Pedicting rainfall and drought at least a year in Advance in Zimbabwe using climatic determinants (Darwin and Southern Oscillation Indices)

Delson Chikobvu[1] and Retius Chifurira[2]

Abstract

In this study, a multiple regression model is developed to explain and predict mean annual rainfall in Zimbabwe. Principal component analysis is used to construct orthogonal climatic factors which influence rainfall patterns in Zimbabwe. The aim of the study is to develop a simple but reliable tool to predict annual rainfall one year in advance using Darwin Sea Level Pressure (Darwin SLP) value of a particular month and a component of Southern Oscillation Index (SOI) which is not explained by Darwin SLP. A weighted multiple regression approach is used to control for heteroscedasticity in the error terms.The model developed has a reasonable fit at the 5% statistical significance level can easily be used to predict mean annual rainfall at least a year in advance.

Key words: Principal Component Analysis, multiple regression, Darwin Sea level Pressure, Southern Oscillation Index, previous year

INTRODUCTION

The prime cause of drought is the occurrence of below normal precipitation, which is affected by various natural phenomena. Firstly, as noted by Panu and Sharma (2002), the most notable large scale climatic variation that occur from one year to another, is the Southern Oscillation climatic condition, which manifests itself in the differential oceanic temperature phenomenon across the tropical Pacific Ocean. The SOI is the difference between seasonally normalised sea level pressures of Darwin (in Australia) and Tahiti (in Mid Pacific). Secondly, Darwin Sea Level Pressures has been found to influence seasonal rainfall patterns in Zimbabwe (Manatsa et al., 2007).

The aim of this study is to develop a simple rainfall predictive model using climatic determinates such as Southern Oscillation Index (SOI) and Darwin Sea Level Pressures (Darwin SLP) for a country such as Zimbabwe at least a year in advance.

[1]University of Free State, Department of Mathematical Statistics and Actuarial Science Box 339, Bloemfontein 9300, RSA. Email:chikobvu@ufs.ac.za

[2]University of Zululand, Department of Mathematical Sciences P. Bag X1001.KwaDlangezwa, 3886, RSA. Email: chifurira@gmail.com

BACKGROUND

Darwin Sea Level Pressure

Sea Level Pressure is the atmospheric pressure at mean sea level either directly measured by stations at sea level or empirically determined when the station is not at sea level. The monthly Darwin SLP values were sourced from the internetwebsite http://www.cpc.ncep.noaa.gov/data/indices/da . In this paper, the Darwin SLP values are not directly measured at Darwin but empirically determined using a reduction formula.

Southern Oscillation Index

The SOI data is obtained from the Internet website*http://www.longpaddock.qld.gov.au.* The SOI is calculated from the monthly or seasonal fluctuations in the air pressure difference of the area between Tahiti (in the mid-Pacific) and Darwin (in Australia). The SOI gives a simple measure of the strength and phase of the difference in sea-level pressure between Tahiti and Darwin.

METHODOLOGY

Multiple regression

Principal Component Analysis

Principal component analysis is a technique used to combine highly correlated factors into principal components that are much less highly correlated with each other. This improves the efficiency of the model.

In this study, the predictive power of Darwin Sea Level Pressure values $(I_1)$and SOI values (I_2) is explored. Two new, uncorrelated factors,$I_1^*$and I_2^*, can be constructed as follows:

Let $I_1^* = I_1$

Then, we carry out a linear regression analysis to determine the parameters $\gamma_1$and γ_2 in the equation:

$$I_2 = \gamma_1 + \gamma_2 I_1^* + \varepsilon_1 \tag{1}$$

$\gamma_1$andγ_2 are the intercept and slope parameters of the regression model respectively and ε_1 is the ‘error’ term, which by definition is independent of $I_1^* = I_1$.

We then set:

$$I_2^* = \varepsilon_1 = I_2 - (\gamma_1 + \gamma_2 I_1^*)$$

By construction I_2^* is uncorrelated with Darwin Sea Level Pressure values (I_1) since$I_2^* = \varepsilon_1$, the residual term in the equation. Changes in I_2^* is interpreted as the change in the observed values of SOI(I_2) that cannot be explained by the observed change in Darwin Sea Level Pressure values (I_1).I_2^*in the rainfall model (equation 1) explains the component of rainfall that cannot be explained by the Darwin SLP.

Another assumption in the rainfall model (equation 1) is that the error terms should be uncorrelated and have constant variance over time. This assumption is likely to be violated in regression models with time series data. Autocorrelation (the error terms being correlated among themelves through time) leads to regression coefficients which are unbiased, inefficient and the standard errors are probably wrong making t tests and F tests unreliable.

Weighted least squares regression

The multiple least squares criterion weighs each observation equally in determining the estimates of the parameters. The procedure treats all of the data equally, giving less precise measured points more influence than they should have and gives highly precise points too little influence. The weighted least squares weighs some observations more heavily than others, giving each data point its proper amount of influence over the parameter estimates, and this maximizes the efficiency of parameter estimation. Weighted least square reflects the behaviour of the random errors in the model.

The model $y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + N_t$

Let $\boldsymbol{Y} = [y_1, y_2, \dots . y_t]'$, $\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2]'$, $\boldsymbol{X} = \begin{bmatrix} 1 & x_{1,1} & x_{2,1} \\ \dots . & \dots . & \dots . \\ 1 & x_{1,t} & x_{2,t} \end{bmatrix}$ and $\boldsymbol{N} = [N_1, N_2 \dots N_t\ ]\ '$ thenthe same model equation can be written as

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{N}$$

Parameter estimates using ordinary least squares can be found as

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^/\boldsymbol{X})^{-1}(\boldsymbol{X}^/\boldsymbol{Y}) = [\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2]'.$$

To find the weighted least squares parameters of the weighted model, we minimise the Weighted Sum of Squared Errors.

$$WSSE = \sum_{t=1}^{n} w_t (y_t - \hat{y}_t)^2$$

$$= \sum_{t=1}^{n} w_t \, (y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{1,t} - \hat{\beta}_2 x_{2,t})^2$$

where $w_t > 0$ is the weight assigned to the t^{th} observation. The weight w_t can be the reciprocal of the variance of that observation's error term, σ_t^2 , i.e.,

$$w_t = \frac{1}{\sigma_t^2}$$

Observations with larger error variances will receive less weight (and hence have less influence on the analysis) than observations with smaller error variances. The estimates are:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^{/}\boldsymbol{W}\boldsymbol{X})^{-1}(\boldsymbol{X}^{/}\boldsymbol{W}\boldsymbol{Y})$$

where $\boldsymbol{W} = [w_1, w_2, \dots w_t]$ is the weight vector.

The biggest disadvantage of weighted least squares is the fact that the theory behind this method is based on the assumption that the weights are known exactly. This is almost never the case in real applications where, instead, estimated weights are used (Carrol and Ruppert, 1988).

RESULTS

Multiple regression

Table 1 shows the results of the multiple regression approach to predict Zimbabwe's mean annual rainfall using the Darwin SLP for March at a lag of one year and the principal component of SOI for September of the same year which is not explained by Darwin SLP. The multiple regression model is:

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1,t-1} + \hat{\beta}_2 x_{2,t-1}$$

Where $\hat{y}_t$ is the predicted mean annual rainfall, $x_{1,t-1}$ is Darwin SLP value for March of the previous year and $x_{2,t-1}$ is I_2^* the component of SOI for September of the previous yearwhich is not explained by the corresponding Darwin SLP value for March.

Table 1: Multiple regression model for predicting mean annual rainfall

$\boldsymbol{Variable}$	$\boldsymbol{coefficient}$	$\boldsymbol{pvalue}$
	$\hat{\beta}_0 = 995.1645$	0.0000
$Darwin_{marc(-1)}$	$\hat{\beta}_1 = -43.324$	0.0162
I_2^*	$\hat{\beta}_2 = 4.5628$	0.0629

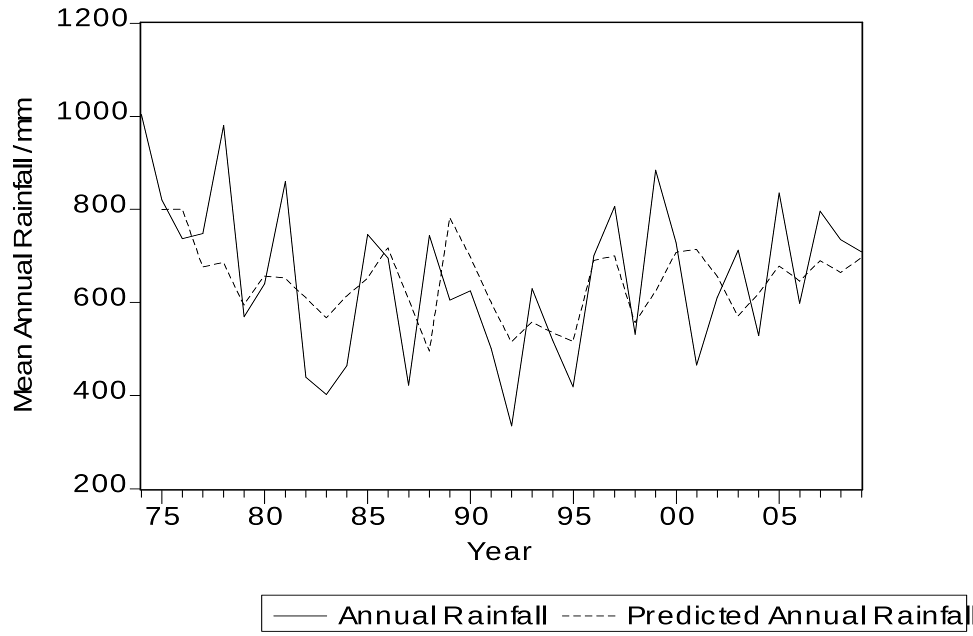

Figure 1: Rainfall versus predicted rainfall using multiple regression model

Weighted regression model

To address the heteroscedacity in the data, weighted regression is used. The weighted linear regression models results for rainfall $\hat{y}_t^*$:

$$\hat{y}_t^* = \hat{\beta}_0^* + \hat{\beta}_1^* x_{1,t-1} + \hat{\beta}_2^* x_{2,t-1}$$

Where $x_{1,t-1}$ is Darwin SLP values for March of the previous year and $x_{2,t-1}$ is I_2^* the component of SOI for September which is not explained by Darwin SLP for March. Various weights are considered in arriving at estimates$\hat{\beta}_0^*$, $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$ using weighted regression.

The model with $\frac{1}{I_2^{*3}}$ as the variance stabilizing weight was selected, because it is the model with the least AIC and BIC. The model is significant at 5% significance level. The estimates of $\hat{\beta}_0^*$, $\hat{\beta}_1^*$and $\hat{\beta}_2^*$ are 730.603, $-$ 13.261 and $-$18.120 respectively. The model has a much improved multiple $R^2 = 0.99$.

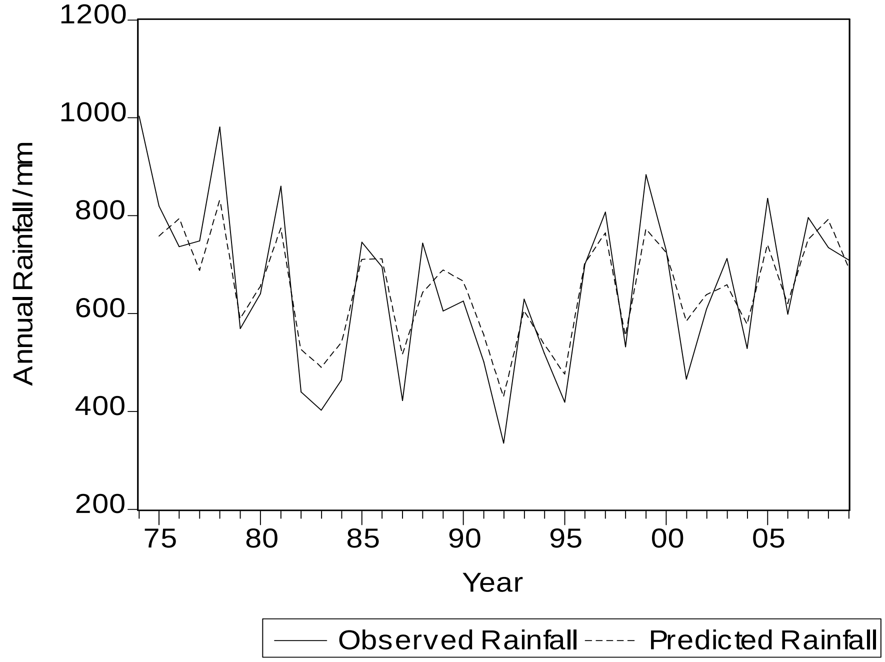

Figure 2: Rainfall versus Rainfall forecast

From figure 2 the model seems to be able to predict in-sample mean annual rainfall. The model shows little variability inbetween forecasts and actual rainfall.The model has a multiple $R^2 = 0.99$. Thus, the model can be used to predict one year ahead mean annual rainfall for Zimbabwe. The out of sample forecasts will be done for the years 2010 to 2013.

REFERENCES

Manatsa, D ,Chingombe, W , Matsikwa, H and Matarira, C. H. (2007). The superior influence of the Darwin Sea Level Pressure anomalies over ENSO as a single drought predictor in Southern Africa.*Theoretical and Applied Climatology D01 10.1007/s00 704 – 007 – 0315 − 3*.www.springerlink.com/index/8703666K421773hpdf

Panu, U.S. and Sharma, T.C. (2002). Challenges in drought research: some perspectives and future directions, *Hydrological Sciences Journal 47(S).*