Study of Some Factors Affecting on the Chickenpox Cases by Using the Partial Linear Regression Model

Nabaa Naeem Mahdi, Auday Taha Raheem, Aseel Abdul Razzak Rasheed
Collage of administration and Economics, Statistics Department, Mustansiriyah University
nabaanaemmahdi@uomustansiriyah.edu.iq
uday_adm@uomustansiriyah.edu.iq
aseelstat@uomustansiriyah.edu.iq

Abstract. Chickenpox is classified as a transmission disease, especially young ages from three months to 15 years old. This study clarifies the effect of the area, population, and the number of the health centers on the number of cases of chickenpox disease in the Rusafa district, Baghdad, Iraq. We use the Partial Linear Model (PLM) that divides the independent variables into two parts (parametric and non-parametric). Moreover, choosing the best model that represents data of the chickenpox disease, using criteria (R2, Bic, Aic) from six models. The results of the study show a positive relation between the number of cases and number of health centers.

Keywords. Partial Linear Model (PLM), Parametric Regression, Nonparametric Regression, Chickenpox disease.

1. Introduction
Chickenpox disease is spread directly through respiratory droplets or skin contact between humans. This means that the risk of developing this disease increases as the population density increases (severity of crowds), which depends on the size and population of the area to be studied. We use the Partial Linear Model (PLM) to find out the effect of some factors on the number of cases of chickenpox disease in each sector in the Rusafa district. The partial linear model is a semi-parametric model, and this model consists of the sum of two compounds. The first component represents the parametric regression model, and the second compound is represented nonparametric regression model. Thus, it contains all the positive advantages found in both the parametric and nonparametric compounds (Jiti, 1995).

Engle et al. (1986) proposed a partial linear model (PLM) when analyzing the relationship between air temperature and the electrical energy. They considered the income variable and the electricity unit variable to be two parametric variables, while the temperature variable would be nonparametric. Jiti and Hau (1995) presented model of the iterative algorithms used of PLM estimates. The parameters of the two models; nonparametric and semi-parametric regression, was estimated by Hardle et al. (2004).

In this study, we formulate several models of partial linear regression to clarify some factors effective on the spread of chickenpox disease in Rusafa district, Baghdad, Iraq. we use several criteria that include Bic, Aic and R2 to determine the best model.

2. Partial Linear Model (PLM)
PLM is a semi-parametric model, which is characterized by being composed of two parts, one part is parametric and another is nonparametric (Jiti, 1995). It was first proposed by Engle, Rice and Weiss as a model that distinguished by its ability to achieve the most important characteristics of the parametric and nonparametric regression models. Also, obtaining the best data curve that is conforming or close to conforming with the dependent variable curve (Hardle et al., 2000).

The methods of estimating this model have varied by researchers. The most used method is the one that tends to estimate the parametric part in the first stage, then the nonparametric part in the second stage, according to the smoothing ways of known nonparametric methods (Ruppert et al., 2003). The model is described by the following equation (Donald and Andrews, 1991):

\[ Y = X^T \beta + m(Z) + \epsilon \quad \ldots (1) \]

\[ E(Y) = X^T \beta + m(Z) \]

where,

\[ Y \]: The response variable.
\[ X^T \beta \]: The linear structure of a parametric.
\[ m(z) \]: The partial nonparametric function.
\[ \epsilon \]: The limit of the random error model.

3. Estimation of the Partial Linear Regression Model

The model includes interact between the parametric and nonparametric part that led to use several methods to estimate the parameters of this model. Some of these methods are follows:

- Residuals Estimation Method.
- Differences Estimation Method (Yatchow Method).
- Speckman-Robinson Estimation Method.
- Speckmen’s method for estimating the nonparametric part development.

Firstly, we define some variables of the model:

\[ Y \]: The number of infections (dependent variable).
\[ x_1 \]: The population of the nonparametric part.
\[ x_2 \]: The number of the health centers (parametric part).
\[ x_3 \]: The area (km²) (parametric part).

3.1. Residuals Estimation Method

The conditional expectation with respect to the nonparametric variables \( Z_i \) is as follows (Hardle et al., 2004):

\[ E(Y/Z) = E(X^T \beta / Z) + E(m(Z) / Z) + E(\epsilon / Z) \quad \ldots (2) \]

Subtract equation (2) From equation (1), yields:

\[ Y - E(Y/Z) = (X^T - E(X^T / Z)) \beta + (m(Z) - E(m(Z) / Z)) + (\epsilon - E(\epsilon / Z)) \]

where as; \[ E(m(Z) / Z) = m(Z) \]

\[ E(\epsilon / Z) = 0 \]

The model becomes as follows:

\[ \bar{Y} - E(\bar{Y}/Z) = (X^T - E(X^T / Z)) \beta + e \quad \ldots (3) \]
From equation (3), we note the amount of nonparametric regression \( E(X^T / Z) \) of the explanatory variable \( (X_i) \) on the nonparametric explanatory variable \( (Z_i) \) is as follows:
\[
E(X^T / Z) = m_X (Z_i) \tag{4}
\]

The nonparametric regression of the dependent variable \( (Y) \) on the nonparametric explanatory variable \( (Z_i) \) is as follows:
\[
E(Y / Z) = m_Y (Z_i) \tag{5}
\]

By substituting equations (4) and (5) into equation (3), we obtain:
\[
Y - m_Y (Z_i) = (X^T - m_X (Z_i))\beta + \epsilon \tag{6}
\]

From equation (6):
\[
D = RB + \epsilon \tag{7}
\]
where, \( D = Y - m_Y (Z_i) \) and \( R = X^T - m_X (Z_i) \).

Equation (7) represents a linear regression equation and its parameters can be estimated by the parametric methods.

\[
\hat{\beta} = (R^T R)^{-1} R^T D \tag{8}
\]

Equation (6) can be written in the form of differences, according to the following equation:
\[
e_1 = e_2 \beta + e \tag{9}
\]
whereas, \( e_1 = Y - m_Y (z) \); \( e_2 = X - m_X (z) \).

The term \( e_1 \) represents the residuals of the observations of the dependent variable \( Y \) and the nonparametric estimators \( m_Y (Z) \). Meanwhile, \( e_2 \) is the residuals of the observations for the explanatory variable \( X \) and the nonparametric estimators \( m_X (Z) \). Thus, the estimation equation that estimate the parameters of the parametric function is as follows: [22]
\[
\hat{\beta}_{res} = (e_2^T e_2)^{-1} e_2^T e_1 \tag{10}
\]
By substituting equation (10) into equation (1), yields:
\[
\hat{\beta}_{res} = Y^T + m(z) + \epsilon \tag{11}
\]
Rearrange equation (11) and put \( A = Y - X^T \hat{\beta}_{res} \), yields:
\[
A = m(z) + \epsilon \tag{12}
\]
Equation (12) represents a nonparametric regression equation with the dependent variable. Meanwhile, \( m(Z) \) represents the nonparametric function, which is estimated by one of the methods of estimating the nonparametric regression. [25]
That estimator $\hat{\beta}_{res}$ adheres to the normal distribution and contains all good properties to OLS: 

$$\hat{\beta}_{res} \approx N(\beta_{res}, \frac{\sigma^2_{\varepsilon}}{n\sigma^2_{\varepsilon}})$$

In the event of heteroscedasticity, it is an estimate $\hat{\beta}_{res}$ as follows:

$$\hat{\beta}_{res/het} = \left[ \hat{\varepsilon}^T_2 \Omega^{-1}_1 \hat{\varepsilon}_2 \right]^{-1} \hat{\varepsilon}^T_2 \Omega^{-1}_1 \hat{\varepsilon}_1$$

... (13)

where, $\Omega_1 = \text{diag}[\varepsilon^2_1, \varepsilon^2_2, \ldots, \varepsilon^2_n]$.

And in the case of autocorrelation, it is an estimate $\hat{\beta}_{res}$ as follows: [22]

$$\hat{\beta}_{res/Acorr} = \left[ \hat{\varepsilon}^T_2 \Omega^{-1}_2 \hat{\varepsilon}_2 \right]^{-1} \hat{\varepsilon}^T_2 \Omega^{-1}_2 \hat{\varepsilon}_1$$

... (14)

$$\Omega_2 = \frac{\sigma^2_n}{1 - p^2} \begin{bmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{n-1} \\ \rho & 1 & \rho & \ldots & \rho^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \ldots & 1 \end{bmatrix}$$

### 3.2. Differences Estimation Method (Yatchow Method)

It is another method for estimating the PLM model that identified by the scientist Yatchow. The two variables ($X$, $Y$) are arranged, according to the basis of the order of the variable ($Z$) 

$$Z_1 \leq Z_2 \leq Z_3 \leq \ldots \leq Z_n$$

so that the first difference is taken as follows (Engle et al., 1986).

$$Y_i - Y_{i-1} = (X_i^T - X_{i-1}^T) \beta + (m(Z_i) - m(Z_{i-1})) + (e_i - e_{i-1})$$

... (15)

In terms of convergence, which means $m(Z_i) - m(Z_{i-1}) = 0$, then:

$$Y_i - Y_{i-1} = (X_i^T - X_{i-1}^T) \beta + (e_i - e_{i-1})$$

... (16)

By supposing, $\Delta \varepsilon = e_i - e_{i-1}$, $\Delta X_i = X_i^T - X_{i-1}^T$, $\Delta Y = Y_i - Y_{i-1}$, then substitute into equation (16), yields:

$$\Delta Y = \Delta X_i \beta + \Delta \varepsilon$$

... (17)

Equation (17) represents the traditional linear regression. The parameters of the parametric part are estimated as follows:

$$\hat{\beta}_{diff} = [\Delta X_i^T \Delta X_i]^{-1} \Delta X_i^T \Delta Y$$

... (18)

By the same way, we can compensate $\hat{\beta}_{diff}$ into equation (1) to estimate the non-parametric part in the second stage. This estimator approximation will achieve all the properties of the desired good OLS estimators:
\[
\hat{\beta}_{\text{diff}} = N\left(\beta_{\text{diff}}, \frac{(1.5)\sigma_{\epsilon}^2}{m\sigma_{\epsilon}^2}\right)
\]

In the event that heteroscedasticity is not uniform or multicollinearity, the estimator's formula \(\hat{\beta}_{\text{diff}}\) are the same formulas of equation (13) and (14) with the same identifiers \(\Omega_1, \Omega_2\).

3.3. Speckman-Robinson Estimation Method

Researchers Speckman and Robinson have proposed two methods of estimate the PLM model. Each method has two stages.

The first one estimates the parametric part of the original data directly using the OLS method and neglecting the effect of the non-parametric part, so as follows:

\[
\hat{\beta}_{SR} = (X^T X)^{-1} X^T Y
\]

Substitute equation (19) into equation (1), yields:

\[
Y - X^T \hat{\beta}_{SR} = m(Z_i) + e
\]

Thus, we can estimate the nonparametric part using traditional non-parametric estimation methods.

Also, using the same method that estimate \(\hat{\beta}_{SR}\) in the case of heteroscedasticity of variance or multicollinearity, which it depends on equations (13) and (14) with the same definition of each case \(\Omega_1, \Omega_2\) so as follows:

\[
\hat{\beta}_{SR/\text{hetro}} = \left[X^T \Omega_1^{-1} X\right]^{-1} X^T \Omega_1^{-1} Y \quad \ldots(21)
\]

\[
\hat{\beta}_{SR/\text{Acorr}} = \left[X^T \Omega_2^{-1} X\right]^{-1} X^T \Omega_2^{-1} Y \quad \ldots(22)
\]

The second method is completely opposite to the first method, where the non-parametric part is first estimated by neglecting the effect of the parametric part. Then, it substitutes into equation (1), which represents PLM so as follows (Muller, 2000):

\[
Y - \hat{m}_{SR}(Z_i) = X^T \beta + e \quad \ldots(23)
\]

This formula represents a traditional linear regression equation, and its parameters can be estimated by any parametric method as follows:

\[
\hat{\beta}_{SR} = (X^T X)^{-1} X^T (Y - \hat{m}_{SR}(Z_i)) \quad \ldots(24)
\]

We can obtain the estimator formula \(\hat{\beta}_{SR}\) in the case of heteroscedasticity of variance or multicollinearity by using equations (13) and (14) with the same definition of each case \(\Omega_1, \Omega_2\) so as follows:

\[
\hat{\beta}_{SR/\text{hetro}} = \left[X^T \Omega_1^{-1} X\right]^{-1} X^T \Omega_1^{-1} (Y - \hat{m}_{SR}(Z_i)) \quad \ldots(25)
\]

\[
\hat{\beta}_{SR/\text{Acorr}} = \left[X^T \Omega_2^{-1} X\right]^{-1} X^T \Omega_2^{-1} (Y - \hat{m}_{SR}(Z_i)) \quad \ldots(26)
\]

Similarities can be seen between the heteroscedasticity of variance and multicollinearity of the mentioned methods, but this similarity in appearance only. Meanwhile, there is a difference in the form of parameters from one method to another.

3.4. Speckmen's method for estimating the non-parametric part development
For motivation, suppose that \( m(.) \) in equation (1) can be parameterized as \((m(z_1), ..., m(z_n))' = W\theta\),

where \( W \) is an \( n \times q \) matrix of full rank and \( \theta \) is additional parameter vector. To assume that \( n \times (p + q) \) matrix \((X, W)\) has full rank, we assume for simplicity that the unit vector \((1, ..., 1)'\) is in the span of \( W \) but not of \( X \), with matrix notation [38].

\[
Y = XB + W\theta + e \quad \text{...(27)}
\]

The normal equations are:

\[
XXB = X'(Y - W\theta) \quad \text{...(28)}
\]

\[
P_W = W(W'W)^{-1}W' \quad \text{Where } P_W \text{ denoted projection on to the column span of } W. \text{Green et al (1985), proposed replacing } P_W \text{ in equation (28) by smoother } M \text{ and simultaneously solving.}\]

\[
B = R(Y - m(Z)) \quad \text{...(29)}
\]

\[
m(Z) = M(Y - XB) \quad \text{Where } R \text{ is a(possibly robust and nonlinear) estimator of treatment effect, Taking } M \text{ to be matrix } S \text{ form Kernel smoothing and letting } R = (XX')^{-1}X' \text{ produces the defining equation for the Green – Jennison - Seheult (GJS)(1985) estimator.}\]

\[
\hat{m}_{GJS}(Z) = S(Y - XB_{GJS}) \quad \text{...(30)}
\]

And

\[
XX\hat{B}_{GJS} = X'(Y - S(Y - XB_{GJS})) \quad \text{...(31)}
\]

Hence

\[
B_{GJS} = (XX'I - S)X'(I - S)Y \quad \text{...(32)}
\]

4. Chickenpox Disease

Chickenpox is a transitional viral disease, very contagious that is transmitted from one person to another through direct skin contact, or by respiratory droplets. Then, it reaches the vesicle fluid or respiratory secretions. Also, it may be transmitted indirectly when using contaminated devices of the patients.

Some of the symptoms of this disease are appearance of mild fever with physical symptoms Then rash in the form of spots that remain a few hours or in the form of vesicles for a period (3-4) days. The cause of this disease is a virus (vzv) that infect human from the age of three months to (15) years. Most of patients, nearly 90%, are infected before reach 15 years old.

The incubation period of the disease is between 2 to 3 weeks. The infection begins two days the rash appears ago, and all pimples are dried. Anyone who suffers from this disease will has immunity to infection again. The disease classifies as a non-fatal, but its dangers the children with HIV.

To prevent and combat this disease, isolate the affected person until the vesicles dry out five days later Then cleaning the sanitary ware well in addition to informing the authorities to take the necessary action (Chin, 2000).

5. Data collection
The number of chickenpox cases are taken from the Baghdad Health Authority, Al-Rusafa Health Department - Transitional Disease Control Division, 2017. Table (1) shows the data that classified according to area and numbers of health centers and population for all sectors of Al-Rusafa, Baghdad. Column six represents the population density that depends on the population and area of the sector.

Table 1. shows the number of cases of transitional chickenpox disease in 2017 in the Rusafa district – Baghdad

| Number | Sector          | No. of health center | population | Area Km² | Density Population | chickenpox |
|--------|-----------------|----------------------|------------|----------|--------------------|------------|
| 1      | Adhamiya        | 8                    | 220743     | 15.8     | 13971              | 794        |
| 2      | Al-aistiqlal    | 11                   | 319940     | 92.2     | 3470               | 913        |
| 3      | Al-Rusafa       | 10                   | 260837     | 62.6     | 4167               | 1147       |
| 4      | New Baghdad     | 8                    | 373116     | 72       | 5182               | 315        |
| 5      | Al-baladiatfirst| 11                   | 675963     | 32       | 21124              | 1725       |
| 6      | Al-baladiatsecond| 11                  | 561688     | 53.8     | 1044               | 639        |
| 7      | Al-Mada'in      | 15                   | 538609     | 424.48   | 3366               | 954        |
| 8      | Sadr City       | 21                   | 1163924    | 47.5     | 24504              | 2532       |
| 9      | Al-shaeb        | 14                   | 551774     | 50.4     | 10948              | 2712       |

Table (1) shows the number of chickenpox cases is positively related with the population density. This means, positively and negatively related with the population and area size, respectively. Also, the few health centers and the high population density lead to an increased possibility of this disease. Al-shaeb sector is the most affected with (2712) cases and population density (10948), while the New Baghdad sector is the lowest affected with (315) cases and population density (5182).

6. Building a Partial Linear Regression Model

The partial (semi-parametric) linear regression model is characterized by the ability to distribute the explanatory (independent) variables over the parametric and non-parametric part of the model. This feature enabled the model to explain the contradictions, such as the area size and the population are large and the number of cases is limited. We distribute these independent variables to know the effect of the parametric and non-parametric on the incidence of this disease. Six PLM models are formulated and use the (i-xplore) to compute these models (Jiti, 1995). We calculate the band width by using the formula that proposed by Scott for more than one variable so as follows (Scott, 1992):

\[ h_k = \sigma \cdot h^{-\frac{1}{4}}, k = 1,2,3,...,l \] ...

(33)

Where,

\[ l \]: The number of explanatory variables.

\[ h \]: Band width.

\[ \sigma \]: The amount of variance of the variable in \( k \).

\[ h_k \]: The smoothing parameter of the variable ink.
6.1. PLM Model with One Variable in the Parametric Part.
We suppose the behavior of one of the variables is linear and create the parametric part. The rest of the variables are behaved as a non-parametric (non-linear), the non-parametric part of the PLM model. Three models of partial regression models are constructed by using equation (1). Table (2) shows the band width of the explanatory variables:

| Table 2. band width for each explanatory variable | Band width |
|-----------------------------------------------|------------|
| Population (X₁) | h₁=519391.85 |
| No. of health center (X₂) | h₂=7.3680161 |
| Area (X₃) | h₃=227171.01 |

Table 3. Shows the models that were built

| Parametric part | Non-parametric part | Mathematical equation |
|-----------------|---------------------|-----------------------|
| X₁ | m (X₂,X₃) | \( \hat{Y} = 0.00134802X₁ + m (X₂,X₃) \) |
| X₂ | m (X₁,X₃) | \( \hat{Y} = 331.354X₂ + m (X₁,X₃) \) |
| X₃ | m (X₁,X₂) | \( \hat{Y} = -0.00326249X₃ + m (X₁,X₂) \) |

From Table (3), we deduce the effect of the variables (X₁ and X₃) that represent the population and area size are weak, in the case of its are parametric. This means, the variables X₃ and X₁ are not efficient when it’s are parametric. Meanwhile, the effect of the variable X₂ (number of health centers) is strong. Overall, the number of chickenpox cases are affected by the number of health centers.

7. Comparison of models
In this section, we make a comparison between models to find the best and determine the variables behavior; parametric or non-parametric. Three criteria are used; the Aic standard, the Bic standard, and the R². Table (4) shows the results.

| Table 4. shows the values of the Aic, Bic, adj R², R² criteria for each PLM |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| Model | Parametric part | Non-parametric part | Aic | Bic | R² | adj R² |
| 1 | X₁ | m(X₂,X₃) | 148.104 | 148.9947 | 0.6055 | 0.2961 |
| 2 | X₁ | m(X₁,X₃) | 142.4184 | 143.3374 | 0.8002 | 0.6317 |
| 3 | X₁ | m(X₁,X₂) | 145.2030 | 146.0001 | 0.6668 | 0.4624 |
| 4 | X₁,X₂ | m(X₃) | 145.1657 | 145.9622 | 0.6679 | 0.4645 |
| 5 | X₁,X₃ | m(X₂) | 146.8176 | 146.8176 | 0.6646 | 0.4304 |
| 6 | X₁,X₃ | m(X₁) | 138.8417 | 139.7075 | 0.8530 | 0.7449 |

From Table (4), we find that the best model is Model No. (6) because it has the lowest values of Aic and Bic standard. Also, it had the highest representation of data through the value of the determination coefficient (R²=0.853) and the adjusted determination coefficient (adj R²=0.744). It
represents the parametric part of the variables $X_2$ (number of health centers) and $X_3$ (area size). The non-parametric part represents the variable $X_1$ (population). The mathematical formula is as follows:

$$\hat{Y} = 361.444X_2 - 0.00515564X_3 + m(X_1)$$

The variable $X_2$ (number of health centers) and $X_3$ (area) have a stable linear behavior, as the increase in one unit of $X_2$ (the number of health centers) leads to an increase in the number of cases of chickenpox ($Y$) by (361.444). Also, the increase of one unit of the variable $X_3$ (the area km^2) leads to a decrease in the number of cases of chickenpox by (-0.00515564). The non-parametric part (non-stationary) represents the variable $X_1$ (population) that has non-linear behavior so as shows in Figure (1).

![GQLM fit, 'noid', n=9](image)

Figure 1: PLM model for parametric part ($X_2, X_3$), nonparametric part ($X_1$)

8. Conclusions and recommendations

The PLM model distinguishes by its high ability to separate between the variables that behave in a linear and non-linear. The study discussed the effect of number of health centers, area size, and population on the number of chickenpox cases. From six models, the best model that took the number of health centers and area size as a linear (stable), while population was non-linear (unstable).

The increase in the number of health centers led to increase in the chickenpox cases, while cases decrease as a result to increase the area size. The number of populations had a non-linear behavior because it depends on the nature of human behavior in terms of human reproduction or movement from one place to another.

Increasing the number of health centers and the area size will help to treat and protect from this disease. This means, reducing the possibility of infection or transmission of the disease.

References

[1] Abdous ,B, Kokonendj,C,Senga.T., 2012,"On Semi Parametric Regression for count explanatory variables", Journal of Statistical Planning and Inference, pp.(1537-1548).

[2] Akkus, O., (2011) , “Xplore package for the popular parametric and semi-parametric single index models ”, Journal of science ,vol.24 , No.4, pp.(753-762) .

[3] Ali, F.A. and Salih, T. A. , (2017) , “Analysis of semi-parametric single-index models by using MAVE-method based on some kernel functions ”, International Journal of Advanced Statistics and Probability , No.5 Vol.1 , P.P (37-43)
[4] Al-Safawi, Safaa Y. and Matti, Nour S., (2011), "Estimating non-parametric regression functions using some methods of smoothing", The Iraqi Journal of Statistical Sciences, (20) is a special issue of the proceedings of the Fourth Scientific Conference of the College of Computer Science and Mathematics.

[5] Al-sharot, Muhammad H. and Khanjar, Muhammad T., (2016), "Comparing some semi-parametric estimators using simulation", Republic of Iraq, College of Computer Science and Mathematics, Al-Qadisiyah University.

[6] Chen & Hung, 1988, "Convergence Rates For parametric Components in a Partly Linear Model", The Annals of Statistics, Vol, 16, No.1, pp. (136-146).

[7] Chiang, C.T. and Huang, M.Y., (2011), "New estimation and inference procedures for a single index conditional distribution model", Journal of Multivariate Analysis, pp.(271–285).

[8] Cuzick & Jack, 1992, "Efficient Estimates in Semi Parametric Additive Regression Models With Unknown Error Distribution", The Annals of Statistics, Vol.20, pp.(1129 – 1136).

[9] Cuzick & Jack, 1992, "Semi parametric Additive Regression", Journal of the Royal Statistical Society Series Methodological, Vol.54, No. 3, pp. (831-843).

[10] Demir, S., Toktamis, O., "On the adaptive Nadaraya-Watson kernel regression estimators", Journal of Mathematics and Statistics Volume 39 (3) (2010), 429 – 437.

[11] Donald W. K., Andrews, 1991, "Asymptotic normality of series estimators for Non parametric And Semi Parametric Regression Models", Econometrica Vol.59, No.2, pp. (307-345).

[12] ElianaChristof, (2014), "Asymptotic Behavior of Nadarya-Watson and Local Linear Estimators Associated With The Single Index Models", The Pennsylvania State University, The Graduate School.

[13] Engle, R. F., Granger, C.W. J., Rice, J. & Weiss, A., 1986, "Semi parametric estimators of the relation between weather and electricity sales", Journal of the American Statistical Association, Vol. 81, No. 394, pp.(310-320).

[14] Eubank R.L., Whitney P., 1989, "Convergence Rates for Estimation in Certain partially linear models", Journal of statistical planning and inference, Vol. 23, pp. (33-43), north Holland.

[15] Eubank, R.L., Speckman, P, 1990, "Curve Fitting by Polynomial Trigonometric Regression", Biometrika, Vol.77, No.1, pp. (1-9).

[16] Gao.Jiti, 1995, "Asymptotic Theory for Partly linear models Communications in Statistics-Theory and Methods", 24:8.

[17] Gao.Jiti, 1995, "The laws of the iterated logarithm of some estimates in partly linear models", Statistics and Probability Letters, 25, pp. (153-162).

[18] Gao.Jiti and Liang.Hau, 1997, "Statistical Inference in Single Index &Partial Nonlinear Models", Annals of the Institute of Statistical Mathematics, Vol.49 Issue 3, pp. (493-517).

[19] German Aneiros & Alejandro Quintela, 2001, "Asymptotic properties in partial linear models under dependence", Sociedad de EstadisticaInvestigacion Operativa Test, Vol. 10, No.2, pp. (333-355).

[20] Haggag. Magda, 2007, "On the Estimation of a Semi parametric Generalized Linear Model", Interstat.atat, Journals.net articles. 0707004, pdf.

[21] Hansen, Bruce E., "Econometrics"(2012), University of Wisconsin.

[22] Hardle W. and Huang L.S., 2013, "Analysis of Deviance in Generalized Partial Linear Models", SFB649, Humboldt-Universitatz Berlin, SpandauerStrabe.1,(D-10178), Berlin.

[23] Hardle W, and Liang, H. & Gao., J., 2000, "Partilly Linear Models", Heidelberg: Physica –Verlag.

[24] Hardle, W, Mammen E. and Muller, M., 1998, "Testing Parametric versus Semi parametric Modeling in Generalized Linear Models", Journal of the American Statistical Association, Vol.93, No.444, pp. (1461-1474).

[25] Hardle W, Muller, M, Sperlich, S &Werwatz, A, 2004, "Nonparametric and Semi parametric Models an Introduction", Spring Edition.

[26] James chin, MD, MPH, 2000, "control of communicable Diseases Manual ", an official Report of the American Public Health Association, 7 h editions.
[27] Jiang, S. (2013) “Variable selection in Semiparametric model” Annals of Statistics, Vol.38, No.1, PP 261-286.
[28] Khalid.C,Djamal.L”Asymptotic results for the linear parameter estimate in partially linear additive regression model”,C. R. Acad. Sci. Paris, Ser. I 349 (2011) 1105–1109.
[29] Liu, X., (2011), “penalized variable selection for semiparametric regression models”, submitted in partial fulfillment of the requirements for the degree doctor of philosophy – University of Rochester, New York.
[30] Moshe, Sandy Q., (2012), "Comparing Backfitting and SIMEX Estimator for an Estimating Aditive nonparametric regression model with application "Master Thesis in Statistics, College of Administration and Economics / University of Baghdad.
[31] Muller, Marlene, 2000, "Semi parametric Extension GLMz", der Humboldt-Universitat zu Berlin.
[32] Peng, H. and Huang, T. , (2011), “Penalized Least squares for single-index models ”, Journal of Statistical Planning and Inference , vol.141 , pp.(1362-1379).
[33] Qasem , Habeb ,S. ,(2018), "Comparison parametric &semi-parametric single index model With application ", Master Thesis, College of Administration and Economics ,Al-Mustansiriya University, Baghdad.
[34] Ruppert, D., Wand, M. and Carroll, R., 2003," Semi parametric Regression", Cambridge University Press.
[35] Saleh,Tariq A., (2016) ,"Some semi-parametric methods in estimating and choosing the variable for a single index model", PhD thesis, College of Administration and Economics, University of Baghdad.
[36] Scott, D.W., 1992,"Multivariate Density Estimation", John Wiley & Son, New York.
[37] Sohail Chand , (2012) , “ on tuning parameter selection of lasso – type Methods-A Monte Carlo study ”, Applied sciences and Technology , IEEE xplore , pp.(120-129).
[38] Speckman .P ,"Kernel Smoothing in Partial Linear Models ” Journal of the Royal Statistical Society. Series B (1988), Vol.50, No.3, pp.413-436.
[39] Su, L. and Zhang Y. , (2012) , “ variable selection in non- parametric and semi-parametric regression model ”, school of Economics , Singapore Management university.
[40] Tarad, Alaa J. ,(2013), "Parametric &semi-parametric regression models (comparative study)",Master Thesis, College of Administration and Economics ,Al-Mustansiriyah University, Baghdad.
[41] Tawfiq A. Al-asadi, Ahmed J. Obaid, Ahmad A. Al-Khayatt, Proposed Method for Web Pages Clustering Using Latent Semantic Analysis, Journal of Engineering and Applied Science, vol.12, no 6, 8279-8277, 2018.