Dynamic Linear State Space modeling Approach with Application to Annual Rainfall Data

Ibrahim Lawal Kane [1], Ibrahim Mu'awiyya Idris[2]

[1] Department of Mathematics and Statistics, Umaru Musa Yar'adua University, Katsina State, Nigeria

[2] Department of Mathematics and Statistics, Umaru Musa Yar'adua University, Katsina State, Nigeria

Abbreviated abstract: The intricate nature of rainfall has been appreciated for decades. This paper is based on state space approach using dynamic linear models (DLM) that provide an adequate model capable of modeling annual rainfall over Katsina, Nigeria from 1949 to 2019. The method allows a natural interpretation of data as the combination of trend, seasonal and regressive components. The validated DLM Model was used for one-step-ahead forecast. Comparison of the observed and predicted series shows the model closely simulate the actual data values.

Related publications:

– Petris, G. *Journal of Statistical Software*, ***36***,12 (2010).

– Umar, S., Lone, M.A. & Goel, N.K. *Water Resour. Manag.* **7**, 59 (2021).

4th Conference on Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022

ibrahim.lawal@umyu.edu.ng

Introduction

- The intricate nature of rainfall has been appreciated for decades. It has being viewed as foremost examples of highly non-linear systems and apprehended as complex systems. In recent years the phenomena of rainfall is not steady.
- Even though, numerous time series models were used in modelling rainfall e.g. (Anthony & Clement 2018; Aieb, et.al., 2020), this paper introduce a different State Space Approach that is powerful however, abandoned; State Space Dynamic Linear model (SSDLM), this method can be applied to non static data without need for a preliminary transformation. It allows natural interpretation of data as combination of trend, seasonal and regressive components. It allows extrapolations to be carried out exactly using Kalman filtering.
- The model is to annual rainfall data from from 1949 to 2019 of Katsina metropolis being in a Tropical Continental region situated in North West zone of Nigeria.
- DLMs are a large class of models that, among other features, allow for time-varying parameters. Thus, they do not need to assume a constant functional relationship between covariates, its embed the temporal dependence within the functional relationship and alleviate the need for numerous lags of a covariate to account for the temporal nature of the data thereby producing cleaner parameter interpretation (Osthus et al., 2014).

4th Conference on Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022

Methods

Consider the annual rainfall data Y_t. It is assume that the series can be written as the sum of independent components such as:

$$Y_t = Y_{1,t} + \cdots +, Y_{j,t} \quad (1)$$

where $Y_{1,t}$ represent a trend component, $Y_{2,t}$ a seasonal component, and so on.

The $i - th$ component in $Y_{i,t}$, $i = 1, \ldots, j$, might be described by a DLM as follows :

$$Y_{i,t} = F_{i,t}\theta_{i,t} + v_{i,t} \qquad v_{i,t} \sim N(0, V_{i,t}) \quad (2)$$

$$\theta_t = G_{i,t}\theta_{i,t-1} + w_{i,t} \qquad w_{i,t} \sim N(0, W_{i,t}) \quad (3)$$

where the $(p_1 \times 1)$ state vectors $\theta_{i,t}$ are distinct and $(Y_{i,t}, \theta_{i,t})$ and $(Y_{j,t}, \theta_{j,t})$ are mutually independent for all $i \neq j$. The components of DLM's are then combined for obtaining the DLM for (Y_t). By the assumption of independence of the components, it is easy to show that $Y_t = \sum_{i=1}^{h} Y_{i,t}$ described by the DLM as;

$$Y_t = F_t\theta_t + v_t, \quad v_t \sim N(0, V_t) \quad (4)$$

$$\theta_t = G_t\theta_{t-1} + w_t, \quad w_t \sim N(0, W_t) \quad (5)$$

where G_t and F_t are known matrices (the *evolution matrix* which pre-multiplies the previous period's state vector and the v_t and w_t are two independent white noise sequences with mean zero and known covariance matrices V_t and W_t respectively. In order to specify a DLM, the parameters; G_t, F_t, V_t and W_t must be specified for each period t (Petris, 2010). The filtering distribution of θ_t is the distribution of $\theta_t / y_1, y_2, \ldots, y_t$, while the smoothing distribution of θ_t at time s is the conditional distribution of $\theta_t / y_1, y_2, \ldots, y_s$, for s ≥ t., under the assumptions that the distributions are Gaussian, therefore completely determine by their means and variances.

The Kalman filter algorithm follows:

$$\alpha_t = c_t + T_t\alpha_{t-1} + R_t\eta_t \quad (6)$$

$$y_t = d_t + Z_t\alpha_t + \varepsilon_t \quad (7)$$

where $\eta_t \sim N(0, Q_t)$ and $R_t \sim N(0, H_t)$. The repeated reference is made in the outcome:

$$\alpha_{t-1} = E\left[\alpha_{t-1} \middle| y_0, \ldots, y_{t-1}\right] \quad (8)$$

$$P_{t-1} = E[(\alpha_{t-1} - a_{t-1})(\alpha_{t-1} - a_{t-1})^T] \quad (9)$$

The estimates of the state vector and its covariance matrix at time t with information available at time $t - 1$, $a_{t|t-1}$ and $P_{t|t-1}$ respectively, are given by the time update equations:

$$a_{t|t-1} = T_t a_{t-1} + c_t \quad (10)$$

$$P_{t|t-1} = T_t P_{t-1} T_t^T + R_t Q_t R_t^T \quad (11)$$

Let $F_t = Z_t P_{t|t-1} Z_t^T + H_t$. If a new observation is available at time t, then $a_{t|t-1}$ and $P_{t|t-1}$ can be updated with the measurement update equations:

$$a_t = a_{t|t-1} + P_{t|t-1} Z_t^T F_t^{-1}(y_t - Z_t a_{t|t-1} - d_t) \quad (12)$$

$$P_t = P_{t|t-1} - P_{t|t-1} Z_t F_t^{-1} Z_t^T P_{t|t-1} \quad (13)$$

For convenience the DLM 's can be written by the joint density of the observations in the form:

$$p(y_1, \ldots, y_n; \psi) = \prod_{t=1}^{n} p(y_t | D_{t-1}; \psi) \quad (14)$$

where $p(y_t | D_{t-1}; \psi)$ is the conditional density of y_t given the data up to time t-1, assuming that ψ is the value of the unknown parameter. Therefore the likelihood equation can be written as;

$$\ell(\psi) = -\frac{1}{2}\sum_{t=1}^{n} log|Q_t| - \frac{1}{2}\sum_{t=1}^{n}(y_t - f_t)' Q_t^{-1} (y_t - f_t) \quad (15)$$

where the f_t and the Q_t depend implicitly on ψ. The expression (15) can be numerically maximized to obtain the maximum likelihood estimator (MLE) of ψ:

$$\hat{\psi} = argmax\ \ell(\psi) \quad (16)$$

4th Conference on Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022

Results and Conclusions

Table 1 Annual Rainfall DLM Parameters

M_0	C_0	FF	V	GG	W	$r = W/V$
0	1.0×10^7	1.0	1.6514×10^4	1.0	4.488×10^3	0.271

The parameters M_0 and C_0 are mean and the variance of the prior distribution respectively, FF is the covariates, V is observational variance, GG is evolution and W is the evolution variance. The behavior of the process is greatly influenced by the signal-to-noise ratio r the ratio between the two error variances. The smaller the r – value the nearer estimation point can be obtained. The estimation and forecasting point obtained recursively by Kalman Filter. Based on the parameters in table 1, the annual Katsina rainfall forecasting model is given by;

$$Y_t = \mu_t + v_t; \quad v_t \sim N(0, 1.6514 \times 10^4)$$

where $\mu_t = \mu_{t-1} + w_t;\ w_t \sim N(0, 4.488 \times 10^3)$ with the Prior distribution,

$(\mu_0 | D_0) \sim N(0,\ 1 \times 10^7)$.

CONCLUSION

Being it a flexible model, it fairly repeats the data behaviour over time with a considerably little variation. The adequacy of the model was realized via its residual, the qq-plot that compare the observed series and the forecasted through plotting their quartiles against each other shows almost straight line which is an indication of similarity in the distributions of observed and the forecasted.

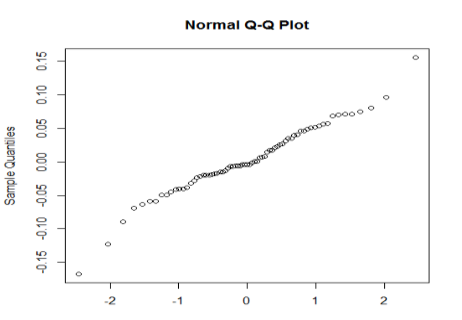

Figure 1

These is indication among others of goodness of fit and the adequacy of the model. The fitted model was used for one-step ahead forecast which the forecast plot shows the forecasted values as true representation of the observed data. Therefore, the model is a notable representation of the annual rainfall data.

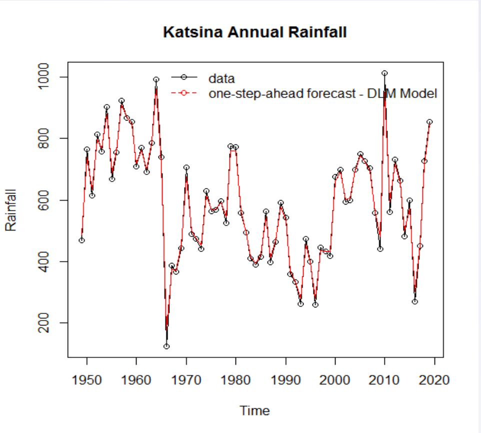

Figure 2 one-step ahead forecast plot

UMARU MUSA YAR'ADUA UNIVERSITY KATSINA

4th Conference on Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022