t - \mu) - a_1(X_{t-1} - \mu)\dots\dots\dots\dots - a_p(X_{p-1} - \mu)]^2 \quad (12)$$

- Sum of squares $= \sum {\epsilon_t}^2(a_1, a_2, ..., a_p, \mu)$
- Differentiate wrt $a_1, a_2, \dots\dots\dots\dots\dots\dots\dots, a_p, \mu$
- p+1 equations in p+1 unknowns

AR(1) Process

$$X_t - \mu = a_1(X_{t-1} - \mu) + \epsilon_t \tag{13}$$

- Minimize $\sum_2^T {\epsilon_t}^2(a, a_1) = \sum_2^T (X_t - \mu - a_1 X_{t-1})^2$
- Differentiate wrt a and a_1 to get the values

MA(q) Process

$$X_t = \epsilon_t + b_1\epsilon_{t-1} + \dots\dots\dots\dots\dots\dots\dots + b_q\epsilon_{t-q} \tag{14}$$

$$\epsilon_1 = X_1 - b_1\epsilon_0 - b_2\epsilon_{-1}\dots\dots\dots\dots\dots\dots\dots - b_q\epsilon_{1-q} \tag{15}$$

$$\epsilon_2 = X_2 - b_1\epsilon_1 - b_2\epsilon_0\dots\dots\dots\dots\dots\dots\dots - b_q\epsilon_{2-q} \tag{16}$$

- Set all pre-sample $\epsilon_i\acute{s} = 0$ and minimize $\sum_1^T \epsilon_t^2$
- Highly non-linear

MA(1) Process

$$X_t = \epsilon_t + b_1\epsilon_{t-1} \tag{17}$$

$$\epsilon_t = X_t - b_1\epsilon_{t-1} \tag{18}$$

$$\epsilon_1 = X_1 \tag{19}$$

$$\epsilon_2 = X_2 - b_1\epsilon_1 \tag{20}$$

- Minimize $\sum_1^T \epsilon_t^2 = \sum_1^T (X_t - b_1\epsilon_{t-1})^2$
- Highly non-linear

ARMA(p,q) Process

$$X_1 - \mu = a_1(X_{t-1} - \mu) + ...a_p(X_{t-p} - \mu) + \epsilon_t + b_1\epsilon_{t-1} + ... + b_q\epsilon_{t-q} \quad (21)$$

$$\epsilon_t = X_1 - \mu - a_1(X_{t-1} - \mu) - .....a_p(X_{t-p} - \mu) - b_1\epsilon_{t-1}........ - b_q\epsilon_{t-q} \quad (22)$$

$\theta_0 = [a_1, a_2, ..., a_p, b_1, b_2, ..., b_q, \mu] \rightarrow$ true but unknown.

- $\epsilon_t(\theta_0) \Rightarrow$ true errors
- But θ_0 is unknown and hence $\epsilon_t(\theta_0)$ is unknown
- Therefore we start with an initial estimate $\Rightarrow \theta^0$ and calculate $\epsilon_t(\theta^0)$

Linearizing the model

$$\epsilon_t(\theta_\circ) = \epsilon_t(\theta^\circ) + \left.\frac{\delta\epsilon_t(\theta)}{\delta\theta}\right|_{\theta^\circ}(\theta_\circ - \theta^\circ)$$

- How to get $\left.\frac{\delta\epsilon_t(\theta)}{\delta\theta}\right|_{\theta^\circ}$

$$\left.\frac{\delta\epsilon_t(\theta)}{\delta\theta_i}\right|_{\theta^\circ} = \frac{1}{\delta}[\epsilon_t(\theta_1{}^\circ, \theta_2{}^\circ, .., \theta_i{}^\circ + \delta, ..., \theta^\circ{}_{p+q+1}) - \epsilon_t(\theta_1{}^\circ, \theta_2{}^\circ, .., \theta_i{}^\circ, ..., \theta^\circ{}_{p+q+1}]$$

$$\epsilon_t(\theta^\circ) = \text{-} \left.\frac{\delta\epsilon_t(\theta)}{\delta\theta}\right|_{\theta^\circ}(\theta_\circ - \theta^\circ) + \epsilon_t(\theta_\circ)$$

..contd

The above is something like this:

$$w_t = \theta_t' \beta + v_t$$

We regress $\epsilon_t(\theta^\circ)$ on - $\left.\frac{\delta \epsilon_t(\theta)}{\delta \theta}\right|_{\theta^\circ}$ & get $(\theta_\circ - \theta^\circ)$

$$\theta^\circ{}_{new} = \theta^\circ + (\theta_\circ - \theta^\circ)$$

- We want to minimise $(\theta_\circ - \theta^\circ)$
- Get $\theta^\circ{}_{new}$. Put it again in equation
- Regress and get new $(\theta_\circ - \theta^\circ)$
- Iterative process

Maximum Likelihood Estimators

- Given n observations

$$y_1, y_2, y_3, \ldots\ldots\ldots\ldots\ldots, y_n$$

  the likelihood L is the probability of obtaining the data actually observed.
- The MLE are those values of the parameters which make the observation of the data most likely event
- Nice asymptotic properties, deals with missing data easily
- Always lower variance than method of moments

Theory

- India accounts for 20% of all world rice production, second only to China (25%).
- Largest agricultural product by value-a staple diet-cum-source of earning for a majority of rural working population.
- Area (mha) under rice cultivation fairly stable (30-40) post-independence.
- Production (mt) up almost five times from 20.58 (1950-51) to 99.18 (2008-09).
- Terrific productivity gains - yield in Kg/Hectare increased from 668 in 1950-51 to 2178 in 2008-09.

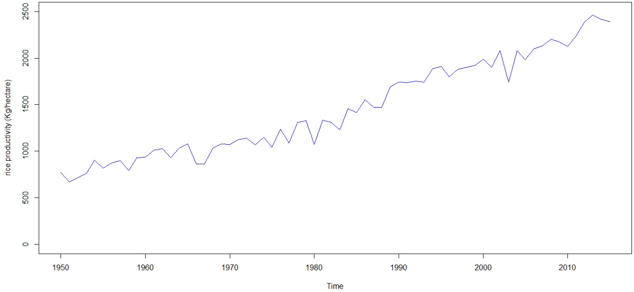
Statement of Purpose

- **Productivity** : Defined as yield of rice in Kg per Hectare.
- Time series modelling of yearly rice productivity from 1950-51 to 2014-15.
- Fitting AR(p) / MA(q) / ARMA(p,q) to the rice productivity data by use of ACF, PACF and AIC criteria.
- Estimating the coefficients of the appropriate model using Maximum Likelihood Estimator (MLE) and Non-Linear Least Squares (NLLS) technique.
- **Data source** : Directorate of Economics and Statistics, Ministry of Agriculture.

Literature Review

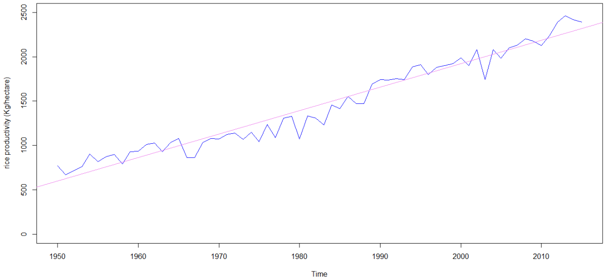
- In *"Application of ARIMA Model for Forecasting Agricultural Productivity in India"*, Purna Chandra Pradhan uses Box-Jenkins identification techniques and forecasts productivity for 35 agricultural products for the period 2011-2015.
- *Foster, et.al* (2007) provide worldwide forecasts of agricultural productivity growth till the year 2040 based on the latest time series evidence on total factor productivity growth for crops, ruminants, and nonruminant livestock.

Our Data: Rice Productivity in India Since Independence (1950/51-2014/15)



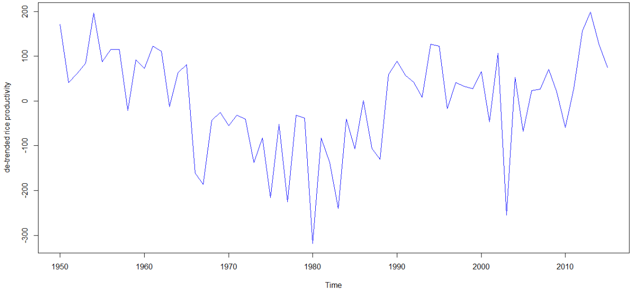
- It looks as though there might be a linear trend...
  - We must check for trend stationarity
- If seasonality is present, differencing may be a crude yet effective way of removing seasonality, as well as most of the trend.

Removing a Linear Trend



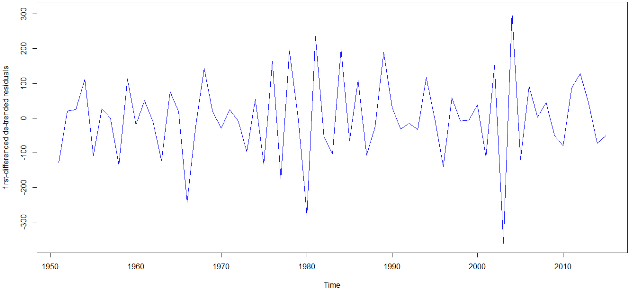
- The KPSS test for trend stationarity is significant.
    - KPSS Level $= 3.2953$, Truncation lag parameter $= 1$, p-value < 0.01

Looking at the De-Trended Component



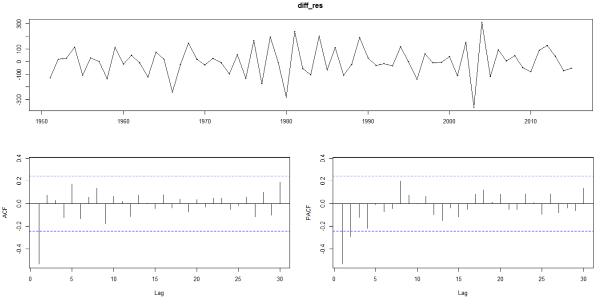
- In order to check for stationarity of this series, we perform the ADF test on it.
  - Dickey-Fuller = -1.5469, Lag order = 4, p-value = 0.7589
  - The R software tells us how many differences needed to get to stationarity with the *ndiffs()* command.

A Look at our Stationary Series



- Dickey-Fuller = -6.4471, Lag order = 3, p-value < 0.01

Identifying an Appropriate Model



- PACF indicates p = 2.
- ACF indicates q = 1.
- R's *auto.arima()* function identifies ARIMA(0,1,1) as the best fit.

AIC Matrix

p/q	0	1	2
0	....	**780.1837**	....
1	788.1823	781.0581	....
2	783.7276	782.9799	....

- Additionally we calculate the AIC for 5 different ARIMA models with first differencing (d=1).
- The table indicates ARIMA(0,1,1) to be the best model according to the AIC criterion.

Estimating an ARIMA(0,1,1) model

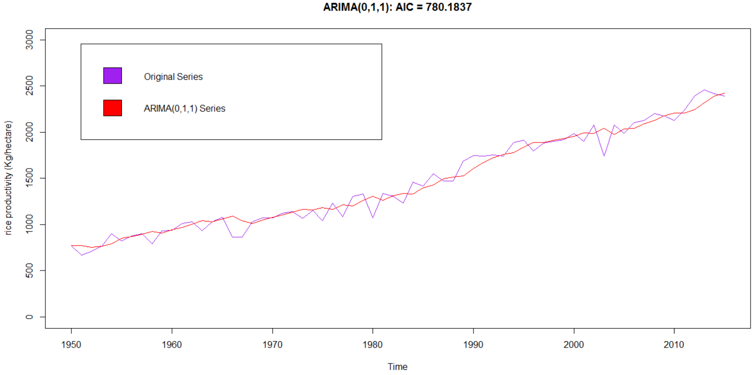
Coefficients:

	MA(1)
	-0.6845
s.e.	0.0797

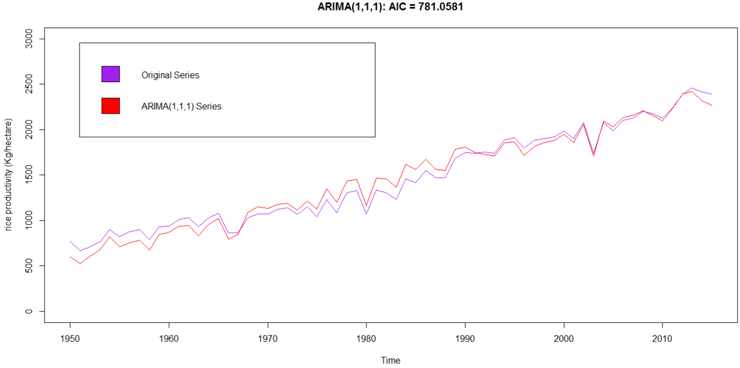
```
sigma^2 estimated as 9038:  log likelihood=-388.09
AIC=780.18   AICc=780.38   BIC=784.53
```

The ARIMA(0,1,1) model gives us a **highly significant MA(1) coefficient with minimum AIC** among possible other models.

Actual Series vs Fitted Series ARIMA(0,1,1)



Actual Series vs Fitted Series ARIMA(1,1,1)



References

- *Box, George; Jenkins, Gwilym (1970). Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day.*
- *Brockwell, Peter J.; Davis, Richard A. (1991). Time Series: Theory and Methods. Springer-Verlag.*
- *P.C. Pradhan, (2013). Application of ARIMA Model for Forecasting Agricultural Productivity in India.*
- *Cris Chattfield. The Analysis of Time Series: An Introduction.*
- *Foster, et.al. (2006). Productivity Growth and Convergence in Crop,Ruminant and Non-Ruminant Production: Measurement and Forecasts.*