Preamble

- As the world moves into more sophisticated time periods, Predicting accurately into the future is the ultimate GOAL

- “At the heart of statistics lies regression”.

- Quantile regressions provides a more complete picture of the conditional distribution(Koenker and Bassett,1978)

Detection of structural changes due to misspecification in a conditional quantile polynomial distributed Lag (QPDL) model using change-point analysis

BY

KWADWO AGYEI NYANTAKYI

Ghana Institute of Management and Public Administration (GIMPA)

PRESENTATED AT SDSS-2019
VENUE: HYATT REGENCY HOTEL
BELLEVUE WASHINGTON STATE

Background

- In recent times, issues relating to risk management and bank performance have dominated the business world with its repercussion felt at regulatory level.

- Loss of large sums of money is some of the issues affecting the global banking and investors.

- The need to assist decision-making at the management, regulatory and governmental levels in making the exact predictions of the occurrence of these events in the future.

Background

- Likewise, this volatility phenomenon occurs in our economic and financial activities causing extreme losses, hence the need to study these fluctuations as they move across the boundaries into the other levels.

- Ordinary least squares regression estimates the mean response as a function of the regressors or predictors, and least absolute deviation regression estimates the conditional median function, which has been shown to be more robust to outliers.

- Although regression analysis has been the mode of estimation, there is now a paradigm shift towards quantile estimation.

Background

- Ordinary least squares regression is not effective in capturing the extreme values or the adverse losses evident in return distributions, these are captured by quantile regressions, which provides a more comprehensive picture of the conditional distribution(Koenker and Bassett, 1978)

- Hence the need to study the transition of one data set into the boundaries of the other which is causing the losses.

- Mathematically, quantiles can be defined as the points taken at regular intervals from the cumulative distribution function (CDF) of a random variable

- Koenker and Bassett (1978), introduced quantile regression to estimate the conditional quantile function of the response, by using the idea of generalized least absolute deviation regression. As a result, they provide much more information about the conditional distribution of a response variable.

Objectives of the Study

Main Objective

- To develop a method for the detection of change points in a parametric and nonparametric setting in time series using conditional quantiles

Specific Objectives

- To develop a method for the estimation of polynomial distributed lags using quantiles

- To detect structural changes in the developed model using change points analysis.

Introduction detection of structural changes

- The question of structural stability of models is very important in predicting the future in various diverse areas of science such as economics, finance, physics, geology, medicine as well as in quality control and agriculture.

- Questions like did a change occur? Did more than one change occur? When did these changes occur? Can be answered by performing a change-point analysis, (Taylor, W. 2000a, 2000b).

- In order to study the structural changes in models came the evolution of change-points which was introduced by the context of quality control by Csorgo and Horvath, (1997)

- Change–points has been employed in finding the possible changes in otherwise independent identically distributed random variables and has been extended to the stability tests of parameters of the regression functions, (Andrews, D. W. K. 1993, Auger,1989).

Methodology Quantiles

Given the multiple regression

$$Y_t = \beta_0 + \beta_1 X_1 + - - - + \beta_n X_n + \varepsilon_t \text{ ....} (1)$$

With ε_t innovations identically distributed with mean zero and variance one.

Given a real valued random variable Y with a distribution function $F_Y(y) = P(Y \leq y)$ by

Koenker (1982) and Zhao, et al., (2008) the τ^{th} quantile of Y is given by

$$Q_Y(\tau) = F_Y^{-1}(\tau) = \inf\{y =: F_Y(y) \geq \tau\} \text{ where } \tau \epsilon [0,1].$$

Thus the conditional quantile $Q(\tau|X)$ are the inverse of the conditional distribution function of the response variables.

Methodology QPDL

The polynomial distributed lag by Almon, (1965) can be written as

$$y_t = \varphi + \sum_{i=0}^{k} \beta_i X_{t-i} + \varepsilon_t \qquad \ldots\ldots(2)$$

With k *number of lags* , and the $\beta_i{}'s$ can be approximated by suitable polynomials. That is

$$\beta_i = a_0 + a_1 i + a_2 i^2 + a_3 i^3 + - - - + a_n i^n$$

Thus for the n^th degree polynomial with k *number of lags* we have

$$y_t = \varphi + \sum_{i=0}^{k}(a_0 + a_1 i + a_2 i^2 + a_3 i^3 + - - - + a_n i^n)\, X_{t-i} + \varepsilon_t$$

$$y_t = \varphi + \sum_{i=0}^{k} a_0 X_{t-i} + \sum_{i=0}^{k} a_1 (i) X_{t-i} + \sum_{i=0}^{k} a_2 (i^2) X_{t-i} + \sum_{i=0}^{k} a_2 (i^2) X_{t-i} +$$

$$\sum_{i=0}^{k} a_3 (i^3) X_{t-i} + - - - + \sum_{i=0}^{k} a_n (i^n) X_{t-i} + \varepsilon_t \qquad \text{--- (3)}$$

We defined the new model QPDL

Conditional Quantile Polynomial Distributed Lag Model

$$y_t = \varphi(\psi_t) + \sum_{i=0}^{k} a_0 \psi_t X_{t-i} + \sum_{i=0}^{k} a_1 \psi_t (i) X_{t-i} + \sum_{i=0}^{k} a_2 \psi_t (i^2) X_{t-i} + \sum_{i=0}^{k} a_2 \psi_t (i^2) X_{t-i} + \sum_{i=0}^{k} a_3 \psi_t (i^3) X_{t-i} + - - - + \sum_{i=0}^{k} a_n \psi_t (i^n) X_{t-i} + \varepsilon_t \quad \ldots\ldots \quad (4)$$

Assuming $\{\varepsilon_t, 1 \leq t \leq n\}$ are independent identically distributed random errors. We can write the lag model as:

$$y_t = \varphi(\psi_t) + a_0 \psi_t \sum_{i=0}^{k} X_{t-i} + a_1 \psi_t \sum_{i=0}^{k} (i) X_{t-i} + a_2 \psi_t \sum_{i=0}^{k} (i^2) X_{t-i} + a_2 \psi_t \sum_{i=0}^{k} (i^2) X_{t-i} + a_3 \psi_t \sum_{i=0}^{k} (i^3) X_{t-i} + - - - + a_n \psi_t \sum_{i=0}^{k} (i^n) X_{t-i} + \varepsilon_t \quad \ldots\ldots(5)$$

Then the QPDL can be simplified as:

$$y_t = \varphi(\psi_t) + a_0(\psi_t)Z_{0t} + a_1(\psi_t)Z_{1t} + a_2(\psi_t)Z_{2t} + a_3(\psi_t)Z_{3t} + -- + a_n(\psi_t)Z_{nt} + \varepsilon_t$$

(6)

The conditional $\alpha - quantile$ function (QPDL) can be written as

$$Q_{y_t}(\alpha|Z_{0t}, - - -, Z_{nt}) = \varphi(\alpha) + a_0(\alpha)Z_{0t} + a_1(\alpha)Z_{1t} + a_2(\alpha)Z_{2t} + a_3(\alpha)Z_{3t} + -- - + a_n(\alpha)Z_{nt} \quad \ldots\ldots\ldots\ldots\ldots\ldots(7)$$

This can simply be written as

$$Q_{y_t}(\alpha|\xi) = Z_{it}^T a(\alpha)$$

Where

$$z_{it}^T = (1, Z_{0t}, - - -, Z_{nt})^T$$

Considering a second degree polynomial, we have

$$Q_{y_t}(\alpha|Z_{it}) = \varphi(\alpha) + a_0(\alpha)Z_{0t} + a_1(\alpha)Z_{1t} + a_2(\alpha)Z_{2t} \quad \ldots(8)$$

The conditional cumulative probabilities of (Y_t). This is given by

$$Pr(Y_t \leq q(Z_{it})|Z_{it} = z) = \alpha.$$

We solve the minimization problem

$$E(|Y_t - q(Z_{it})|_\alpha |Z_{it} = z) = \min_{f \in L'(u)} E(|Y_t - f(Z_{it})|_\alpha |Z_{it} = z)$$

For a QPDL (2) we have

$$Y_t = \varphi(\alpha) + a_0(\alpha)Z_{0t} + a_1(\alpha)Z_{1t} + a_2(\alpha)Z_{2t} + \varepsilon_t. \text{ -- (9)}$$

Therefore solving the minimization problem for the estimates for φ, a_0, $a_1$and $a_2$we have

$$S(\varphi, a_0, a_1 a_2) = \sum_{t|y_t \geq aZ_{it}}^{k} \alpha(|Y_t - \varphi + a_0 Z_{0t} + a_1 Z_{1t} + a_2 Z_{2t}|) +$$

$$\sum_{t|y_t < aZ_{it}}^{k} (1-\alpha)(|Y_t - \varphi + a_0 Z_{0t} + a_1 Z_{1t} + a_2 Z_{2t}|)$$

$$= \sum_{t|y_t \geq aZ_{it}}^{k} \alpha\left(\left|Y_t - Z_{it}^T \omega\right|\right) + \sum_{t|y_t < aZ_{it}}^{k} (1-\alpha)\left(\left|Y_t - Z_{it}^T \omega\right|\right) \text{ ……(10)}$$

Where $\omega = \left(\varphi(\alpha), a_0(\alpha), a_1(\alpha), a_2(\alpha)\right)$ denotes the coefficients

Detection of structural changes methodology

- We propose the use of both Binary Segmentation (BinSeg) and Cumulative Sum (cusum) methods for detecting the structural changes in the QPDL model because they are capable of detecting multiple changes (Mueller, 1992),

- The advantage of the BinSeg method is that it is computationally efficient, (Eckley et al., 2011).

- Both methods are powerful tools that better characterize the changes, control the overall error rate, robust to outliers, more flexible and simple to use, (Efron et al., 1993).

Testing for Structural Changes

From the proposed model, we assume two (2) conditional 2nd degree polynomial $\alpha - quantile$ regimes

$$\boldsymbol{Y_{1t} = \varphi(\alpha) + a_{1,0}(\alpha)Z_{1,0t} + a_{1,1}(\alpha)Z_{1,1t} + a_{1,2}(\alpha)Z_{1,2t} + \varepsilon_{1t}, \quad t = 1, 2, \dots, T_1} \quad \dots(1)$$

$$\boldsymbol{Y_{2t} = \varphi(\alpha) + a_{2,0}(\alpha)Z_{2,0t} + a_{2,1}(\alpha)Z_{2,1t} + a_{2,2}(\alpha)Z_{2,2t}, +\varepsilon_{2t}, \quad t = 1, 2, \dots, T_2} \quad \dots(2)$$

With $\{\varepsilon_{1t}\}$ and $\{\varepsilon_{1t}\}$ i.i.d unobservable innovations with

$$E(\varepsilon_{1t}) = 0 \; and \; E(\varepsilon_{2t}) = 0 \;, V(\varepsilon_{1t}) = \sigma_1^2 \; and \; V(\varepsilon_{2t}) = \sigma_2^2$$

We further assume that $E(\varepsilon_{1t}) \; and \; E(\varepsilon_{2t})$ are i.i.d innovations which are independent of each other.

Let $\hat{a}_{1,0}(\alpha)$ and $\hat{a}_{2,0}(\alpha)$ be the estimates of $a_{1,0}(\alpha) \; and \; a_{2,0}(\alpha)$ respectively. Then we also define

$$\boldsymbol{Y_{1t}^* = Y_{1t} - \overline{Y}_1 \; and \; Y_{2t}^* = Y_{2t} - \overline{Y}_1}$$

Testing Procedure for the Detection of the Change-Point

From the proposed QPDL model, let us now consider a conditional 2^{nd} degree polynomial $\alpha - quantile$ model with a change after an unknown time point

$$\mathbf{1 \leq k^* = k^*(n) \leq n}$$

$$\boldsymbol{X_{it} = \begin{cases} Z_{it} & t \leq k^* \\ Y_{it-k^*} & t > k^* \end{cases}}$$

Where $\{Y_{it}\}$ is some QPDL which differs distributionally from $\{Z_{it}\}$. The unknown parameter k^* is called the change-point.

We are now interested in the testing problem

$$\boldsymbol{H_0 : k^* = n \text{ vs. } H_1 : k^* < n}$$

Our testing procedures are based on various functionals of the partial sums of estimated residuals with respect to the model (6.1)

$$\hat{S}n(k) = \sum_{t=p+1}^{k} \hat{\varepsilon}_t$$

$$= \boldsymbol{\sum_{t=p+1}^{k} \left(X_{it} - f(Z_{it}, \widehat{\beta}_n)\right).} \qquad \text{...(6)}$$

where $\hat{\beta}_n$ is the least-squares estimator of β_0 (assuming the null hypothesis holds true).

Precisely we minimize the nonlinear least squares (NLLS) with respect to β.

$$Q_n(\beta) = \sum_{t=p+1}^{k} \left(X_t - f(Z_{it}, \hat{\beta}_n)\right)^2$$

$$= \sum_{t=p+1}^{k} q_n(\beta) \qquad \text{...(7)}$$

Testing Procedure for the Detection of the Change-Point

The modified test statistics are of the form

$$T_{n1} = \max_{p<k<n} \left(\sqrt{\left(\frac{n-p}{k(n-p-k)}\right)} \left|\hat{S}_n(k)\right| \right), \quad \text{.......(6)}$$

$$T_{n2}(q) = \max_{p<k<n} \left(\frac{1}{\sqrt{n-p}\, q\left(\frac{k}{n-p}\right)} \left|\hat{S}_n(k)\right| \right), \quad \text{.......(7)}$$

$$T_{n3}(G) = \max_{p+G<k<n} \frac{1}{\sqrt{G}} \left(\left|\hat{S}_n(k) - \hat{S}_n(k-G)\right| \right), \quad \text{.......(8)}$$

$$T_{n1} = \max_{p+G<k<n-p-G} \frac{1}{\sqrt{2G}} \left(\left|\hat{S}_n(k+G) - 2\hat{S}_n(k) + \hat{S}_n(k-G)\right| \right), \quad \text{.......(9)}$$

$$T_{n4}(r) = \frac{1}{n-p} \sum_{k=p+1}^{n-1} \frac{1}{r(k/(n-p))} \left(\frac{1}{\sqrt{n-p}} \hat{S}_n(k) \right)^2, \quad \text{.......(10)}$$

where q(·) and r(·) are weight functions defined on (0,1) specified below and G < n. In this case, we obtain a consistent change-point estimator which is related to the test statistics. (Stockis, J. P. et al., 2010; Tadjuidje K., J. et al., 2011).

Application to QPDL Using Change-Point Analysis

- Using rubber data, we apply both Binary Segmentation and Cumulative Sum methods to investigate the structural changes in the QPDL model.

Results and Discussion

Detection of structural Changes for the correctly specified QPDL model

Table 1: Conditional quantile polynomial distributed lag (QPDL), tau=0.25 for cusum and BinSeg methods. (SIC (Penalty) = 7.742402)

tau	Change type	Positions of the change						Max no. of change
0.25	Cusum	16.00	41.00	44.00	46.00	24.00		5
	BinSeg	16	19	23	42	44	48	5
	Mean	1572.5	1000.9	1539.4	1028.4	1432.9	1829.5	
	Variance	52268.88	42726.38	36481.89	22493.61	570.39	30727.15	

- Both the cumulative sum (Cusum) and the Binary Segmentation (BinSeg) methods detected maximum of 5 structural changes.
- **The minimum change for both methods occurred at the 16th position.**
- The highest maximum for Cusum is 46th and that of BinSeg is 48th .

There are changes in mean production with minimum of 1000.9, occurring at 19th position and the maximum mean of 1829.5 with a change at 48th position.

Graphical interpretation

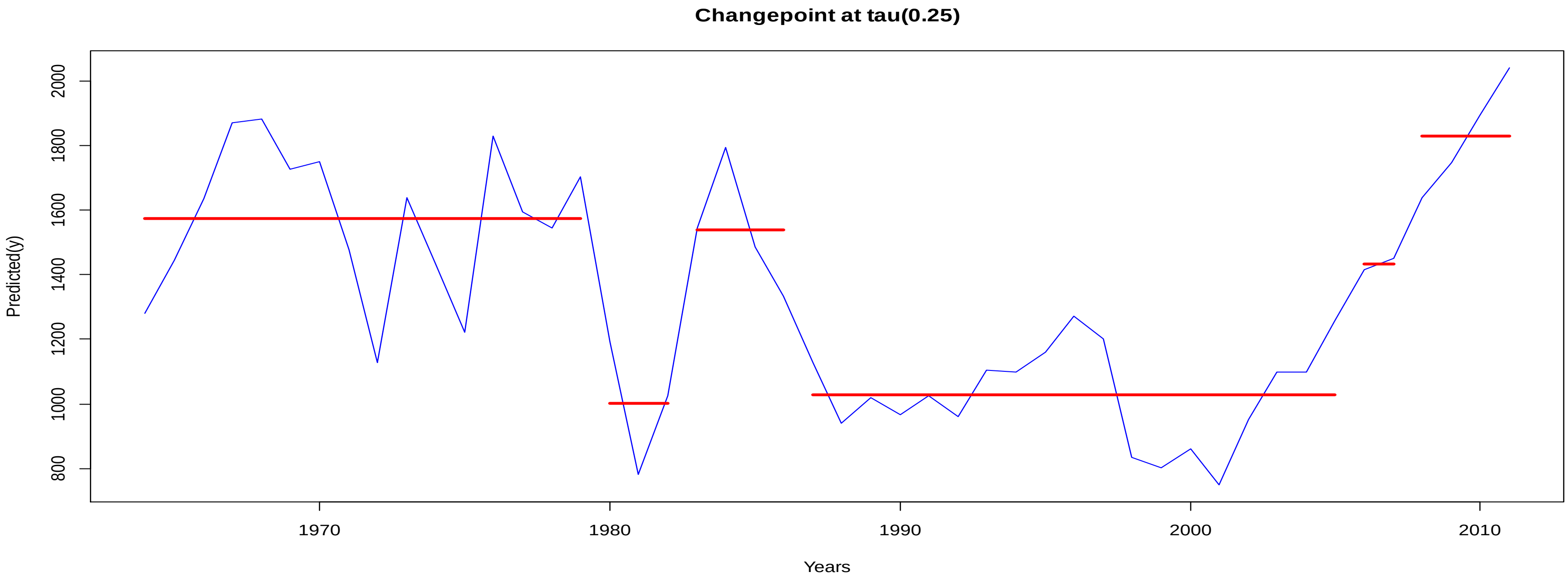

Figure 1: Plot showing the changes in production of Predicted $\hat{y}$ against Years of tau=0.25

- **The 1st change occurred in 1979**,
- The 2nd change in production occurred in 1982 with a high increase in average production again.
- The 3rd change shows another drop in average production which lasted from 1977 to 2005 and then the 4th change in average production occurred.

Table 2: Conditional quantile polynomial distributed lag (QPDL) tau=0.75 cusum and BinSeg methods. (SIC (Penalty) = 7.742402)

tau	Change type	Positions of the change					Max no. of change
0.75	Cusum	17.00	42.00	45.00	24.00	19.00	5
	BinSeg	17	23	42	44	48	4
	Mean	2488.207	1987.198	1362.633	2166.128	2965.220	
	Variance	164495.50	533186.09	86739.77	18553.91	146672.70	

- an optimum of 5 changes for Cusum and 4 for BinSeg.
- The minimum average production 1362.6 occurred at the 42nd position with a variance of 86739.8
- the maximum average production 2965.2 occurring at the 48th position with variance of 146672.7.

Graphical interpretation

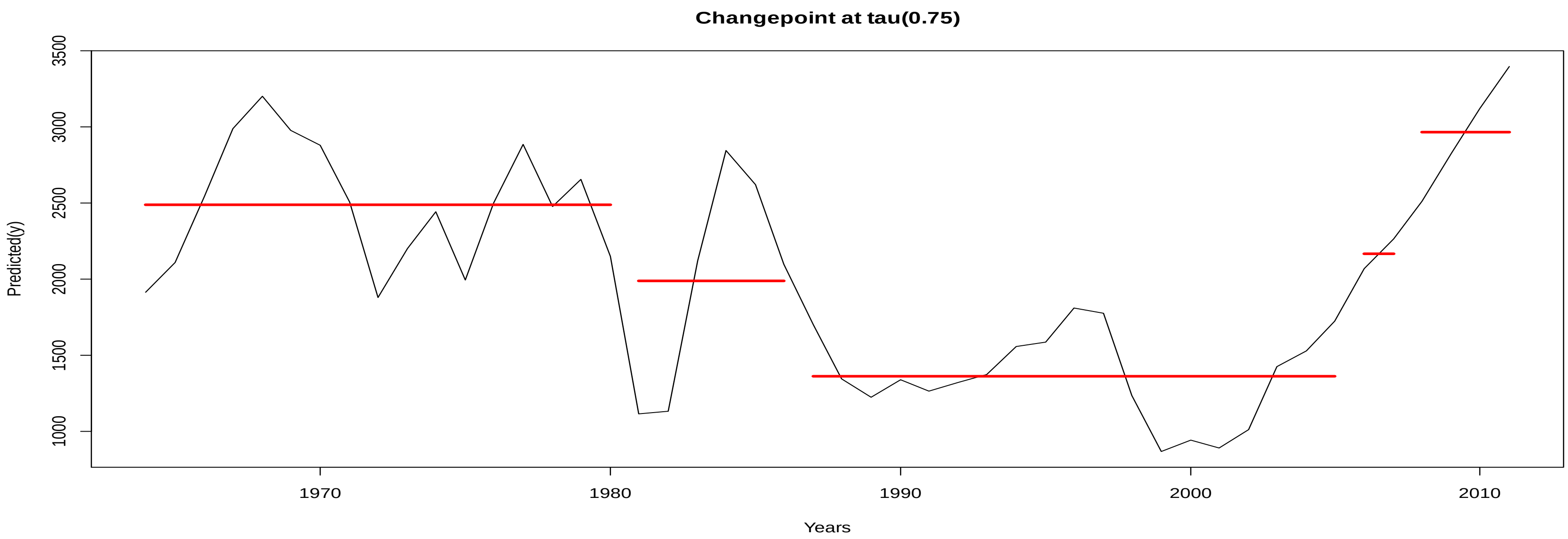

Figure 2: Plot showing the changes in production of Predicted $\hat{y}$ against Years of tau=0.75

- **The 1st change occurred in 1981**, with a drop in average production between 1981 and 1986.
- The 2nd change in average production occurred in 1986 and 2005
- The 3rd change shows another an average increase in production which lasted from 2005 to 2007

Detection of structural Changes for the misspecified QPDL model

Table 3: Misspecified Conditional quantile polynomial distributed lag (QPDL), tau=0.25 for cusum and BinSeg methods. SIC (Penalty)= 7.700295.

tau	Change type	Positions of the change						Max no. of change
0.25	Cusum	23.000	41.000	45.000	43.000	15.000		5
	BinSeg	23	31	33	39	41	47	5
	Mean	1494.72	974.70	1109.92	828.48	956.63	1463.46	
	Variance	51195.76	7094.85	233.45	2160.53	905.68	49666.63	

- There is a shift in position of the **1st change from the 16th position from the correctly specified model to 23rd position for the misspecified model** for both methods.
- The change in mean production was 956.6, occurring at 41st position with variance of 905.7.
- The maximum average production of 1494.7 with a variance of 51195.8 occurred at 23rd position.

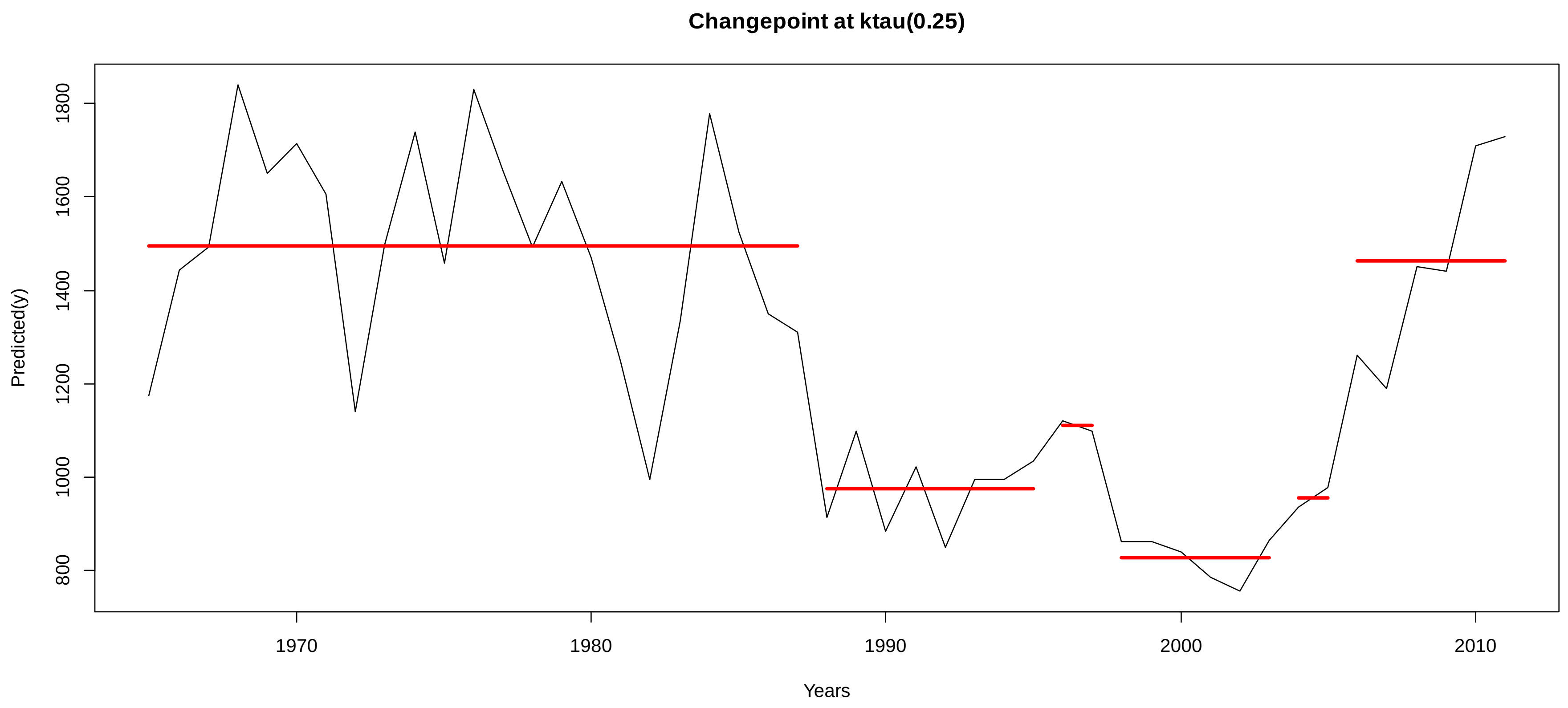

Figure 3: Plot showing the changes in production of Predicted $\hat{y}$ against Years of tau=0.25.

- A shift of the change in the average production which **occurred in 1979 for the correctly specified QPDL model to 1988 in the misspecified QPDL model.**
- There was high drop in the mean production between 1988 and 1996
- There was a drop again between 1998 and 2003.

Table 4: Conditional quantile polynomial distributed lag (QPDL) misspecified tau=0.75 for cusum and BinSeg methods. SIC (Penalty)= 7.700295.

tau	Change type	Positions of the change						Max no. of change
0.75	Cusum	21.000	41.000	45.000	43.000	42.000		5
	BinSeg	15	21	23	41	43	47	5
	Mean	2372.246	2048.590	1891.935	1471.076	2141.091	2594.873	
	Variance	200317.63	364103.87	189.03	56549.01	40653.84	44469.31	

- Similarly, there is a **shift in position of the 1st change from the 16th position from the correctly specified model to 21st position for Cusum and 15th position for BinSeg for the misspecified model.**
- The change in minimum mean production was 1471.1, occurring at 41st position with variance of 56549.
- The maximum average production of 2594.9 with a variance of 44469.3 occurred at 47th position.

is Figure 4, showing the various positions and the time period for which each change occurred.

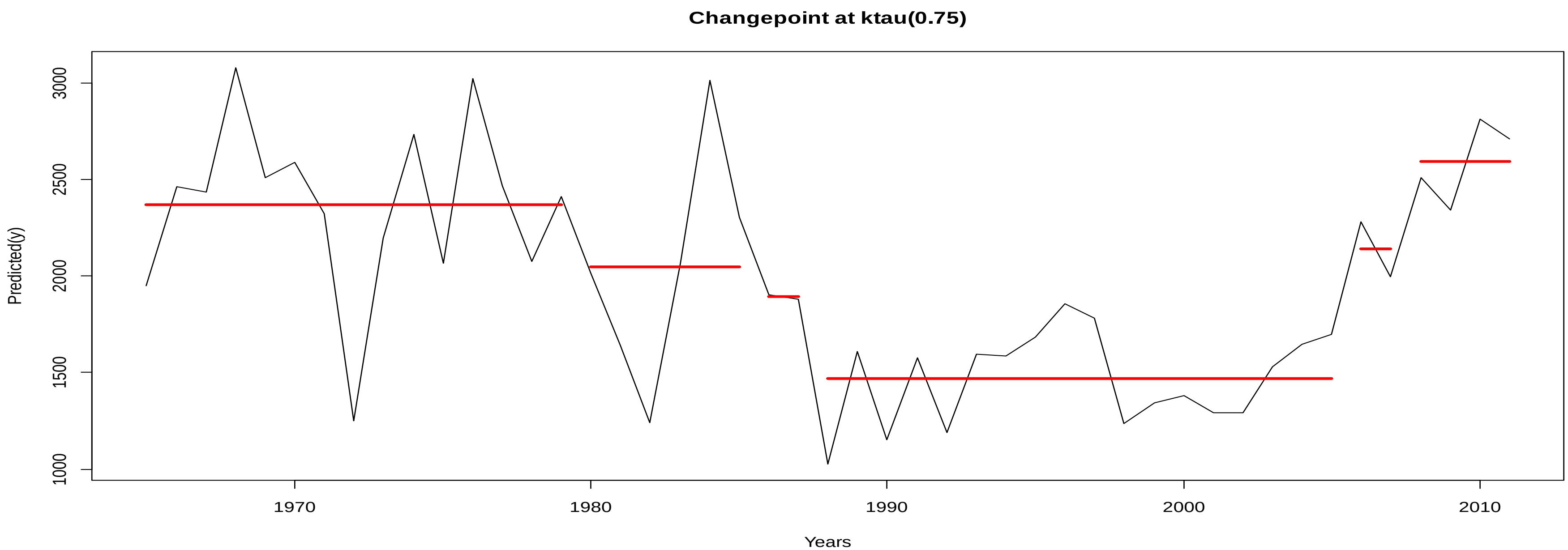

Figure 4: Plot showing the changes in production of Predicted $\hat{y}$ against Years of tau=0.75.

- Similarly, there is a shift of the change in the average production which **occurred in 1981 for the correctly specified QPDL model to 1979 in the misspecified QPDL model** which lasted till 1985
- There was a further drop in average production 1987.
- There was high drop in the mean production between 1988 and 2005, and then an increase occurred for a short period in 2006.
- There was a further high increase in production from 2007 till 2011.

Conclusions

- We observe that both the Cusum and the BinSeg methods detected the structural changes for both the correctly specified and the misspecified QPDL model.

- The Cusum method gives the exact positions where the structural changes occurred and the BinSeg gives the approximated positions where the changes occurred.

- Both methods were able to detect the shift in time for both the mean and variance for the missspecified QPDL model, hence both methods were better for predicting structural stability in a QPDL models.

- The impact of this is that, when there are changes made to a data knowingly or unknowingly, they can be detected, as well as when these changes were effected.

- We further observed that both methods were powerful tools that better characterizes the changes, controls the overall error rate, robust to outliers, more flexible and simple to use.

THANK YOU

*Final Conclusions

- Secondly, we applied neural network to the QPDL model. The results showed that neural network was useful alternative for estimating the QPDL model. We observed very high R-square values with very low AIC values across the quantiles. We observe the effects of the quantile on the dependent variables through the various plots.

- Thirdly, we applied the change-point algorithm using Binary Segmentation and Cumulative Sum methods to the QPDL model. The Cusum method gave the exact positions where the structural changes occurred and the BinSeg gave the approximated positions where the changes occurred.

- Both methods were able to detect the shift in time for both the mean and variance for the missspecified QPDL model, hence both methods were better for predicting structural stability in a QPDL models.

- The impact of this is that, when there are changes made to a data knowingly or unknowingly, they can be detected, as well as when these changes were affected.

References

- Stockis, J.-P., Franke, J., and Tadjuidje K., J. On geometric ergodicity of charme models. Journal of Time Series Analysis, 31:141–152, 2010.
- Wu, Y. Inference for change point and post change mean after a CUSUM test. Springer, New York, 2004.
- M¨uller, H.-G. (1992) Change points in nonparametric regression analysis. Annals of Statistics, 20:737–761.
- Andrews, D. W. K. (1993) Test for parametric instability and structural change with unknown change point. Journal of theoretical probability, 61:821–856.
- Horvath, L., Huskova, M., Kokoszka, P., and Steinebach, J.(2004) Monitoring changes in linear models. J. Statist. Plann. Inference, 126:225 - 251.

Reference

- Kirch, C. and Tadjuidje K., J.(2011) Testing for parameter stability in nonlinear autoregressive models. Preprint,
- Csorgo, M., Horvath., L. (1997) Limit Theorems in Change-Point Analysis, Wiley E. S. Page (1954) Continuous Inspection Schemes, Biometrika 41(1/2), 100–115
- Auger, I. E. And Lawrence, C. E. (1989) Algorithms for the Optimal Identification of Segment Neighborhoods, Bulletin of Mathematical Biology 51(1), 39–54
- Efron, Bradley and Tibshirani, Robert (1993), An introduction to the Bootstrap, Chapman & Hall, New York.
- Hinkley, D. V. (1971), "Inference about the change-point from cumulative sum tests," Biometrika, 58 3, 509-523.