Comparison between the Local Polynomial Kernel Method and cubic spline to Estimating Time-Varying Coefficients Model

Noor Abdul-Kareem Fayadh[1]
College of Administration and Economic, Wasit University, Iraq
nab393820@gmail.com

Wafaa Jaafar Hussain[2]
College of Administration and Economic, Wasit University, Iraq
wjaffer@uowasit.edu.iq

Abstract—The research deals with the study of the varying coefficient model as well as the Time-Varying Coefficient Model because of their remarkable interest in recent years. Methods Two methods were used, local polynomial kernal method and the cubic spline method. The simulation method was used for comparison. Compared to three mathematical functions that represented the variable coefficients, different sample sizes, and different levels of the model's standard deviation, the cubic spline method was best.

Keywords— Smoothing, Varying Coefficient Model, Variable Selection, Smoothing Parameter, Cubic Spline

1. Introduction

Generalized linear regression models are parametric models that are based on analyzing the relationship between variables on several assumptions, and among these assumptions are those that state that the relationship between variables is linear. Assumptions, which make those linear models unrealistic from an applied point of view, and to obtain more realistic models, several parametric models were proposed that aim to make these assumptions more flexible, taking into account the possibility of applying those models to avoid placing restrictions when modeling the relationship between variables. Recently, researchers have relied on nonparametric modeling, which does not impose any restrictions when formulating the model, except that the nonparametric models may fail to include some of the previous information that is available in the sample, which leads to a large variance estimator of the unknown regression function. The other problem that nonparametric methods suffer from is the dimensionality, which makes the nonparametric methods weak in practice when there is more than one explanatory variable. There is a special case of variable coefficient models characterized by the fact that the variable coefficients of the model are a function of the time variable t, and they are known as time-varying coefficient models. Thus, the research will include varying Coefficient model and time-varying Coefficient model and two methods for estimating time-varying Coefficient model, namely the local polynomial kernal method and the cubic spline method

Some of those who used this model can be mentioned Such as (Fan 2008) [3]

(Huang 2004)[6] , (Hastie 1993) [5] , (Cai 1999) [1].

2. Theoretical Aspect

(1.2) Varying Coefficient Model [1] [3] [5] [6] [7] [10]

The formula of generalized additive models is shown, but rather the following formula, which assumes that the response is a linear synthetic for a set of unknown functions.

$$y_i = \sum_{j=1}^{p} \beta_j(x_i) + e_i \quad , i = 1,2,\dots,n \qquad (1)$$

Since y is the response variable and $\beta_j(x_i)$are undefined functions in terms of explanatory variables, and e_i represents the random error variable with mean zero and variance σ^2

There are cases in which the coefficients of the explanatory variables are not fixed but rather a function that depends on another variable, which makes them have variable values The most appropriate model to describe these relationships is known as the time-varying coefficients model, and its mathematical formula is written as follows :

$$Y_i = \sum_{j=1}^{p} \beta_j(U_i)X_{ij} + \varepsilon_i \qquad , i = 1,2,\cdots n \qquad (2)$$

Since y, i = 1,2,..., n is the response variable , and X_{ij} are the explanatory variables, and $\beta_j(U_i)$ They represent the coefficients of the explanatory variables which are a function of the variable U, which is a variable ε_i It is a random error with a mean of zero and variance σ^2 Where the basic idea of these models is to allow the regression coefficients to change easily (interaction) with another variable, while the model shown in formula (1) in which the coefficients of the explanatory variables are fixed there is a special case of the varying coefficient model, which is characterized by the fact that the model parameters are a function of the time variable t they are known as the time-varying Coefficient Model and are used especially in longitudinal studies by allowing to explore the time-varying effect of the explanatory variables on the response Y_i , as follows

$$y_i = \sum_{j=1}^{p} \beta_j(t) X_{ij} + \varepsilon_i \text{ , } i = 1,2,\cdots n \qquad (3)$$

The model described above is called the dynamic generalized linear model. It is a semi-parametric model, which means that it avoids the problem of high dimensions that happens when there are more than one variable that explains something.There are several methods that can be followed in estimating the time-varying coefficient model shown in formula (3), and the following is a detailed description of those methods

2.2 cubic spline [1] [4] [9]

In the case of estimating time-varying parameters in a Time-Varying Coefficient Model. the sum of the remaining squares will be calculated plus a limit called the "roughness penalty". As shown in the following formula :

$$\sum_{i=1}^{n} \left[y_i - \sum_{j=1}^{p} X_{ij}\, \beta_j(t_i) \right]^2 + \lambda \int_a^b \left\{ \beta_j^{(2)}(t_i) \right\}^2 dt \quad i = 1,2,\dots,n \text{ , } j = 1,2,\dots,p \qquad (4)$$

λ: Smoothing Parameter

$\beta_j(t)$: the unknown function to be estimated.

Here is where we reduce the equation(4) to get the estimator value $\hat{\beta}_j(t_i)$ where the first part of the equation is the sum of the squares of the residuals (RSS). As for the second penalty from the above equation, it represents the roughness penalty. which contains the smoothing parameter (λ) and has an important role in determining the value of the unpaved roughness penalty, i.e. when ($\lambda \rightarrow 0$) this means that the roughness penalty does not appear in the equation, and when ($\lambda \rightarrow \infty$) this means that the smoothing parameter is very large, and this produces a stable curve for linear regression.

This parameter also plays an important role in balancing bias and variance $[a, b]$ Suppose that the design time points that are used as knots are ($T_1, T_2, \dots, T_P$) The function β is a cubic spline if the following conditions are present:

- In each penalty period $[T_j, T_{j+1}]$ the function is a polynomial.
- The first derivative, the second derivative, and the function (β) are all continuous functions.

Let β be the parameter vector $\beta = [\beta_1, \beta_2, \dots, \beta_p]^T$ is an ($p\times1$) It is characteristic of the cubist spline.

$\beta^{(2)}(b) = 0$ and $\beta^{(2)}(a) = 0$ knowing that:

$\beta^{(2)}(a)$: The second derivative at the term a

$\beta^{(2)}(b)$: The second derivative at the term b

Then the vector S is the vector of the second derivative of the vector β where:

$S=\left[S_2, \ \dots \ , S_{p-1}\right]^T$

Of degree ((p-2)×1) we assume that $\beta_j = \beta_j(t_i)$ and $S_j = \beta_j^{(2)}(t_i)$, $j = 1,2,\dots,p \quad i = 1,2,\dots,n$

Since the shape of the curve β depends on the above two vectors (S_j, β_j) by two matrices we can define Vectors Let these be matrices (D,E)
Let a vector be the difference between T_{j-1} and T_j my agencies

$a_{j=T_{j+1}-T_j} \quad j=1,2,\dots,p$

Suppose E is an array with a degree of p×(p-2)) its elements e_{ij} and it is calculated as :

$e_{(j-1,j)} = a^{-1}(j-1)$

$e_{(j,j)} = a^{-1}(j-1) - a_j^{-1}$

$e_{(j+1,j)} = a_j^{-1}$

$e_{(i,j)} = 0 \ \forall \quad |i-j| \ \geq 2 \ \ i=1,2,\dots,n \ \ j=2,\dots,p-1$

We also suppose that (D) symmetric matrix Of degree (p-2)× (p-2) its elements are calculated as follows:

$d_{jj} = (a_{j-1} + a_j)/3 \quad , \quad j = \ 2,3 \dots, p-1$

$d_{(j,j+1)} = d_{j+1,j} = \frac{a_j}{6} \quad , \quad j = 2,3,\dots,p-2$

$d_{ij} = 0 \ \forall \ \ |i-j| \geq 2 \ \ , \ \ i=1,2,\dots,n \ \ j=2,\dots,p-1$

And since (D^{-1}) is known, we can define the matrix L as follows :

$$L = E \ D^{-1}E^T \qquad (5)$$

According to the opinion of researchers (Green and Silverman) , if the following condition is met:
E^T β=DS The vectors B, S are spline cubic. When the above condition is satisfied, the square of the second derivative of the β function can be calculated as follows:

$$\int_a^b (\beta^{(2)}(t))^2 dt = S'D\,S = B'L\,B \qquad (6)$$

The formula (4) can be rewritten as follows:

$$[y - x\beta]'[y - x\beta] + \lambda\beta' L\beta \qquad (7)$$

Where the estimator of the cubic spline $\hat{\beta}$ represents my agency:

$$B^{\wedge}_{\lambda} = (x^T x + \lambda L)^{-1} X'Y \qquad (8)$$

$$\beta^{\wedge}_{\lambda} = R_{\lambda} x' y \qquad (9)$$

$\beta^{\wedge}_{\lambda}$: The cubic spline estimator representing the following vector

$\beta^{\wedge} = [\beta^{\wedge}_{\lambda}(t_1), \ldots, \beta^{\wedge}_{\lambda}(t_n)]'$

R_{λ} : smothing matrix of degree (p×p)

2.2.1 smoothing parameter selection:

The Smoothing parameter (λ) plays an important role in smoothing the slice as (λ) has a strong bias effect. And the variance, as the decrease in the smoothing parameter leads to a decrease in bias and an increase in variance, and vice versa. To obtain a good value for the parameter (λ) by applying the following boot parameter selection method.

1. Generalized Cross-Validation[9]

The General Forensic Crossing Standard (GCV) was proposed by Wahba (1977). and Craven and Wahba (1979) The formula for the GCV standard is as follows:

$$GCV_{\lambda} = \frac{n^{-1} \sum_{i=1}^{n} [y_t - y^{\wedge}_t]^2}{\left[1 - tr\frac{(R_{\lambda})}{n}\right]^2} \qquad (10)$$

$$GCV_{\lambda} = \frac{n^{-1}\, SSE}{\left(1 - \frac{df}{n}\right)} \qquad (11)$$

for the residual sum of squares $\hat{\beta}$ divided by the correction factor $n[1 - tr(R_{\lambda})]^2$ Therefore, the main idea is to determine the value of Boot parameter (GCV) Decrease the sum of the squares of the residuals. To determine the best value for the smoothing parameter..

(3-2) Local Polynomial Kernel Method[8] [9] [10]

The Local Polynomial Kernel Method is widely used in non-parametric due to its smoothness, and also the fact that it can be utilized with any design. This method is based on the function expansion $\hat{\beta}(t)$ for local polynomials through polynomials from a specific degree with fixed points t_0 , and the expansion of β(t) can occur by local polynomial from degree p, as shown below:

$$\beta(t_i)=a_0 + a_1(t_i - t_0) + a_2(t_i - t_0)^2 + \cdots + a_p(t_i - t_0)^p: \qquad (12)$$

The hereunder can be assumed as below:

$$a_j = \frac{1}{j!}\,\beta^j(t)$$

α_j : is the local Polynomial parameter that will be estimated.

And by alternating a_j and α_j in equation 1, we can obtain:

$$a_j = \alpha_j \qquad j = 0, \ldots, p$$

$$\beta(t_i) = \alpha_0 + \alpha_1(t_i - t_0) + \alpha_2(t_i - t_0)^2 + \cdots + \alpha_p(t_i - t_0)^p \qquad (13)$$

the compensation of $\beta(t_i)$ as much as being to it in the hereunder equation (3) :

$$y_i = \sum_{j=1}^{p}\left[\left(\alpha_0 + \alpha_1(t_i - t_0) + \alpha_2(t_i - t_0)^2 + \cdots + \alpha_p(t_i - t_0)^p\right)x_{ij}\right] + e_i \qquad (14)$$

$$e_i = y_i - \sum_{j=1}^{p}\left[\left(\alpha_0 + \alpha_1(t_i - t_0) + \cdots + \alpha_p(t_i - t_0)^p x_{ij}\right)\right] \qquad (15)$$

We can also use the weighted least square (WLS) , by squaring the equation(18) and multiply it with $K_h(t_i - t_0)$, then the total on both parts can be inserted as shown below :

$$\sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n}\left[y_i - \sum_{i=1}^{n}\left[\left(\alpha_0 + \alpha_1(t_i - t_0) + \cdots + \alpha_p(t_i - t_0)^p\right)X_{ij}\right]\right]^2 \times K_h(t_i - t_0) \qquad (16)$$

whereas

$$K_h(t_i - t_0) = \frac{1}{h}K(t_i - t_0)$$

That can be acquired through counting the Kernal function and specifying the invariable h that is called band width with a positive value (h>0), where :

$$I_h(t_0) = [K_h - h\,, K_h + h]$$

In order to create a sloping curve function harmony in pion (t_0), we rely on the core function that helps us to set the observations' weights within this extent $I_h(t_0)$.

By changing the equation (19) to the Matrix's formula as shown below:

$$x = \begin{bmatrix} 1 & (t_1 - t_0) & \dots & (t_1 - t_0)^p \\ \vdots & \vdots & & \vdots \\ 1 & (t_n - t_0) & \dots & (t_n - t_0)^p \end{bmatrix}, \qquad \alpha = [\alpha_0, \alpha_1, \cdots, \alpha_p]$$

$$w= \begin{bmatrix} K_h(t - t_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & K_h(t - t_1) \end{bmatrix}, \qquad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

So, the equation can be typed as shown hereunder:

$$y = x\alpha + e \qquad (17)$$

where :

x: is the design matrix from the degree (p+1)× n

α: is the parameters vector that should be estimated from the degree (p+1)×1

w: a diameter matrix for the core weights from the degree n×n.

The estimated formula of the polynomial parameters can be obtained by decreasing the weighted least squares, as indicated in the below equation:

$$\min_\alpha (y - x\alpha)W(y - x\alpha) \qquad (18)$$

And the derived equals zero, according to the formula mentioned above. The estimated formula for the parameter vector is polynomial as indicated hereunder:

$$\hat{\alpha} = (x'wx)^{-1}x'wy \qquad (19)$$

To derive the function estimator $\hat{\beta}^{(j)}$ (t), we propose that e_j is known as a unit vector with (p+1) dimension, that its first input is one and the output is zero.

$$\hat{\beta}^{(j)}(t) = j!\ e_{j+1}\hat{\alpha} \qquad (20)$$

$$\hat{\beta}^{(j)}(t) = j!\ e_{j+1}\ (x'wx)^{-1}x'wy \qquad (21)$$

To figure out the function's rate $\beta^{(j)}(t)$) at (j=0), which means a zero derived for that function:

$$\hat{\beta}(t) = e'_1\ (x'w\,x)^{-1}x'wy \qquad (22)$$

By alternating the $\hat{\beta}(t)$in equation (3), we obtain the value of $\hat{y}$ as indicated herunder:

$$\hat{y}_i = \sum_{j=1}^{p} \hat{\beta}(t)\ X_{ij} \qquad (23)$$

2. Experimental Aspect

Simulation experiments were carried out using three sample sizes (50, 80, and 150) and with a frequency of 1000 for each experiment, as follows:

1. enerate the explanatory variable so that it follows a normal distribution with zero mean and variance σ^2
2. Generate the variable t so that it follows a normal distribution with zero mean and variance σ^2
3. Generate random error with a mean of zero and three levels of standard deviation (1,1.5,2)
4. Generating the dependent variable with zero mean and variance σ^2 .

The models used in the simulation [9]

1. $\beta(t) = \mathrm{Sin}(2\pi t)$
2. $\beta(t) = (2u - 1)^2$
3. $\beta(t) = \exp(2u - 1) - 1$

Depending on the functions of the variable coefficients mentioned above, the formula for the time-varying coefficient model can be written as follows :

1. y=$(\sin(2\pi t)\ )$x +e
2. y=$((2u-1)^2\ )$x +e
3. y=$(\exp(2u-1)-1)x + e$

Then, the variables generated in the time-varying coefficient models are offset for each standard deviation level as well as for each of the covariate models. A comparison was made between the cubic slice method and the local polynomial kernel method. This is done by using the MSE comparison benchmark.

TABLE 1. shows the standard value (MSE) for all Standard deviation levels, and all samples' sizes in the used non-Parametric levels to estimate the first Time-Varying Coefficient Models.

N	Standard deviation levels for the boundary of error	nonparametric methods	
		LPK	CS
n=30	1	2.425333105	0.878629543
	1.5	0.527315539	0.373426888

	2	0.405726987	0.329363166
n=100	1	1.268415209	0.327636219
	1.5	0.37311243	0.149351827
	2	0.190026046	0.149082479
n=250	1	0.155630851	0.040041725
	1.5	0.068617483	0.045865677
	2	0.180347029	0.177521113

TABLE 2. shows the standard value (MSE) for all Standard deviation levels, and all samples' sizes in the used non-Parametric levels to estimate the second Time-Varying Coefficient Models

N	**Standard deviation levels for the boundary of error**	**nonparametric** methods	
		LPK	**CS**
n=30	1	0.757640964	0.308924692
	1.5	0.499133549	0.315313638
	2	0.356393578	0.320845967
n=100	1	0.303739342	0.149178776
	1.5	0.307044624	0.135863983
	2	0.141044215	0.137477782
n=250	1	0.200368726	0.035275927
	1.5	0.061657798	0.037585694
	2	0.076987562	0.037384848

TABLE 3. shows the standard value (MSE) for all Standard deviation levels, and all samples' sizes in the used non-Parametric levels to estimate the third Time-Varying Coefficient Models.

N	**Standard deviation levels for the boundary of error**	**nonparametric methods**	
		LPK	**CS**
n=30	1	0.351025479	0.32348244
	1.5	0.604802509	0.443702774
	2	0.782182839	0.679312743
n=100	1	0.195861781	0.089802015
	1.5	0.43393289	0.241820254
	2	0.296446759	0.235021258
n=250	1	0.050196069	0.037676291
	1.5	0.054689128	0.044525405
	2	0.136858748	0.112385781

4. CONCLUSIONS

1. Comparing the nonparametric estimation methods used to estimate the model with time-varying parameters showed that the best method was the cubic spline if it met the MSE criterion with the lowest value for all samples and standard deviation levels
2. As the sample size increased, the value of MSE decreased while the standard deviation stayed unchanged.
3. It was found that, if the sample size stayed unchanged, the MSE would increase proportionally to the standard deviation.

5. REFERENCES

[1] Cai, Z., Fan, J., & Li, R. (1999). Generalized varying-coefficient models.

[2] Fan, J., & Gijbels, I. (2018). Local polynomial modelling and its applications. Routledge.

[3] Fan, J., & Zhang, W. (2008). Statistical methods with varying coefficient models. Statistics and its Interface, 1(1), 179.

[4] Green, P. J. and Silverman, B. W. (1994). "Nonparametric Regression and Generalized Linear Models:a Roughness Penalty Approach". Chapman and Hall, London .

[5] Hastie, T., & Tibshirani, R. (1993). Varying-coefficient models. Journal of the Royal Statistical Society: Series B (Methodological), 55(4), 757-779.

[6] Huang, J.Z., Wu, C.O. and Zhou, L. (2004). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. Biometrika, 89, pp.111-128.

[7] Kiebel, S., & Holmes, A. (2011). The general linear model. Statistical Parametric Mapping: The Analysis of Functional Brain Images, 101-125.

[8] Wand, M. P., & Jones, M. C. (1994). Kernel smoothing. CRC press.

[9] Wu, H., & Zhang, J. T. (2006). Nonparametric regression methods for longitudinal data analysis: mixed-effects modeling approaches. John Wiley & Sons.

[10] Xue, L., & Qu, A. (2012). Variable selection in high-dimensional varying-coefficient models with global optimality.

Article submitted 22 August 2022. Published as resubmitted by the authors 27 Dec. 2022.