Modified State Space Modelling with Nonlinear State Equation

R. A. Yemitan[1]; O. I. Shittu[2]

Department of Statistics,
University of Ibadan,
Ibadan, Nigeria.
e-mail: ryemitan@gmail.com[1]; oi.shittu@hotmail.com[2]

***Abstract*— The Classical State Space Models (CSSM) allow researchers to model a time series (y_t), explained by a vector of stochastic variables (SV), using the Kalman Filter (KF) technique. The evolution of the Structural Equation (SE) is commonly assumed to be linear particularly of Markov process, which does not address non-linear dynamics of today's world challenges, hence, gives less precise estimates and low forecast performance. This paper introduces the Modified SSM (MSSM) with Smooth Transition Autoregressive (STAR) models as SE within the nonlinear framework to ameliorate the limitations of the linear SE. The MSSM was applied to Nigeria's CPI and GDP data as well as three simulated data sets. The MSSM captured the nonlinearity of the systems using LM test, with results for CPI, GDP and n=500 as (0.33, 0.06 and 0.36) showing nonlinearity while for n=250 and n=1000 were (0.00 and 0.03,), which implies linearity. The AIC for CPI in terms of the MSSM and CSSM were (23.02 and 25.84), MAPE; (0.43 and 0.60) and RMSE; (2.13 and 3.65) respectively. Results for GDP were AIC; (565.14 and 970.19), MAPE; (0.77 and 3.80) and RMSE (2.18 and 10.54) respectively. Results for n=500 gave AIC; (3,957 and 5,161), MAPE; (-14.9 and 80.5) and RMSE; (12.6 and 41.2) respectively. The Modified State Space Model with its key attributes has improved inference for nonlinear phenomena. The Smooth Transition Autoregressive model has been shown to aid the evolution of a given system.**

Keywords: *Kalman Filter, Classical State Space Models, Modified State Space Models..*

I. Introduction

The goal of the paper is to introduce the smooth transition autoregressive (STAR) modelling approach for estimating nonlinear (especially intrinsically nonlinear) systems in the state space modelling framework. The STAR model as developed by [9] can be adjusted to suit linear and non-linear systems. This unique feature forms the basis of rethinking the Kalman filtering methodology. Hence, we describe the Kalman filter (KF), existing KF for non-linear systems and present our methodology of Kalman Filtering using the STAR methodology.

II. Research Methodology

A. The Kalman Filter

KF also known as linear Gaussian state space model (or just state space model – SSM) can be described the two system of equations.

$$x_{t+1} = \varphi x_t + \omega_t \qquad (1)$$
$$y_t = H x_t + v_t \qquad (2)$$

Time is indexed by the discrete index t. The output y_t is a linear function of x_{t+1}which is linearly dependent on its previous state,x_t. Hence, (1) is termed the measurement or observation equation and (2) is referred to as the state or transition equation. Both the state and the measurement noise $\omega_t$and v_t are zero-mean normally distributed random variables with covariance matrices Q and R respectively. Only the output of the system is observed, the state and all the noise variables are hidden. The KF estimation procedure is detailed in [7] and [1].

B. Existing KF for Nonlinear systems

The Extended Kalman filter (EKF) and the Unscented Kalman Filter (UKF) have been the widely used KF for nonlinear systems.

The EKF as well as the UKF are used to solve the estimation problem for any intrinsically linear systems. The considered nonlinear system is represented by:

$$x_{t+1} = f(x_t) + \omega_t \qquad (3)$$
$$y_t = h(x_t) + v_t \qquad (4)$$

where x_{t+1}, y_t, v_t, and ω_t remain as explained in KF above. The nonlinear mapping f(.) and h(.) are assumed to be continuously differentiable with respect to x_t. Like KF, the EKF and UKF procedure for estimation are the same once the state equation (3) has been linearized or an approximate linear alternative is available. The UKF has been proved superior to the EKF when solving the nonlinear system in recent years. Nonetheless, the limitation of these approaches of the KF to intrinsically linear systems calls for alternative approaches of the KF to any nonlinear system.

C. The Modified State Space Model with Smooth Transition Autoregressive State Equation

Given equations 1 and 2, the STAR of [7] is used as the transition equation. Hence the model takes the form

$$y_{t+1} = Hx_{t+1} + v_{t+1} \dots (1^*)$$
$$x_{t+1} = \varphi_1 x_t[1 - G(x_t)] + \varphi_2 x_t G(x_t) + Q_t\omega_t \dots (2^*)$$

Where $v_t$and w_t are uncorrelated, zero-mean, white random process. They are uncorrelated with the initial statex_0. $Q_t$is a known matrix and x_t and x_{t+1} are known vectors, non-linear function of the state $G(x_t)$ is the transition function of the state bounded between 0 and 1. This property makes it possible to estimate the two extreme states also a continuum of states that lie between those two extremes

D. Filter Development

From equations (1^*) and (2^*), the system of the form in (2^*) is considered andω_t is the *system input or system disturbance* term with zero mean, white noise random process that is uncorrelated with the initial state x_0 and has a covariance matrix Ω

The aim is to get the value of x_{t+1} which is not directly measurable or observable. However, y_{t+1} is the only measurable quantity that is related to $x_{t+1}$by(1^*)

$$y_{t+1} = Hx_{t+1} + v_{t+1}$$

Where 'H' is a known constant, v_{t+1} is a zero mean, white noise process that is uncorrelated with w_t, the initial state and has a covariance ofR_{t+1}. $v_{t+1}$is the measurement or observation noise. Therefore, x_{t+1} is a function of the measurements$y_1, y_2, \dots y_{t+1}$ in the form

$$\hat{x}_{t+1} = g(y_1, y_2, \dots y_{t+1}) = g(.)$$

Where $g(.)$ is in the traditional State Space Model (SSM) an AR(1) or random walk with drift process. This is not so MSSM, but can be possible by making some assumptions in the desired STAR process. Hence, the introduction of $g(.)$ to be a smooth transition autoregressive process instead of the traditional random walk model, i.e.

$$x_{t+1} = \varphi_1 x_t[1 - G(x_t)] + \varphi_2 x_t G(x_t) + Q_t w_t \dots (2^*)$$

In (2^*), the aim is to choose $G(s_t)$ (linear or non-linear) that will be the best estimate of x_{t+1}.

Therefore, x_{t+1} is estimated using the STAR model and using the specification and estimation procedure by [9]. Hence, the need for an estimate of x_{t+1} to develop the Kalman filter for the *redefined MSSM.*

E. Predicted Estimate of the State

To achieve this, the conditional expectation of x_{t+1} at time x_t is required i.e $G(x_{t+1};\, \theta)$, where θ is the vector of all the parameters in the model i.e $\theta = (\varphi_1, \varphi_2, \gamma, c)'$

Hence we consider the predictable configuration of the models given by

$$G(x_{t+1};\, \theta) = \varphi_1 x_t[1 - G(x_t)] + \varphi_2 x_t G(x_t) \dots (3^*)$$

$G(x_{t+1};\, \theta)$is the deterministic part of the model and can be referred to as the *SKELETON OF THE MODEL* as in [4]. This skeleton contains useful properties of non-linear time series models and can be obtained from analysing the skeleton as well as its associated difference. This will be used to build the Kalman filter for the MSSM.

Let $x_{t+1} = G(x_t;\, \theta)$ with ε_t assumed to be zero. We then differentiate equation (3^*) to obtain the equilibrium and fixed point in the skeleton, say$x^* = G(x^*;\, \theta)$, since x^* is unknown.

F. Assumptions for Determining Equilibrium Fixed Point Estimate

- The sequence $(say, x_0, x_1, x_2, \dots)$ generated from (3^*) should converge to x^* for values of x_0 close tox^*. This is called the *LOCALLY STABLE EQUILIBRIUM.*
- The sequence generated above converges to x^* for all initial values of x_0. This is termed the *GLOBALLY STABLE EQUILIBRIUM.*

Hence, according to the theory of equilibrium in non-linear systems, non-linear difference equations can have different types of equilibrium points, categorised as

- Single (Stable or Unstable) equilibrium
- Multiple equilibrium
- No equilibrium at all

An equilibrium point must satisfy certain stability criterion to be significant physically. Also equilibrium is said to be stable if close or nearby solutions stay nearby for all future time. However, the equilibrium position cannot be identified exactly but approximately in dynamic systems such as the case in state space modelling. Nonetheless, equilibrium must be stable to be physically meaningful.

For example, Suppose $x^* \epsilon \mathcal{R}^n$ is an equilibrium point for the differential equation.

$$x' = F(x)$$

Then x^* is a stable equilibrium if ∀ neighbourhood $\mathcal{O}$ of x^* in $\mathcal{R}^n$, there is a neighbourhood $\mathcal{O}_1$ of x^* in $\mathcal{O}$ such that all solution x_t with $x(0) = x_0$ in $\mathcal{O}_1$ is defined and remain in $\mathcal{O} \forall t > 0$

Hence, a necessary and sufficient condition for an equilibrium of 3^* to be stable (locally) is given as

$$\left|\frac{\partial G(x^*;\, \theta)}{\partial x}\right| < 1$$

Therefore, for the points x_t close to x^*

$$x_{t+1}^* - x^* = G(x_t;\, \theta) - G(x_t^*;\, \theta) \approx \frac{\partial G(x_t^*;\, \theta)}{\partial x}(x_t - x^*)$$

Hence,

$$|x_{t+1} - x^*| < |x_t - x^*|$$

This implies

$$\frac{\partial G(s_t^*;\, \theta)}{\partial x} < 1$$

This shows that x_{t+1} will be closer to x^* than x_t if $G(x_t^*;\, \theta)$ is a contraction in the neighbourhood of $x = x^*$.
Therefore, we derive the difference equation of (3*)

$$\frac{\partial G(x_t^*;\, \theta)}{\partial x} = \hat{x}_t^*[\gamma(\varphi_2 - \varphi_1)]G(x_t^*;\gamma,\theta)[1 - G(x_t^*;\gamma,\theta)] + \varphi_1[1 - G(x_t^*;\gamma,\theta)] + \varphi_2 G(x_t^*;\gamma,\theta) \ldots (4^*)$$

Therefore,

$$\hat{x}_t^*[\gamma(\varphi_2 - \varphi_1)]G(x_t^*;\gamma,\theta)[1 - G(x_t^*;\gamma,\theta)] + \varphi_1[1 - G(x_t^*;\gamma,\theta)] + \varphi_2 G(x_t^*;\gamma,\theta) = 0$$

$$\hat{x}_t^*[\gamma(\varphi_1 + \varphi_2)]G(x_t^*;\gamma,\theta)[1 - G(x_t^*;\gamma,\theta)] = \varphi_1[1 - G(x_t^*;\gamma,\theta)] + \varphi_2 G(x_t^*;\gamma,\theta)$$

Hence, the estimate of s_t for the equation

$$\hat{x}_t^* = \frac{\varphi_1[1 - G(x_t^*;\gamma,\theta)] + \varphi_2 G(x_t^*;\gamma,\theta)}{[\gamma(\varphi_1 + \varphi_2)]G(x_t^*;\gamma,\theta)[1 - G(x_t^*;\gamma,\theta)]} \ldots (5^*)$$

Given the estimate of the prediction, $\hat{x}_t^*$, in (5*), the measurement equation (1*) is updated and the state equation (2*) is redefined with the new information in y_{t+1}. This gives a filtered or optimal estimate of the state up to the last estimate at time t.
Updating the state equation to get an improved estimate is achieved through the blending of the predicted estimate and the new observation y_{t+1}.
By illustration,

$$\hat{x}_t = (I - K_t H_{t+1})\hat{x}_t^* + K_t v_{t+1}$$

where, $I - K_t H_{t+1}$ is the blending factor of the predicted estimate and $K_t v_{t+1}$ is the blending factor for the new observation y_{t+1}. This can be rewritten as

$$\hat{x}_t = \hat{x}_t^* + K_t(y_{t+1} - H_{t+1}\hat{x}_t^*) \quad \ldots (7^*)$$

The optimal MMSE of x_t is therefore required to optimise the system. This is achieved by finding the optimal value of K_t.

To achieve the optimal value of K_t, the later part of (7*) is considered and this contains the new residual, $y_{t+1} - H_{t+1}\hat{x}_t^*$, which is a measurement noise, however, the aim is to optimize the state. Hence, the error generated in the "a posteriori filtered state vector" estimate, $\hat{x}_t^*$, is used. That is;

$$e_t = trueestimated - predictedestimate = \hat{x}_t - \hat{x}_t^*$$

This gives the ***filtered state vector covariance matrix***

$$P_t = E[e_t e_t'] = E[(\hat{x}_t - \hat{x}_t^*)(\hat{x}_t - \hat{x}_t^*)']$$

where $\hat{x}_t$ takes the form in 7*
Hence,

$$P_t = E\{[\hat{x}_t - \hat{x}_t^* - K_t(y_{t+1} - H_{t+1}\hat{x}_t^*)][\hat{x}_t - \hat{x}_t^* - K_t(y_{t+1} - H_{t+1}\hat{x}_t^*)]'\}$$

and y_{t+1} takes the form in 1*
This implies that,

$$\begin{aligned}P_t &= E\{[(\hat{x}_t - \hat{x}_t^*) - K_t(Hx_{t+1} + v_{t+1} - H_{t+1}\hat{x}_t^*)][Same]'\}\\
&= E\{[(\hat{x}_t - \hat{x}_t^*) - K_t H(\hat{x}_t - \hat{x}_t^*) - K_t v_{t+1}][Same]'\}\\
&= E\{[(I - K_t H_{t+1})(\hat{x}_t - \hat{x}_t^*)(\hat{x}_t - \hat{x}_t^*)'(I - K_t H_{t+1})' - (I - K_t H_{t+1})(\hat{x}_t - \hat{x}_t^*)v_{t+1}' K_t'] - K_t v_{t+1}(\hat{x}_t - \hat{x}_t^*)'(I - K_t H_{t+1})' + K_t v_{t+1} v_{t+1}' K_t'\}\\
&= (I - K_t H_{t+1})E[(\hat{x}_t - \hat{x}_t^*)(\hat{x}_t - \hat{x}_t^*)'](I - K_t H_{t+1})' - (I - K_t H_{t+1})E[(\hat{x}_t - \hat{x}_t^*)v_{t+1}']K_t' - K_t E[v_{t+1}(\hat{x}_t - \hat{x}_t^*)'](I - K_t H_{t+1})' + K_t E[v_{t+1}v_{t+1}']K_t'\}\end{aligned}$$

Where $E[v_{t+1}] = 0$ by definition
Therefore, the ***filtered state vector covariance matrix***,

$$P_t = (I - K_t H_{t+1})P_t^*(I - K_t H_{t+1})' + K_t R_t K_t', \qquad \forall K_t \ldots (8^*)$$

Where, $P_t^* = E[(\hat{x}_t - \hat{x}_t^*)(\hat{x}_t - \hat{x}_t^*)']$ and $R_t = E[v_{t+1}v_{t+1}']$
This shows that the ***filtered state vector covariance matrix***, P_t, is also dependent on K_t and further reveals the importance of the optimal value of K for both $\hat{x}_t$ and P_t.

The minimization of K will require summing the diagonal entries of the state vector covariance matrix (Since the diagonal entries are the variances). This is achieved by the use of some matrices techniques, i.e. taking the derivatives of the trace of P_t, (TrP_t) with respect to (w.r.t.) K_t to get the optimal K_t.
Hence, rewriting (8*),

$$P_t = P_t^* - K_t H_{t+1} P_t^* - P_t^* H_{t+1}' K_t' + K_t(H_{t+1} P_t^* H_{t+1}' + R_t)K_t' \ldots (8_a^*)$$

and differentiating w.r.t K_t,

$$\frac{d(TrP_t)}{dK_t} = \frac{d}{dx}Tr(P_t^*) - \frac{d}{dx}Tr(K_t H_{t+1} P_t^*) - \frac{d}{dx}Tr(P_t^* H_{t+1}' K_t') + \frac{d}{dx}Tr(K_t(H_{t+1} P_t^* H_{t+1}' + R_t)K_t')$$

Hence,

$$K_t = (H_{t+1} P_t^*)(H_{t+1} P_t^* H_{t+1}' + R_t)^{-1} \ldots (9^*)$$

This is the ***Kalman gain*** and can be substituted in 8 to obtain the optimal estimate of the state error covariance matrix.
Substituting (9*) into (8_a^*) gives,

$$P_t = P_t^* - (H_{t+1}P_t^*)'(H_{t+1}P_t^*H_{t+1}' + R_t)^{-1}H_{t+1}P_t^* - P_t^*H_{t+1}'[(H_{t+1}P_t^*)(H_{t+1}P_t^*H_{t+1}' + R_t)^{-1}]' + (H_{t+1}P_t^*)(H_{t+1}P_t^*H_{t+1}' + R_t)^{-1}(H_{t+1}P_t^*H_{t+1}' + R_t)[(H_{t+1}P_t^*)(H_{t+1}P_t^*H_{t+1}' + R_t)^{-1}]'$$

where $(H_{t+1}P_t^*H_{t+1}' + R_t)^{-1}(H_{t+1}P_t^*H_{t+1}' + R_t) = I$

$$P_t = P_t^* - (H_{t+1}P_t^*)'(H_{t+1}P_t^*H_{t+1}' + R_t)^{-1}H_{t+1}P_t^* \ldots (8_b^*)$$

Therefore,

$$P_t = P_t^* - K_tH_{t+1}P_t^* \ldots (8_c^*)$$

Hence, $(8_a^*), (8_b^*), (8_c^*)$ are three different equations for P_t. However, $(8_b^*)\ and (8_c^*)$ are the only valid equations for the optimal K_t. This implies that any of the two equations can be used to optimise the system. Therefore, any of $(8_b^*) and (8_c^*)$ can be used in (9*) to optimize the state equation (2*) and system (1*).

III. Results

In summary, the state update and measurement update are as follows

A. State Update:

$$\hat{x}_t^* = \frac{\varphi_1[1 - G(x_t^*;\gamma,\theta)] + \varphi_2 G(x_t^*;\gamma,\theta)}{[\gamma(\varphi_1 + \varphi_2)]G(x_t^*;\gamma,\theta)[1 - G(x_t^*;\gamma,\theta)]}$$

$$P_t = (I - K_tH_{t+1})P_t^*(I - K_tH_{t+1})' + K_tR_tK_t', \qquad \forall K_t$$

where, $P_t^* = E[(\hat{x}_t - \hat{x}_t^*)(\hat{x}_t - \hat{x}_t^*)'$ and $R_t = E[v_{t+1}v_{t+1}']$

B. Measurement update:

$$\hat{x}_t = \hat{x}_t^* + K_t(y_{t+1} - H_{t+1}\hat{x}_t^*)$$

$$P_t = P_t^* - (H_{t+1}P_t^*)'(H_{t+1}P_t^*H_{t+1}' + R_t)^{-1}H_{t+1}P_t^*$$

where $(H_{t+1}P_t^*)'(H_{t+1}P_t^*H_{t+1}' + R_t) = K_t$

IV. VDiscussion

The Nigerian Consumer Price Index (CPI) between January 1995 - December 2015 and Gross Domestic Products (GDP) data between Q1 1988 and Q4 2013 were collected from the National Bureau of Statistics data portal to test the practicability of the Modified State Space Model (MSSM) as well as its performance against the Classical State Space Model (SSM).

In addition to the real data, simulated data samples of sizes 250, 500, and 1; 000, from a logistics function were also used to verify the superiority of the MSSM against the SSM.

After the evaluation the Akaike Information Criterion (AIC), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) were used as assessment criteria to evaluate the MSSM and SSM.

For the SSM, the "KFAS" package in the R package was used to obtain the results while for the MSSM, the authors developed the code for the model but used the "tsdyn" R package for the Smooth Transition Autoregressive (STAR) model.

The Lagrange Multiplier (LM) test statistic was used to validate the nonlinearity of the data sets at $p < 0.05$ before estimating the Modified State Space Model (MSSM). The Predicted State (PS), Blending Equation (BE), Kalman Gain (KG), and Filtered State Covariance (FSC) which are key attributes of the MSSM were derived using the STAR model to tune the Kalman Filter.

The LM test result revealed that the CPI data was linear and the GDP data was nonlinear. The assessment tests for the CPI, GDP and the simulated samples of 250, 500 and 1000 are presented in Table 1 below:

TABLE 1: SUMMARY OF DIAGONOSTIC RESULTS

	Table Column Head		
	Measure	*SSM*	*MSSM*
CPI	AIC RMSE MAPE	25.84 3.65 0.60	23.02 2.13 0.43
GDP	AIC RMSE MAPE	23.25 13.84 33.29	11.44 7.67 23.17
n = 250	AIC RMSE MAPE	2546.6 127.4 39.3	1941.6 66.8 11.7
n = 500	AIC RMSE MAPE	5161.7 80.5 41.2	3957.7 -14.9 12.6
n = 1000	AIC RMSE MAPE	10289.60 44.7 40.9	6550.30 21.2 6.4

a. This results were generated using the R Package for the two models

V. Conclusion

After implementation of MSSM and CSSM using the collected samples, the model diagnostics results for the MSSM are less than the CSSM results. It was noted that the MSSM is efficient in the estimation of both linear and nonlinear systems. Hence, the MSSM is therefore preferred in the estimation and inference for phenomena exhibiting nonlinear relationships.

References

[1] Durbin J. and Koopman S. J., Time Series Analysis by State Space Methods, 1st Edition, Oxford University Press, Oxford, 2001.

[2] Durbin J. and Koopman S. J., Time Series Analysis by State Space Methods, 2nd Edition, Oxford University Press, Oxford, 2012.

[3] Harvey A. C., Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge: Cambridge University Press, 1989.

[4] Chan K. S. and Tong H., "On estimating thresholds in autoregressive models, Journal of Time Series Analysis", 7, 1986, 179-194.

[5] Hamilton J. D, "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", Econometrica 57, 1989, 357-384.

[6] Hansen B. E., "Threshold Autoregression in Economics (PDF), Statistics and Its Interface", 2011, 4: 123127.

[7] Kalman R. E., "A New Approach to Linear Filtering and Prediction Problems, Journal of Basic Engineering", 1960, 82: 35.

[8] Kim C. J., and Nelson C. R, "State-Space Models with Regime-Switching: Classical and Gibbs-Sampling Approaches with Applications", MIT Press, 1999.

[9] Terasvirta T., "Specification, estimation, and evaluation of smooth transition autoregressive models", Journal of the American Statistical Association, 1994, 89, 208218.