International Journal of Sciences
Research Article
(ISSN 2305-3925)
Volume 2, Issue July 2013
http://www.ijSciences.com

Modelling Monthly Rainfall Data of Port Harcourt, Nigeria by Seasonal Box-Jenkins Methods

Ette Harrison Etuk[1] ✎, Imoh Udo Moffat[2], Benjamin Ele Chims[3]

[1] Department of Mathematics/Computer Science, Rivers State University of Science and Technology, Nigeria
[2] Department of Mathematics, Statistics and Computer Science, University of Uyo, Nigeria
[3] Department of Mathematics and Statistics, Rivers State Polytechnic, Bori, Nigeria

Abstract: Brief review of literature of the well documented seasonal Box-Jenkins modelling is done. Rainfall is a seasonal phenomenon the world over. For illustrative purposes, monthly rainfall as measured in Port Harcourt, Nigeria, is modelled by a $(5, 1, 0)x(0, 1, 1)_{12}$ seasonal ARIMA model. The time-plot shows no noticeable trend. The known and expected seasonality is clear from the plot. Seasonal (i.e. 12-point) differencing of the data is done, then a nonseasonal differencing is done of the seasonal differences. The correlogam of the resultant series reveals the expected 12-monthly seasonality, and the involvement of a seasonal moving average component in the first place and a nonseasonal autoregressive component of order 5. Hence the model mentioned above. The adequacy of the modelled has been established.

Keywords: Seasonal Time Series, ARIMA model, rainfall, Port Harcourt

1. Introduction

A time series is defined as a set of data collected sequentially in time. It has the property that neighbouring values are correlated and this tendency is called *autocorrelation.* A time series is said to be stationary if it has a constant mean and variance. Moreover the autocorrelation is a function of the lag separating the correlated values and is called the *autocorrelation function* (ACF).
A stationary time series $\{X_t\}$ is said to follow an *autoregressive moving average model of orders p and q* (designated ARMA(p, q)) if it satisfies the following difference equation

$$X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - \dots - \alpha_p X_{t-p} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q} \qquad (1)$$

Or

$$A(L)X_t = B(L)\varepsilon_t \qquad (2)$$

where $\{\varepsilon_t\}$is a sequence of random variables with zero mean and constant variance, called *a white noise process*, and the α_i's and β_j's constants; $A(L) = 1- \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p$ and $B(L) = 1 + \beta_1 L + \beta_2 L^2 + \dots + \beta_q L^q$ and L the backward shift operator defined by $L^k X_t = X_{t-k}$.

If p = 0, the model (1) becomes a *moving average model of order q* (designated MA(q)). If, however, q = 0 it becomes an *autoregressive process of order* p (designated AR(p)). Besides stationarity, invertibility is another important necessity for a time series. It ensures the uniqueness of the model covariance structure and, therefore, allows for meaningful expression of current events in terms of the past history of the series [1].

An AR(p) model may be more specifically written as

$$X_t + \alpha_{p1} X_{t-1} + \alpha_{p2} X_{t-2} + \dots + \alpha_{pp} X_{t-p} = \varepsilon_t$$

The sequence of the last coefficients $\{\alpha_{ii}\}$ is called the *partial autocorrelation* function (PACF) of {Xt}. The ACF of an MA(q) model cuts off after lag q whereas that of an AR(p) model is a combination of sinusoidals dying off slowly. On the other hand the PACF of an MA(q) model dies off slowly whereas that of an AR(p) model cuts off after lag p. AR and MA models are known to exhibit some duality relationships.

Parametric parsimony consideration in model building entails the use of the mixed ARMA fit in preference to either the pure AR or the pure MA fit. Conditions for stationarity and invertibility for model (1) or (2) are that the equations A(L) = 0 and B(L) =0 should have roots outside the unit circle respectively.

Often, in practice, a time series is non-stationary. Box and Jenkins [1] proposed that differencing of appropriate order could make a non-stationary series {Xt} to become stationary. Let degree of differencing

✎ Ette Harrison Etuk (Correspondence)
✉ ettetuk@yahoo.com

necessary for stationarity be d. Such a series {Xt} may be modeled as

$$A(L)\nabla^d X_t = B(L)\varepsilon_t \qquad (3)$$

where $\nabla = 1 - L$ and in which case $A(L)\nabla^d = 0$ shall have unit roots d times. Then differencing to degree d renders the series stationary. The model (3) is said to be an *autoregressive integrated moving average model of orders p, d and q* and designated ARIMA(p, d, q).

2. Seasonal ARIMA Models

A time series is said to be seasonal of order d if there exists a tendency for the series to show periodic behaviour after every time interval d. The time series $\{X_t\}$ is said to follow a multiplicative (p, d, q)x(P, D, $Q)_s$ *seasonal ARIMA model* if

$$A(L)\Phi(L^s)\nabla^d\nabla^D_s X_t = B(L)\Theta(B^s)\varepsilon_t \qquad (4)$$

where Φ and Θ are polynomials of order P and Q respectively. That is,

$$\Phi(L^s) = 1 + \phi_1 L^s + \ldots + \phi_P L^{sP} \qquad (5)$$

$$\Theta(L^s) = 1 + \theta_1 L^s + \ldots + \theta_q L^{sQ} \qquad (6)$$

where the ϕ_i and the θ_j are constants such that the zeros of the equations (5) and (6) are all outside the unit circle for stationarity or invertibility respectively. Equation (5) represents the autoregressive operator whereas (6) represents the moving average operator. Here $\nabla_s = 1 - L^s$.

Existence of a seasonal nature is often evident from the time plot. Moreover for a seasonal series the ACF or correlogram exhibits a spike at the seasonal lag. Box and Jenkins[1] and Madsen [2] and Etuk[3] are a few authors that have written extensively on such models. A knowledge of the theoretical properties of the models provides basis for their identification and estimation.

The purpose of this paper is to fit a seasonal ARIMA model to the Port Harcourt monthly rainfall totals. Osarumwense[4] has modelled the quarterly rainfall data as a (0, 0, 0)x(2, 1, $0)_4$ seasonal ARIMA model. Olofintoye and Sule[5] fitted the trend line y = 0.3903x – 587.5125 which is indicative of a positive trend for rainfall. A few other researchers who have published research results on Port Harcourt rainfall are Chiadikobi et al.[6], Dike and Nwachukwu[7] and Salako[8].

3. Materials and Methods

The data for this work are monthly rainfall totals from 1990 to 2006 obtainable from the meteorological centre of Port Harcourt International Airport.

3.1. Determination of the orders d, D, P, q and Q:

Seasonal differencing is necessary to get rid of the seasonal trend. If there is secular trend non-seasonal differencing will be necessary to remove it. In order for the model not to be too complicated it has been advised that orders of differencing d and D should add up to at most 2 (i.e. $d + D < 3$). The involvement of a seasonal AR component is suggestive if the ACF of the differenced series has a positive spike at the seasonal lag; if, however, it has a negative spike at the seasonal lag then a seasonal MA term is suggestive.

As already mentioned above, an AR(p) model has a PACF which truncates at lag p and an MA(q) has an ACF which truncates at lag q. In practice $\pm 2/\sqrt{n}$ where n is the sample size are the non-significance limits for both functions.

3.2. Model Estimation

The fact that items of a white noise process are involved in an ARIMA model entails a nonlinear iterative process in the estimation of the parameters. An optimization criterion like least error sum of squares, maximum likelihood or maximum entropy is used. An initial estimate is usually chosen. Each iterative step is expected to be an improvement of the last one until the estimate converges to an optimal one. However, for pure AR and pure MA models linear optimization techniques exist (See for example, Box and Jenkins[1], Oyetunji[9]. There are efforts to propose and adopt linear methods to estimate ARMA models (See for example, Etuk[10, 11]. We shall use Eviews software which employs the least squares approach involving nonlinear iterative techniques.

3.3. Diagnostic Checking

The estimated model should be tested for goodness-of-fit. This shall be done by some analysis of the residuals of the model. Should the model be correct, the residuals would be uncorrelated and would follow a normal distribution with mean zero and constant variance. The autocorrelations of the residuals should not be significantly different from zero.

4. Results and Discussion

The time plot of the original series RAINFALL in Figure 1 shows seasonality as expected but no secular trend. Seasonal (i.e. 12-month) differencing of the series produces a series SDRAINFALL with seasonality (see Figure 2). Non-seasonal differencing of SDRAINFALL yields a series DSDRAINFALL with seasonality (See Figure 3). Its ACF in Figure 4 has a negative spike at lag 12 revealing a seasonality of lag 12 and a seasonal MA component of order one to the model. The PACF shows spikes at the first five lags suggesting a non-seasonal AR component of order five. We therefore propose the seasonal model

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \alpha_3 X_{t-3} + \alpha_{t-4} X_{t-4} + \alpha_5 X_{t-5} = \varepsilon_t + \beta_{12}\varepsilon_{t-12} \qquad (7)$$

where X = DSDRAINFALL. The estimation of the model is summarized in Table 1. The fitted model is given by

$$X_t + 0.84X_{t-1} + 0.67X_{t-2} + 0.62X_{t-3} + 0.45X_{t-4} + 0.23X_{t-5} = \varepsilon_t - 0.89\varepsilon_{t-12}$$

All coefficients are significantly different from zero, each being larger than twice its standard error. As high as 72% of the variation in DSDRAINFALL is explained by the model. In Figure 5, the fitted model agrees closely with the actual data. In Figure 6, the correlogram of the residuals indicates model adequacy since virtually all the autocorrelations are non-significant.

5. Conclusion

The literature of seasonal Box-Jenkins modelling is briefly reviewed. A $(5, 1, 0)x(0, 1, 1)_{12}$ seasonal autoregressive integrated moving average model is fitted to rainfall data in Port Harcourt. The model has been shown to be adequate.

References

[1] Box, G. E. P. and Jenkins, G. M. (1976)." Time Series Analysis, Forecasting and Control", *Holden-Day*, San Francisco.

[2] Madsen, H. (2008)." Time Series Analysis", *Chapman & Hall/CRC*, London.

[3] Etuk, E. H. (2013). A Seasonal Time Series for Nigerian Monthly Air Traffic Data. International of Review and Research in Applied Sciences, 14, 3, pp. 596 – 602.

[4] Osarumwense, O.I. (2013). Applicability of Box Jenkins SARIMA Model in Rainfall Forecasting: A Case Study of Port-Harcourt South South Nigeria. *Canadian Journal on Computing in Mathematics, Natural Sciences, Engineering amd Medicine*, 4, 1, pp. 1-4-.

[5] Olofintoye, O. O. And Sule, B. F. (2010). Impact of Global Warming on the Rainfall and Temperature in the Niger Delta of Nigeria. *USEP-Journal of Research Information in Civil Engineering*, 7, 2, pp. 33 – 48.

[6] Chiadikobi, K. C., Omoboriowo, A. O., Chiaganam, O. I., Opatola, A. O. And Oyebanji, O. (2011). Flood Risk Assessment of Port Harcourt, Rivers State, Nigeria. *Advances in Applied Research*, 2, 6: pp. 287 – 298.

[7] Dike, B. U. And Nwachukwu, B. A. (2003). Analysis of Nigerian Hydrometeoroligical Data. *Nigerian Journal of Technology*, 22, 1, pp. 29 – 38.

[8] Salako, F. K. (2007). Rainfall seasonality in the Niger Delta Belt, Nigeria. *Journal of Geography and Regional*, 52, 2, pp. 51 – 60.

[9] Oyetunji, O. B. (1985)." Inverse Autocorrelations and Moving Average Time Series Modelling". *Journal of Official Statistics*, 1, pp. 315 – 322.

[10] Etuk, E. H. (1987). *On the Selection of Autoregressive Moving Average Models.* An unpublished Ph. D. Thesis, Department of Statistics, University of Ibadan, Nigeria.

[11] Etuk, E. H. (1998)." An Autoregressive Integrated Moving Average (ARIMA) Model: A Case Study". *Discovery and Innovation*, 10, 1 & 2, pp. 23 – 26.

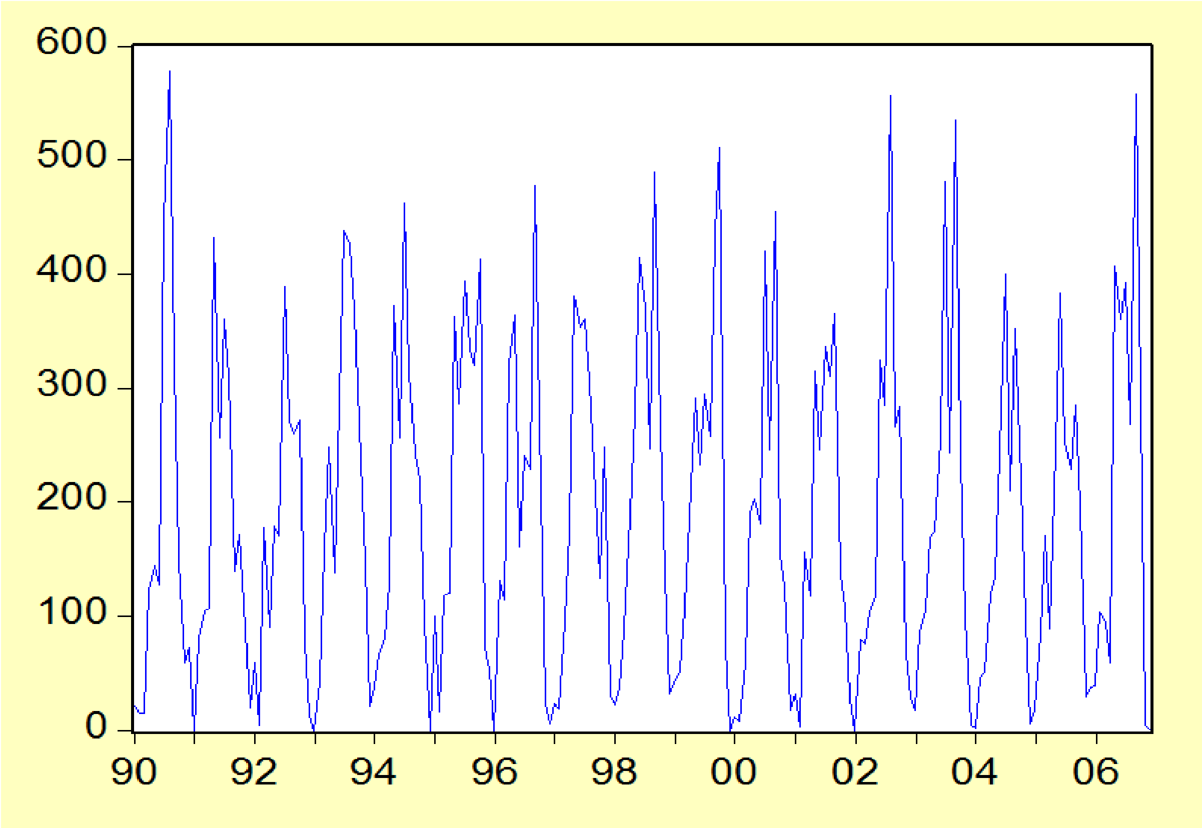


FIGURE 1: RAINFALL

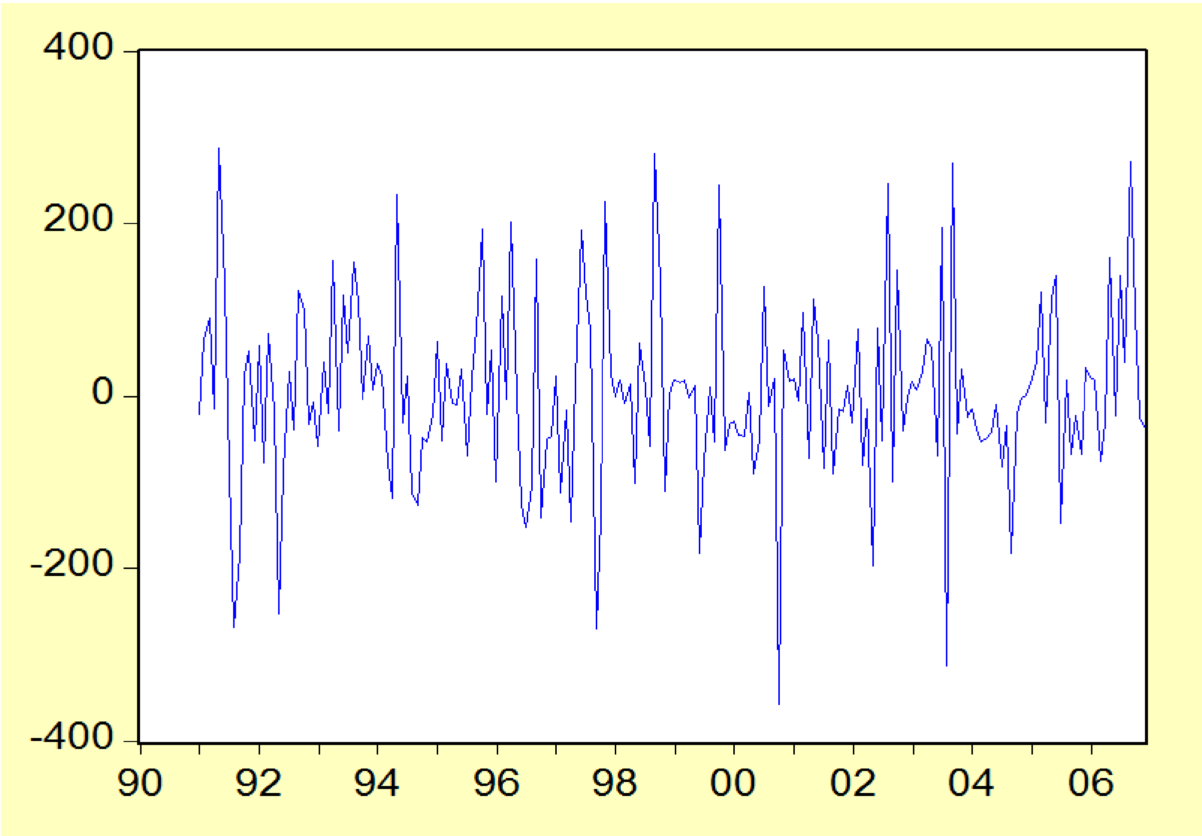


FIGURE 2: SDRAINFALL

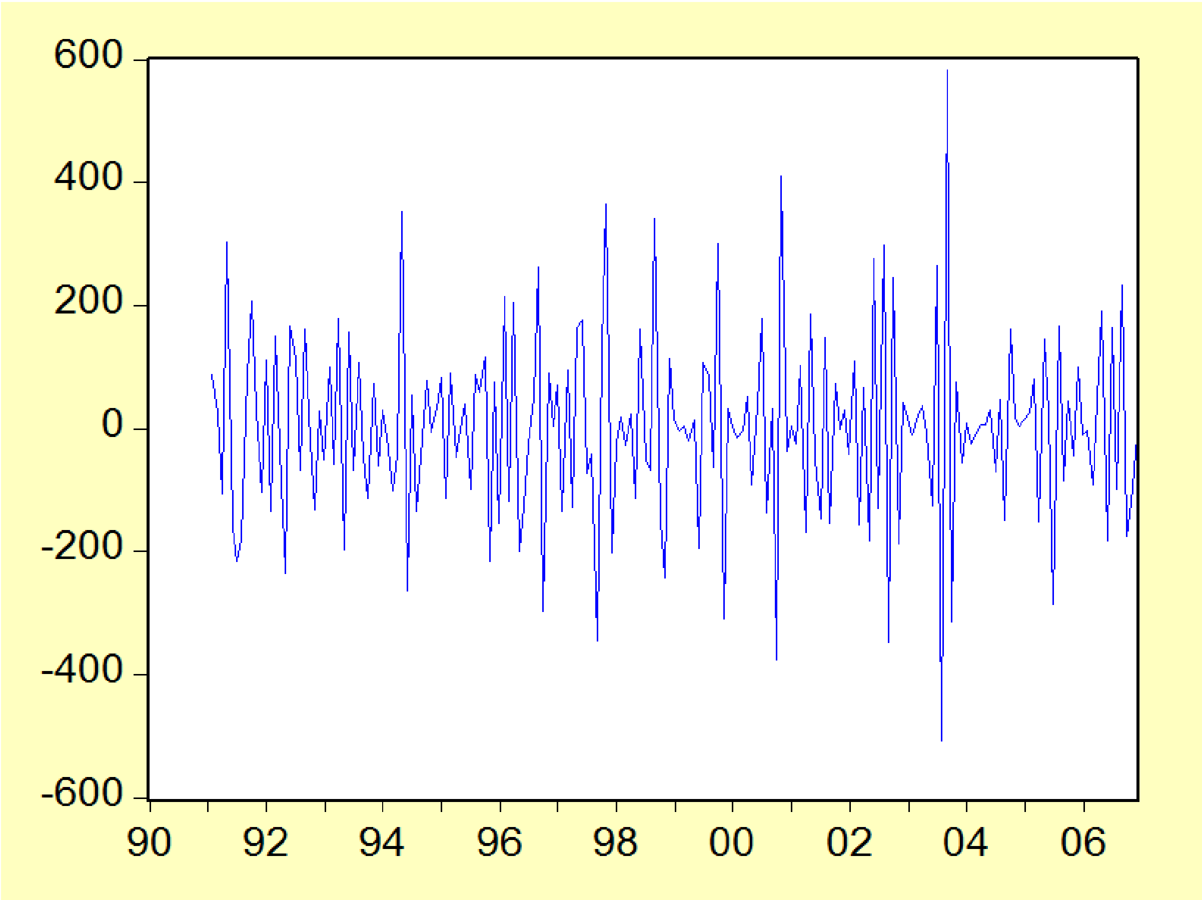
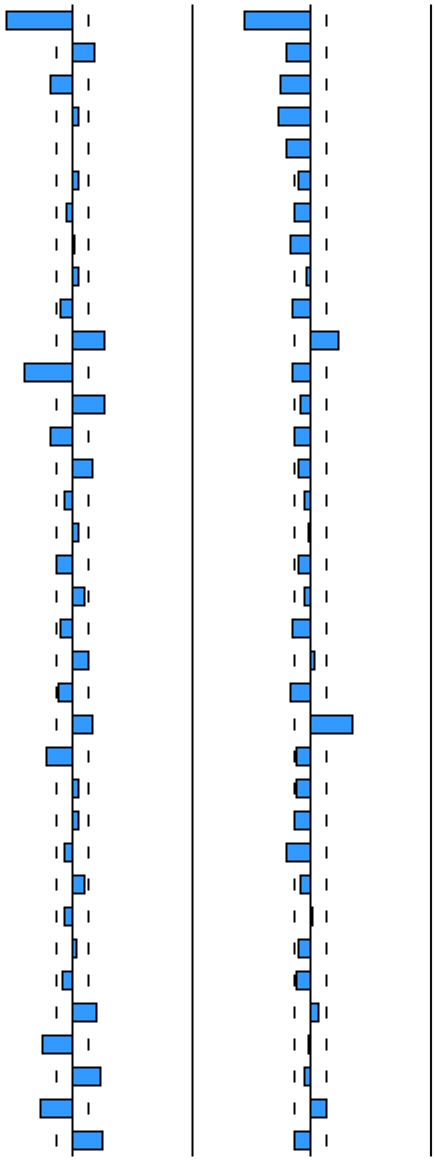


FIGURE 3: DSDRAINFALL

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob



Lag	AC	PAC	Q-Stat	Prob
1	-0.570	-0.570	63.059	0.000
2	0.189	-0.202	70.003	0.000
3	-0.186	-0.266	76.820	0.000
4	0.054	-0.271	77.391	0.000
5	-0.003	-0.210	77.393	0.000
6	0.054	-0.111	77.972	0.000
7	-0.056	-0.144	78.596	0.000
8	0.009	-0.167	78.611	0.000
9	0.055	-0.030	79.230	0.000
10	-0.112	-0.159	81.775	0.000
11	0.281	0.253	97.979	0.000
12	-0.420	-0.158	134.27	0.000
13	0.280	-0.095	150.49	0.000
14	-0.184	-0.139	157.57	0.000
15	0.174	-0.101	163.88	0.000
16	-0.060	-0.057	164.64	0.000
17	0.050	-0.010	165.18	0.000
18	-0.132	-0.114	168.88	0.000
19	0.115	-0.053	171.70	0.000
20	-0.103	-0.155	173.97	0.000
21	0.135	0.027	177.94	0.000
22	-0.115	-0.166	180.82	0.000
23	0.179	0.363	187.84	0.000
24	-0.222	-0.119	198.74	0.000
25	0.047	-0.116	199.24	0.000
26	0.047	-0.137	199.74	0.000
27	-0.058	-0.206	200.50	0.000
28	0.111	-0.082	203.31	0.000
29	-0.070	0.011	204.44	0.000
30	0.036	-0.101	204.73	0.000
31	-0.086	-0.124	206.43	0.000
32	0.199	0.063	215.63	0.000
33	-0.265	-0.014	232.07	0.000
34	0.244	-0.042	246.00	0.000
35	-0.285	0.133	265.14	0.000
36	0.254	-0.140	280.54	0.000

FIGURE 4: CORRELOGRAM OF DSDRAINFALL

Dependent Variable: DSDRAINFALL
Method: Least Squares
Date: 03/05/12 Time: 20:51
Sample(adjusted): 1991:07 2006:12
Included observations: 186 after adjusting endpoints
Convergence achieved after 19 iterations
Backcast: 1990:07 1991:06

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.839691	0.057295	-14.65545	0.0000
AR(2)	-0.671853	0.070072	-9.587996	0.0000
AR(3)	-0.622000	0.073157	-8.502286	0.0000
AR(4)	-0.449151	0.069994	-6.417028	0.0000
AR(5)	-0.229516	0.058884	-3.897737	0.0001
MA(12)	-0.885785	0.000120	-7376.413	0.0000

R-squared	0.715492	Mean dependent var	-0.897849
Adjusted R-squared	0.707589	S.D. dependent var	153.9822
S.E. of regression	83.26588	Akaike info criterion	11.71368
Sum squared resid	1247977.	Schwarz criterion	11.81774
Log likelihood	-1083.372	F-statistic	90.53437
Durbin-Watson stat	2.122132	Prob(F-statistic)	0.000000

Inverted AR Roots	.34+.73i	.34 -.73i	-.39+.57i	-.39 -.57i
	-.73			
Inverted MA Roots	.99	.86+.49i	.86 -.49i	.49 -.86i
	.49+.86i	.00 -.99i	-.00+.99i	-.49 -.86i
	-.49+.86i	-.86 -.49i	-.86+.49i	-.99

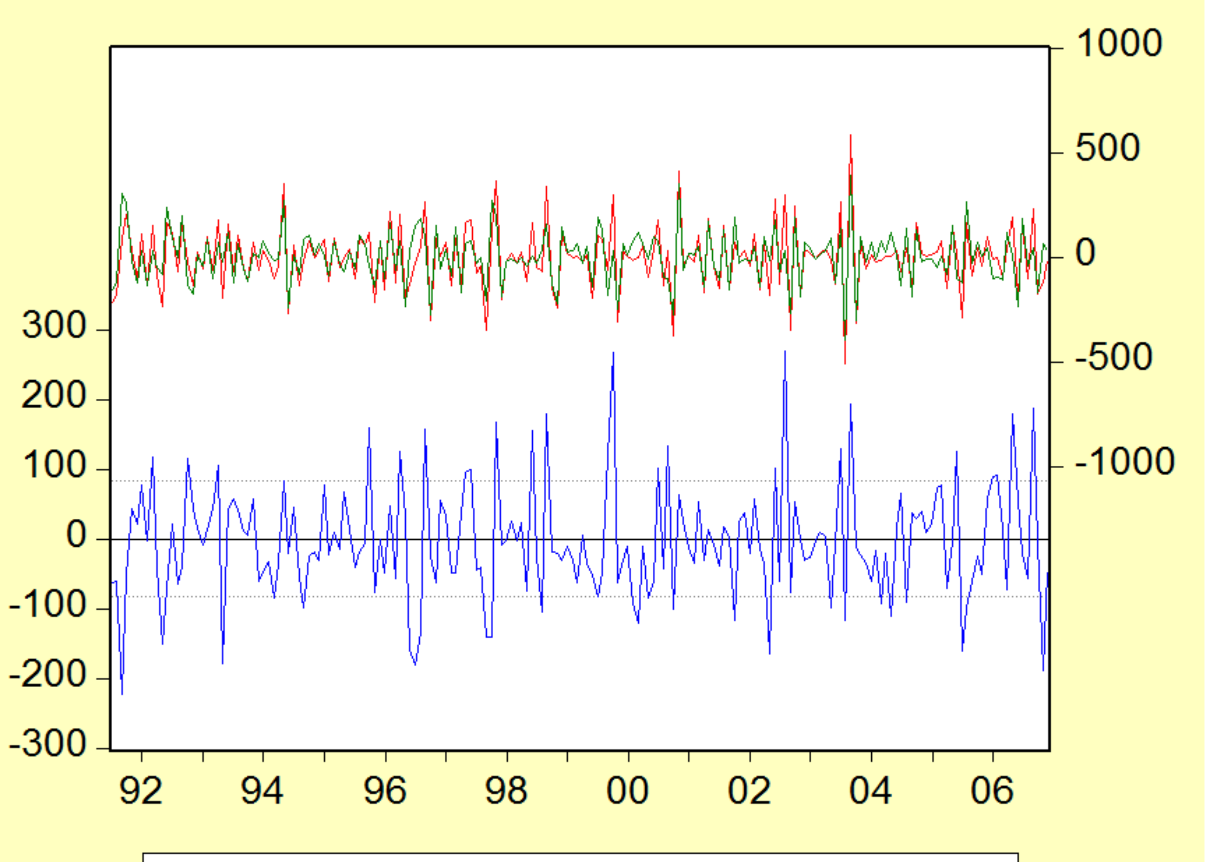
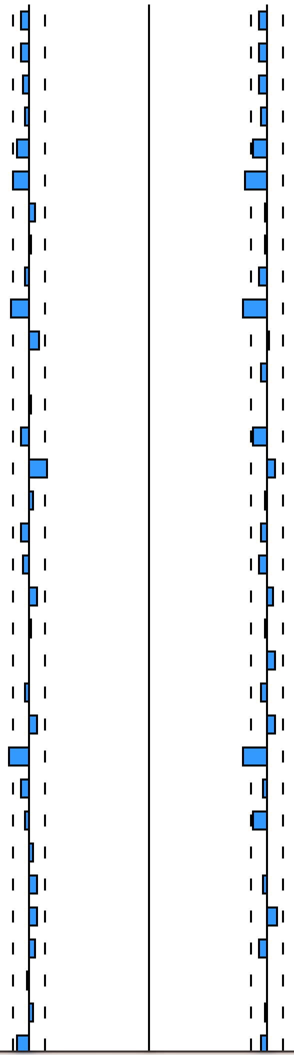


Fig. 5: Residual — Actual — Fitted

Date: 03/31/12 Time: 15:45
Sample: 1991:07 2006:12
Included observations: 186
Q-statistic probabilities adjusted for 6 ARMA term(s)

Autocorrelation | Partial Correlation



	AC	PAC	Q-Stat	Prob
1	-0.066	-0.066	0.8305	
2	-0.065	-0.069	1.6221	
3	-0.050	-0.060	2.0965	
4	-0.033	-0.046	2.3014	
5	-0.114	-0.129	4.7973	
6	-0.147	-0.181	9.0070	
7	0.045	-0.010	9.4052	0.002
8	0.018	-0.024	9.4721	0.009
9	-0.027	-0.063	9.6203	0.022
10	-0.150	-0.203	14.084	0.007
11	0.098	0.012	15.996	0.007
12	-0.002	-0.057	15.996	0.014
13	0.019	-0.007	16.070	0.024
14	-0.075	-0.123	17.221	0.028
15	0.148	0.080	21.698	0.010
16	0.026	-0.012	21.839	0.016
17	-0.070	-0.046	22.857	0.019
18	-0.047	-0.077	23.323	0.025
19	0.077	0.052	24.570	0.026
20	0.016	-0.009	24.625	0.038
21	-0.007	0.059	24.636	0.055
22	-0.027	-0.055	24.791	0.074
23	0.062	0.058	25.604	0.082
24	-0.178	-0.218	32.414	0.020
25	-0.079	-0.028	33.759	0.020
26	-0.030	-0.123	33.956	0.026
27	0.039	-0.007	34.293	0.034
28	0.060	-0.040	35.093	0.038
29	0.061	0.085	35.926	0.042
30	0.053	-0.070	36.565	0.048
31	-0.010	-0.007	36.587	0.063
32	0.030	-0.022	36.798	0.078
33	-0.110	-0.045	39.544	0.056

FIGURE 6: CORRELOGRAM OF RESIDUALS