International Journal of Applied Mathematics
& Statistical Sciences (IJAMSS)
ISSN (P): 2319–3972; ISSN (E): 2319–3980
Vol. 7, Issue 8, Aug–Sep 2019; 15–30

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

DIFFERENCE FOR DIFFERENCE ESTIMATION METHOD FOR SEMIPARAMETRIC PARTIALLY LINEAR REGRESSION MODELS

Magda M. M. Haggag

Associate Professor, Head of the Department of Statistics, Faculty of Commerce, Damanhour University, Egypt

ABSTRACT

This paper proposes a new difference-based estimation method for estimating the semi parametric partially linear model (PLM). This method is called the difference for difference (DFD) estimation method, which is proposed by the author, for estimating the residual variance in nonparametric regression models. In this work, the DFD estimation method is used for estimating both the parametric component and the residual variance. A numerical study has been shown that the proposed DFD estimation gives best results compared to other existing difference methods; in the form of less mean squared error of parametric component and less residual variance of the fitted model.

KEYWORDS: *Difference-Based Estimation, Mean Squared Error, Partially Linear Models, Residual Variance, Semi Parametric Regression*

Article History

Received:* *29 Jun 2019* | *Revised:* *14 Aug 2019* | *Accepted:* *31 Aug 2019

INTRODUCTION

A partially linear regression model (PLM) is a well-known semi parametric regression model, which contains both parametric and non-parametric components. Different estimation schemes are used in estimating PLM, such as kernel smoothing, spline smoothing, and difference-based method. (See for kernel and spline smoothing methods: Silverman (1985), Engle et al., (1986); Heckman (1986); Robinson (1988); Speck man (1988); Buja et al., (1989); Hastie and Tibshirani (1990); Wahba (1990); Severini and Wang (1992); Zhou et al (1998); Opsomer and Ruppert (1999); Hardle et al., (2000; and 2004); Muller (2001); Ruppert et al., (2003); and Holland (2017)). (See for difference-based method: Yatchew (1997; and 2003); Liu (2010); Wang et al., (2011); and Zhou et al. (2018). The interest in this work, is to use the difference-based estimation approach for estimating the PLM.

Consider the following partially linear model (PLM),

$$Y = X^T \beta + f(U) + \varepsilon \,, \tag{1}$$

where X is an *(nxp)* matrix of regressors in the parametric component of the model, U is an *(nxq)* matrix of regressors in the nonparametric component, β is a *(px1)* vector of unknown parameters, *f(U)* is an unknown function (a nonparametric function) from R^q to R, ε is an independent vector of random errors with mean zero and finite variance σ^2. A PLM in (1) is a semiparametric model since it contains parametric and nonparametric components. PLM is a preferred regression model than a fully parametric model and a fully nonparametric one, since PLM is a more flexible than the first one and combat the curse of dimensionality which is a well-known problem in nonparametric regression models. The difference estimation method is based on removing the nonparametric component of PLM in (1), i.e., removing the function *f(U)*. The differencing method will be applied to estimate both the parametric component, β, and the residual variance σ_{ε}^{2}.

The difference-based estimation is applied to the nonparametric model to estimate the residual variance. Hall et al., (1990), proposed a new difference sequence to estimate the residual variance. This sequence is called the optimal sequence, since it is derived by minimizing the asymptotic mean square error of the difference variance.The optimal sequence is computed for $m=1,2,\ldots,10$ by Hall et al., (1990), and their outputs will be developed, in this work, to estimate the PLM.

Dette et al., (1998) introduced a difference sequence which is called the ordinary difference sequence by Hall et al., (1990). This sequence is used for numerical differentiation and employed to reduce the bias in small sample sizes. The ordinary sequence will be developed for estimating the PLM.

Haggag (2019), proposed new difference estimation for the residual variance in nonparametric regression, called the difference-for-difference (DFD) estimation method. The DFD method gave best results for the nonparametric regression, and will be developed in this work for estimating the PLM. The three difference sequences will bedeveloped and compared for the PLM.

The rest of this paper is organized as follows section (2) presents the difference-based estimation method for the partially linear model (PLM), and its various difference sequences. Section (3) introduces the proposed difference-for-difference estimation for PLM and its theoretical properties. Section (4) considers the simulation study under different settings. The conclusions of this work are presented in section (5).

The Difference-Based Estimation for Partially Linear Models

Yatchew (1997; and 2003) proposed the difference-based estimation method to estimate the PLM. Their interest was on estimating the parametric component of the model, and they showed that using higher difference order the nonparametric component *f(Z)*will be removed. Also, they found that the relative efficiency of the estimator converges to the asymptotic efficiency, when using higher differences.

Wang et al., (2011) proposed a difference-based approach for estimating the PLM. This approach is based on differencing the parametric component, then estimating the nonparametric one using by using any nonparametric method such as kernel or wavelet.

Differencing the Data Give the Following First-Difference Equation

$$y_{i+1}-y_{i}=\left(x_{i+1}-x_{i}\right)\beta+\left[f\left(u_{i+1}\right)-f\left(u_{i}\right)\right]+\varepsilon_{i+1}-\varepsilon_{i}\text{, i=1,2,…,n} \quad (2)$$

From (2), the m[th] difference order can be generalized as follows:

$$\sum_{j=0}^{m}d_{j}y_{i+j}=\left(\sum_{j=0}^{m}d_{j}x_{i+j}x_{i+1}\right)\beta+\sum_{j=0}^{m}d_{j}f\left(u_{i+j}\right)+\sum_{j=0}^{m}d_{j}\varepsilon_{i+j}, i=m,\ m+1,\ldots,n-1. \quad (3)$$

where, $d_j$is the j^{th} difference sequence for $j=0,1,2,\ldots,m$. The first order difference is defined as: $d_1=\{d_0,d_1\}$,the second order difference is: $d_2=\{d_0,d_1,d_2\},\ldots$,and the m order differences is:$d_m=\{d_0,d_1,d_2,\ldots,d_m\}$. The estimation of (3) is based on a difference sequence of real numbers $\{d_j\}, j=0,1,2,\ldots,m$, with two constraints as follows:

$$\sum_{j=0}^{m} d_j = 0, \quad \sum_{j=0}^{m} d_j^2 = 1, \tag{4}$$

where m is the order of the differences.According to the constraints in (4), the first-order sequence will be as follows for all difference methods:

$$\{d_j\} = \{d_0, d_1\} = \left\{\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right\}. \tag{5}$$

The Following Assumptions will be considered throughout the Working in this Paper.

Assumption 1: The design points of u are equidistant such that $u_i = \frac{i}{n}$, $1 \le i \le n$.

Assumption 2: The error term ε_i in (1), is an independent and identical (iid) distributed with mean=zero, and variance= σ^2, with $E(\varepsilon^2) \prec \infty$, and also, $E(\varepsilon^4) \prec \infty$.

Assumption 3: It is assumed that the maximal distance between any two adjacent observations, i.e., $\max|u_{i+1} - u_i| = O\left(\frac{1}{n}\right)$

Assumption 4: The nonparametric function $f(u)$ in (1) has a bounded first derivative.

The difference sequences differ according to the type of differences. Three types of difference sequences will be considered in this paper as follows.

The Difference Sequence of Hall Et Al. (1990)

Hall et al. (1990) proposed a difference sequence called the optimal sequence, they used it in estimating the residual variance in nonparametric regression and they found that the optimal sequence $\{d_j\} = \{d_0, d_1, \ldots, d_m\}$ can be obtained by minimizing the following quantity:

$$\delta = \sum_{k=1}^{m} \left(\sum_{j=0}^{m-k} d_j d_{j+k} \right)^2, \tag{6}$$

And the numerical sequences are computed according to the constraints in (4) as shown in Table (1). These sequences will be developed, in this work, for estimating the PLM as “Hall sequences”.

Table 1: The Computed Difference Sequence of Hall et al., (1990)

M	Optimal Sequence $(d_0, d_1, \ldots, d_m)$
1	(0.7071,-0.7071)
2	(0.8090,-0.5000,-0.3090)
3	(0.1942,-0.2809, 0.3832,-0.8582)
4	(0.2708,-0.0142, 0.6909,-0.4858,-0.4617)
5	(0.9064,-0.2600,-0.2167,-0.1774,-0.1420,-0.1103)
6	(0.2400, 0.0300,-0.0342, 0.7738,-0.3587,-0.3038,-0.3472)
7	(0.9302,-0.1965,-0.1728,-0.1506,-0.1299,-0.1107,-0.0930,-0.0768)
8	(0.2171, 0.0467,-0.0046,-0.0348, 0.8207,-0.2860,-0.2453,-0.2260,-0.2879)
9	(0.9443,-0.1578,-0.1429,-0.1287,-0.1152,-0.1025,-0.0905,-0.0792,-0.0687,-0.0588)
10	(0.1955, 0.0539, 0.0104,-0.0140,-0.0325, 0.8510,-0.2384,-0.2079,-0.1882,-0.1830)

Source: Hall et al., (1990).

The Difference Sequence of Dette Et Al., (1998)

Dette et al. (1998) introduced the following difference sequence:

Table 2: The Computed Difference Sequence of Dette et al., (1998)

M	Ordinary Sequence $(d_0, d_1, \dots, d_m)$
1	(0.7071,-0.7071)
2	(0.4082,-0.8165,0.4082)
3	(0.2236,-0.6708, 0.6708,-0.2236)
4	(0.1195,-0.4780, 0.7170,-0.4780, 0.1195)
5	(0.0630,-0.3145, 0.6299,-0.6299, 0.3150,-0.0630)
6	(0.0329,-0.1974, 0.4935,-0.6580, 0.4935,-0.1974, 0.0329)
7	(0.0171,-0.1195, 0.3591,-0.5985, 0.5985,-0.3591, 0.1195,-0.0171)
8	(0.0088,-0.0704, 0.2464,-0.4928, 0.6160,-0.4928, 0.2464,-0.0704, 0.0088)
9	(0.0045,-0.0405, 0.1620,-0.3780, 0.5670,-0.5670, 0.3780,-0.1620, 0.0405,-0.0045)
10	(0.0023,-0.0230, 0.1035,-0.2760, 0.4830,-0.5796, 0.4830,-0.2760, 0.1035,-0.0230, 0.0023)

$$d_j = (-1)^j \binom{2m}{m}^{-\frac{1}{2}} \binom{m}{j}, \text{ j=0, 1, 2, } \quad ,\text{m} \tag{7}$$

The sequence d_j in (7), is called the ordinary difference sequence by Hall et al., (1990). This sequence is employed to reduce the bias in small sample sizes.The ordinary sequence is computed according to the constraints in (4) as shown in Table (2). These sequences and will be developed, in this work, for estimating the PLM as "Dette Sequences".

Haggag (2019) New Optimal Difference Sequence

Haggag (2019) proposed the difference-for-difference (DFD) method for estimating the residual variance in nonparametric regression, and the corresponding DFD sequence is found to be as follows:

$$d_j = \begin{cases} (-1)^{j+2}(m^*)^{-\frac{1}{2}}, & j = 0, m \\ (-1)^{j+2}(m)(m^*)^{-\frac{1}{2}}, & j = 1,2,\dots,(m-1) \\ 0, & otherwise \end{cases}, \tag{8}$$

and for even m is found to be:

$$d_j = \begin{cases} (-1)^{j+2}\left(\frac{m}{2}\right)(m^*)^{-\frac{1}{2}}, & j = 0, m \\ (-1)^{j+2}(m)(m^*)^{-\frac{1}{2}}, & j = 1,2,\dots \quad ,(m-1) \\ 0, & otherwise \end{cases}, \tag{9}$$

where,

$$m^* = \left[m^2(m-1) + 2\right] \tag{10}$$

Table 3: The Computed DFD Sequence of Haggag (2019)

M	Dfd Sequence $(d_0, d_1, ..., d_m)$
1	*(0.7071,-0.7071)*
2	*(0.4082,-0.8165,0.4082)*
3	*(0.2236,-0.6708, 0.6708,-0.2236*
4	(0.2828,-0.5657, 0.5657,-0.5657,0.2828)
5	(0.0990,-0.4951, 0.4951,-0.4951, 0.4951,-0.0990)
6	(0.2223,-0.4447, 0.4447,-0.4447, 0.4447,-0.4447,0.2223)
7	(0.0581,-0.4068, 0.4068,-0.4068, 0.4068,-0.4068, 0.4068,-0.0581)
8	(0.1886,-0.3771, 0.3771,-0.3771, 0.3771,-0.3771, 0.3771,-0.3771, 0.1885)
9	(0.0392,-0.3530, 0.3530,-0.3530, 0.3530,-0.3530, 0.3530,-0.3530, 0.3530,-0.0392)
10	(0.1665,-0.3330, 0.3330,-0.3330, 0.3330,-0.3330, 0.3330,-0.3330, 0.3330,-0.3330, 0.1665)

The DFD sequences in (8) and (9) are derived to satisfy the two constraints in (4), and shown in table (3). When m=1, the DFD sequence in (8) will be the unique sequence d_j in (5). For m=2, the DFD sequence in (8) will be the ordinary sequence d_j as shown in Table (2). Also, for *m=3*, the ordinary sequence and the DFD sequence d_j are coincident as shown in Table (2). It is found that the ordinary sequence and the DFD sequence d_j differ for *m >3*. (See Tables (2) and (3)

Estimating the Partially Linear Models Using the Difference approach

Consider the PLM defined in (1), the generalized difference equation in (3) can be written as:

$$\tilde{Y} = \tilde{X}^T \beta + \tilde{f}(U) + \tilde{\varepsilon}, \tag{10}$$

where,

$\tilde{Y} = \Delta Y$, $\tilde{X} = \Delta X$, m and $\tilde{\varepsilon} = \Delta \varepsilon$. Δ is an (n-m)×n matrix of differences defined as:

$$\Delta = \begin{bmatrix} d_0 & d_1 & d_2 & . & . & . & d_m & 0 & 0 & . & . & . & . & 0 \\ 0 & d_0 & d_1 & d_2 & . & . & . & d_m & 0 & 0 & . & . & . & 0 \\ . & . & . & & & & & . & . & & & & . & . \\ . & . & . & & & & & . & . & & & & . & . \\ . & . & . & & & & & . & . & & & & . & . \\ 0 & 0 & 0 & . & . & . & . & & d_0 & d_1 & d_2 & . & d_m & 0 \\ 0 & 0 & 0 & . & . & . & . & & & d_0 & d_1 & . & . & d_m \end{bmatrix}$$

where $d_0, d_1, d_2, \ldots, d_m$ are the differencing weights satisfying the constraints in (4).

According to assumptions 1:4, the difference model in (10) can be written as:

$$\tilde{Y} = \tilde{X}^T \beta + \tilde{\varepsilon}, \tag{11}$$

The difference estimator of β is defined as follows:

$$\hat{\beta}(diff) = \left(\tilde{X}^T \tilde{X}\right)^{-1} \tilde{X}^T \tilde{Y}, \tag{12}$$

where $\tilde{X}$ is an (n-m)×p matrix of design points on regressors, and $\tilde{Y}$ is an (n-m)×1 vector of response variable. Also, the residual variance can be written as:

$$\hat{\sigma}^2(diff) = \tilde{Y}^T P \tilde{Y} / (n-m-p), \tag{13}$$

where P is an (n-m)×(n-m) matrix such that:

$$P = I - \tilde{X}\left(\tilde{X}^T \tilde{X}\right)^{-1} \tilde{X}^T, \tag{14}$$

And I is an (n-m) × (n-m) identity matrix. The estimator of the parametric components, β, and the residual variance will differ according to the type of the difference sequence. Therefore, three difference estimators will be obtained and compared as follows:

- $\hat{\beta}(opt)$ and $\hat{\sigma}^2(opt)$ using the optimal sequences in table 1.
- $\hat{\beta}(ord)$ and $\hat{\sigma}^2(ord)$ using the ordinary sequences in table 2.
- $\hat{\beta}(DFD)$ and $\hat{\sigma}^2(DFD)$ using the optimal sequences in table 3.

Theorem 1

Consider the assumptions 1:4 and the DFD sequence in (8), which computed in table (3), then:

$$\hat{\beta}(DFD) \rightarrow N(\beta, \Lambda)$$

Where Λ is the covariance of $\hat{\beta}(DFD)$ defined as:

$$\Lambda = \sigma^2\left(1 + 2\sum_{k=1}^{m}\delta_k^2\right)\left(X^T X\right)^{-1}$$

$$\hat{\sigma}^2(DFD) \xrightarrow{p} \sigma_\varepsilon^2.$$

The Proof in Appendix (A).

A Simulation Study

Simulation Design

To study the effect of the difference sequence on the estimation of PLM, three estimators of β will be considered as follows:

- $\hat{\beta}(Hall)$ which uses the sequences of Hall et al. (1990) shownin Table (1).
- $\hat{\beta}(Dette)$ which uses the sequences of Dette et al (1998) shown in Table (2).
- $\hat{\beta}(DFD)$ which uses the sequences of Haggag (2019) shown in Table (3).

Two sample sizes are used: n=50, and 200. The parameter vector β = (0.2, 0.3, 0.5), and X_j ~ *Uniform (0, 1)*. The nonparametric function used in this study is *f(u)= 2 Sin (wπu)*, which used by Zhou et al., (2018), with weights w = 0, 2, 4, and 6, which represents the smoothness of *f(u)*, and presented as f_0, f_2, f_4, *and* f_6. Also, the design points are chosen equidistant such that $x_i = \frac{i}{n}$ for i=1, 2,…,n. the errors are generated such that $\varepsilon_i \sim N(0,1)$. The procedures are repeated 1000 times and the average MSE of each parameter estimator is computed for $\hat{\beta}(Hall)$, $\hat{\beta}(Dette)$, and $\hat{\beta}(DFD)$. Also, the residual sum of Squares for different models are presented. The results are recorded for 240 models: three different estimation methods of β, four nonparametric function f(u) according to the value of w, ten difference sequences m, and two sample sizes, i.e. (3×4×10×2=240 models).

Simulation Results

Simulation Results of the Difference Estimator of Β

The effect of the proposed difference sequence, introduced by Haggag (2019), on the estimation of the PLM is considered and compared with the two difference sequences introduced by Hall et al., (1990), and Dette et al (1998). The results are shown in Tables (4): (7) and Figures 1:8, for the sequences $1 \leq m \leq 10$ and for the sample sizes, n = 50 and n = 200. When n=50, best results are obtained, in the form of less MSE, for DFD when $1 \leq m \leq 7$ and for all values of weight for f(u). Bad results are obtained, in the form of large values of MSE, for Hall especially when m = 4, 10, and w = 0, and when m = 4, 8, 10, and w = 2, 4, 6. Best results are obtained for Dette when m < 7, and for values of w. when n=200, DFD gave best results compared to Hall and Dette for $m \geq 4$, and for all values of w = 0, 2, 4, and 6.

Table 4: The Mean Squared Error (MSE) of the Parameter Vector Β, in the PLM, Using $\hat{\beta}(Hall)$, $\hat{\beta}(Dette)$ and $\hat{\beta}(DFD)$. The Results are Recorded of Order $1 \leq M \leq 10$, Four Weights (0,2,4,6) for the Nonparametric Function F (U), and Sample Size N = 50

Difference order (m)	MSE of $\hat{\beta}(Hall)$				MSE of $\hat{\beta}(Dette)$				MSE of $\hat{\beta}(DFD)$			
	f_0	f_2	f_4	f_6	f_0	f_2	f_4	f_6	f_0	f_2	f_4	f_6
1	0.0261	0.0228	0.0198	0.0171	0.0261	0.0228	0.0198	0.0171	0.0261	0.0228	0.0198	0.0171
2	0.0367	0.0240	0.0144	0.0077	0.0208	0.0208	0.0216	0.0208	0.0208	0.0208	0.0210	0.0216
3	0.1145	0.0356	0.0038	0.0005	0.0035	0.0035	0.0034	0.0029	0.0035	0.0035	0.0034	0.0029
4	0.2576	0.1429	0.0826	0.0704	0.0126	0.0126	0.0127	0.0129	7.7e-5	9.2e-5	0.0002	0.0008
5	0.1213	0.0495	0.0146	0.0037	0.0292	0.0293	0.0293	0.0293	0.0033	0.0125	0.0225	0.0249
6	0.0192	0.0136	0.0119	0.0150	0.1303	0.0086	0.0105	0.0298	0.0013	0.0011	0.0004	6.4e-5
7	0.0249	0.0250	0.0251	0.0250	0.0397	0396	0.0391	0.0396	0.0035	0.0002	0.0016	3.7e-5
8	0.0238	0.6972	0.6274	0.0912	0.0397	0.0256	0.0257	0.0256	0.1817	0.1859	0.2081	0.2395
9	0.0928	0.0044	0.0021	1.7e-5	0.0404	0.0404	0.0404	0.3356	0.3356	0.2105	0.2216	0.3825
10	0.4708	1.7117	0.7312	0.1454	0.2421	0.2421	0.2421	0.2421	0.2967	0.3015	0.3219	0.3351

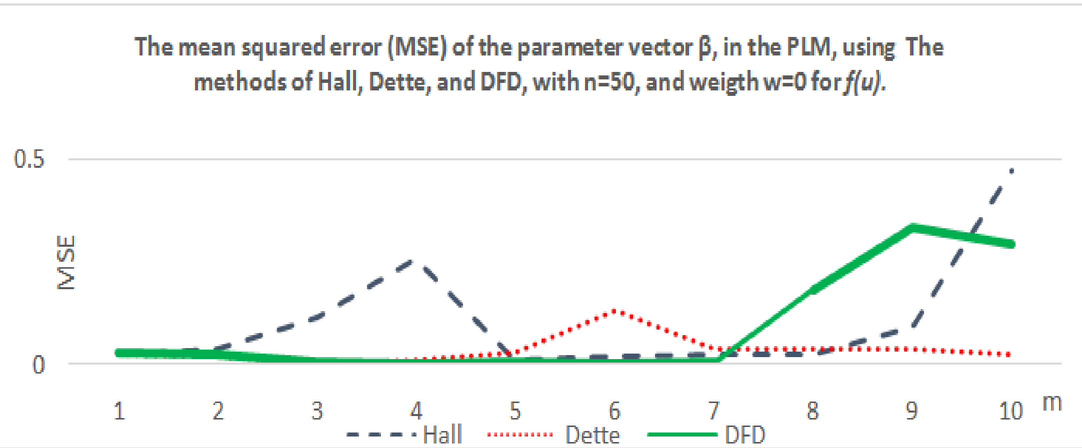

Figure 1: The Mean Squared Error (MSE) of the Parameter Vector Β, in the PLM, Using the Methods of Hall, Dette, and DFD, With N = 50 and Weight W = 0 For F (U).

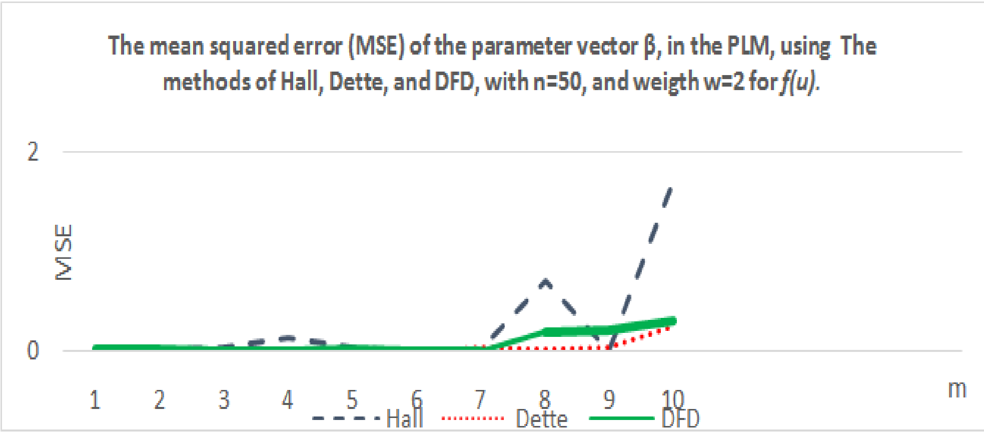

Figure 2: The Mean Squared Error (MSE) of the Parameter Vector Β, in the PLM, Using the Methods of Hall, Dette, and DFD, with N = 50 and Weight W = 2 for F (U).

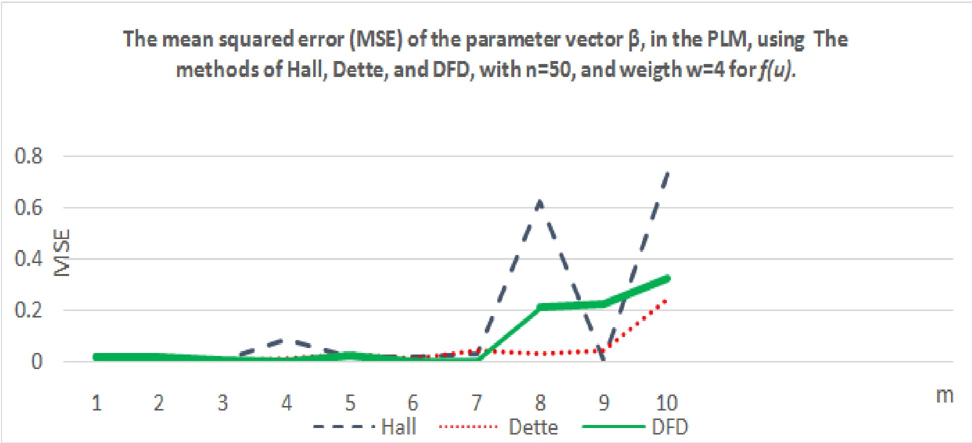

Figure 3: The Mean Squared Error (MSE) of the Parameter Vector Β, in the PLM, Using the Methods of Hall, Dette, and DFD, with N = 50 and Weigth W = 4 for F (U).

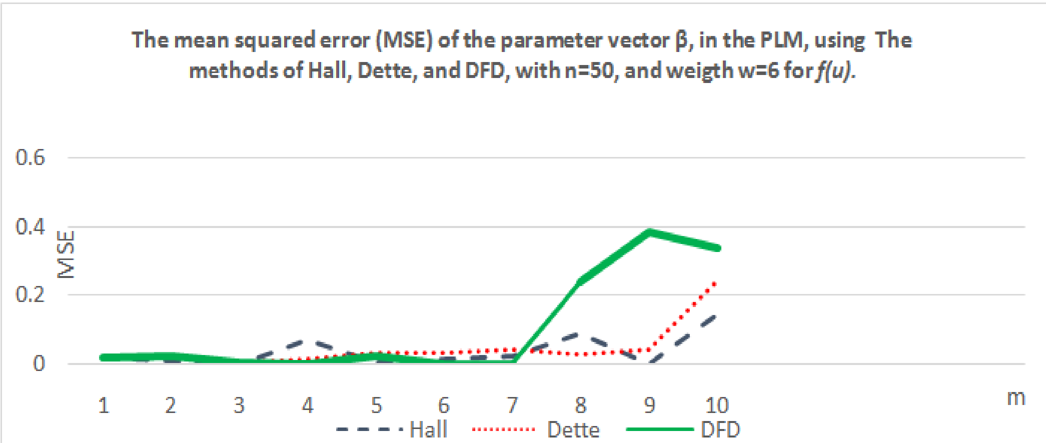

Figure 4: The Mean Squared Error (MSE) of the Parameter Vector Β, In The PLM, Using the Methods of Hall, Dette, and DFD, with N = 50 and Weight W = 6 for F (U).

Table 5: The Mean Squared error (MSE) of the Parameter Vector β, in the PLM, using $\hat{\beta}(Hall)$, $\hat{\beta}(Dette)$ and $\hat{\beta}(DFD)$. The Results are Recorded for Difference Sequences of Order $1 \leq m \leq 10$, Four Weights (0, 2, 4, 6) for the Nonparametric Function f(u), and Sample Size n = 200

Difference Order (M)	MSE of $\hat{\beta}(Hall)$				MSE of $\hat{\beta}(Dette)$				MSE of $\hat{\beta}(DFD)$			
	f_0	f_2	f_4	f_6	f_0	f_2	f_4	f_6	f_0	f_2	f_4	f_6
1	0.0085	0.0074	0.0079	0.0085	0.0085	0.0074	0.0079	0.0085	0.0085	0.0074	0.0079	0.0085
2	0.0034	0.0039	0.0043	0.0048	0.0200	0.0200	0.0200	0.0200	0.0200	0.0200	0.0200	0.0200
3	0.0684	0.0684	0.0684	0.0684	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625
4	0.0328	0.0436	0.0555	0.0684	0.0248	0.0248	0.0248	0.0248	0.0028	0.0028	0.0028	0.0029
5	0.0004	4.7e-5	0.0010	0.0032	0.0017	0.0017	0.0017	0.0017	0.0008	0.0007	0.0005	0.0004
6	0.0004	0.0036	0.0029	0.0023	0.0054	0.0097	0.0152	0.0152	0.0012	0.0012	0.0012	0.0013
7	0.0372	0.0052	0.0053	0.0052	0.0067	0.0067	0.0067	0.0067	0.0015	0.0012	0.0010	0.0008
8	0.0372	0.0311	0.0258	0.0218	0.0053	0.0053	0.0053	0.0053	0.0009	0.0010	0.0010	0.0010
9	0.0012	0.0001	0.0028	0.0087	0.0054	0.0054	0.0054	0.0054	0.0024	0.0015	0.0008	0.0004
10	0.0197	0.0282	0.0370	0.0445	0.0029	0.0029	0.0029	0.0029	0.0044	0.0045	0.0045	0.0046

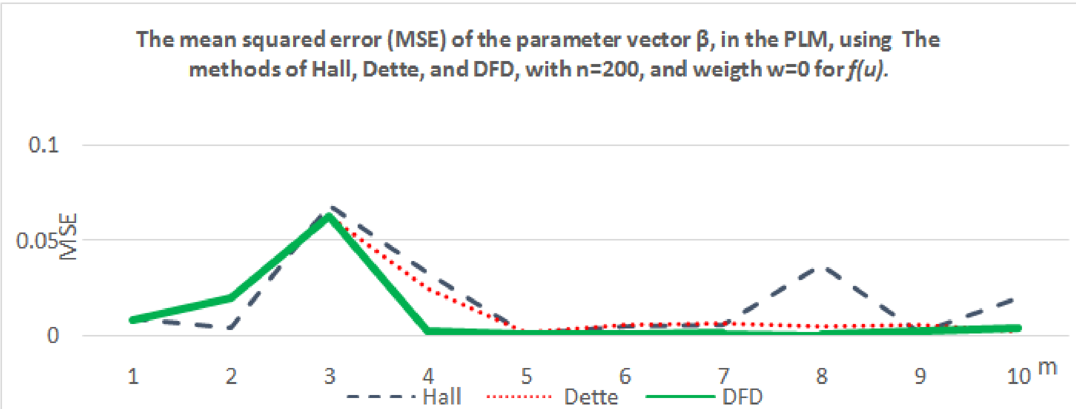

Figure 5: The Mean Squared Error (MSE) of the Parameter Vector Β, In The PLM, Using the Methods of Hall, Dette, and DFD, with N = 200, and Weight W = 0 for F (U).

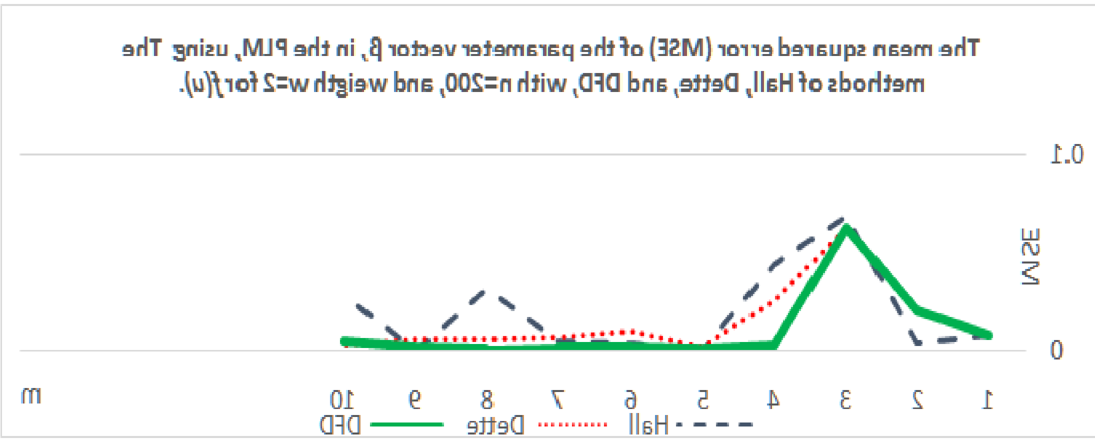

Figure 6: The Mean Squared Error (MSE) of the Parameter Vector Β, in the PLM, Using the Methods of Hall, Dette, and DFD, with N = 200, and Weight W = 2 for F (U).

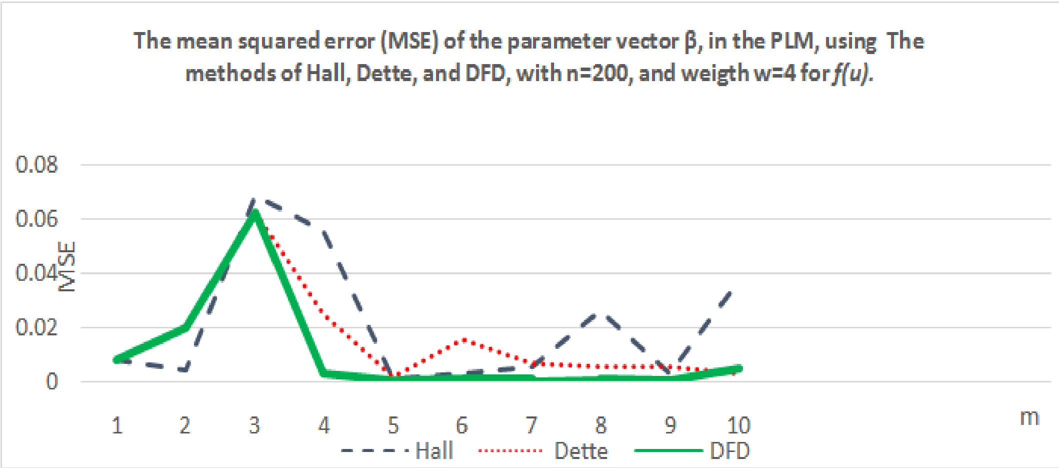

Figure 7: The Mean Squared Error (MSE) of the Parameter Vector Β, in the PLM, Using the Methods of Hall, Dette, and DFD, with N = 200 and Weigth W = 4 For F (U).

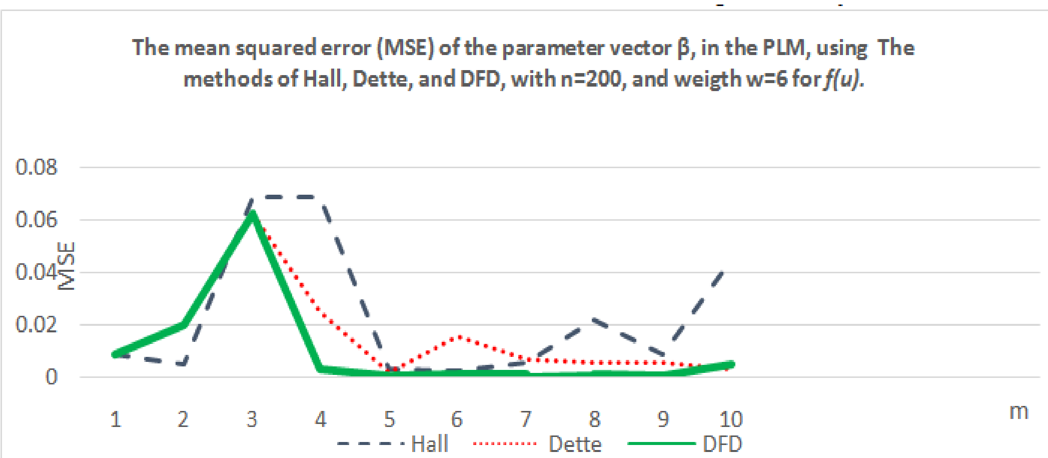

Figure 8: The Mean Squared Error (MSE) of the Parameter Vector Β, in the PLM, Using the Methods of Hall, Dette, and DFD, with N = 200, and Weight W = 6 for F (U).

Table 6: The Estimated Variance $\hat{\sigma}^2$ for the Three Fitted PLM Using Different Parametric Estimation Methods $\hat{\beta}(Hall)$, $\hat{\beta}(Dette)$ and $\hat{\beta}(DFD)$, and Using the Nonparametric Function *F (U)* with Four Different Weights (0, 2, 4, 6). The Results are Recorded for Difference Sequences of Order 1 ≤ M ≤ 10, Four Weights (0, 2, 4, 6) for the Nonparametric Function *F (U)*, and Sample Size N = 50

Difference Order (M)	$\hat{\sigma}^2(diffl)$ Using $\hat{\beta}(Hall)$				$\hat{\sigma}^2(diffl)$ Using $\hat{\beta}(Dette)$				$\hat{\sigma}^2(diffl)$ Using $\hat{\beta}(DFD)$			
	f_0	f_2	f_4	f_6	f_0	f_2		f_6	f_0	f_2	f_4	f_6
1	4.12	3.13	2.28	1.59	4.12	3.13	2.28	1.59	4.12	3.13	2.28	1.59
2	8.30	6.08	4.27	2.89	1.18	1.73	01.14	1.04	1.18	1.73	1.14	1.04
3	24.5	17.4	12.4	9.51	5.0e-	8.9e-	05.0e-	8.9e-	5.0e-	8.9e-	5.0e-	8.9e-
4	19.2	13.5	10.1	9.39	11.98	11.98	11.99	12.02	2.21	2.19	2.06	1.78
5	14.7	8.75	5.05	3.34	13.78	13.78	13.79	13.78	0.90	1.28	1.59	1.65
6	2.99	0.58	0.18	1.05	31.87	16.16	8.05	5.81	0.14	0.14	0.12	0.09
7	13.2	5.99	2.80	2.28	12.60	12.60	12.6	12.60	0.03	0.07	0.09	0.06
8	2.16	0.09	0.19	4.39	12.60	9.78	9.78	9.78	0.08	0.08	0.09	0.12
9	12.2	4.07	1.75	2.50	6.24	6.24	6.24	6.24	0.13	0.06	0.07	0.16
10	0.74	0.19	0.26	9.59	2.98	2.98	2.98	2.98	0.41	0.42	0.46	0.49

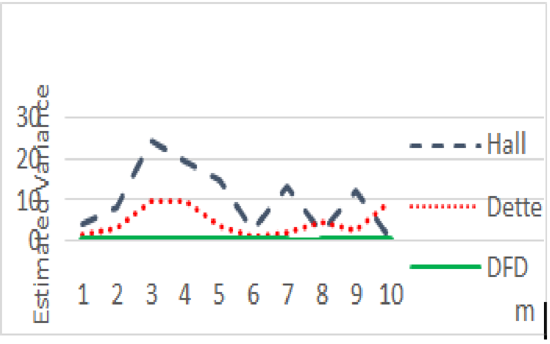

Figure (9–1): w = 0

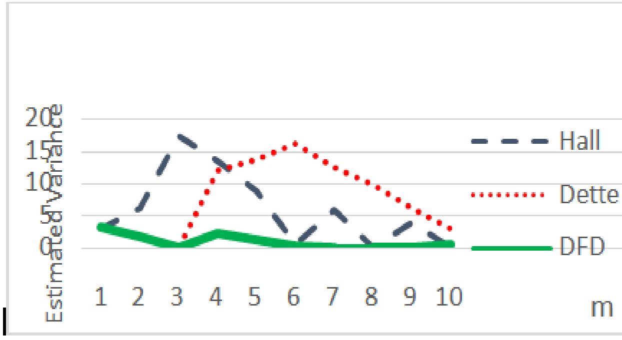

Figure (9–2): w = 2

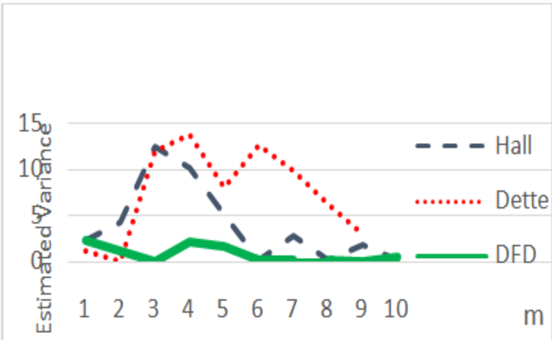

Figure (9–3): w = 4

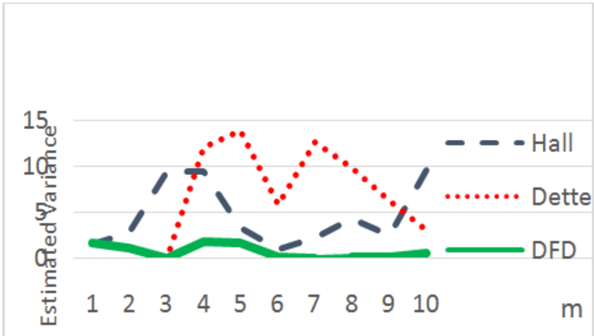

Figure (9-4): w = 6

Figure 9: The Estimated Variance $\hat{\sigma}^2$ for the three Fitted PLM using Different Parametric Estimation Methods, $\hat{\beta}(Dette)$ and $\hat{\beta}(DFD)$, and using the Nonparametric Function *f(u)* with Four Different Weights (0, 2, 4, 6). The Results are Recorded for Difference Sequences of Order 1≤m ≤10, Four Weights (0, 2, 4, 6) for the Nonparametric Function *f(u)*, and Sample Size n=50.

Table 7: The Estimated Variance $\hat{\sigma}^2$ for the three Fitted PLM Using Different Parametric Estimation Methods $\hat{\beta}(Hall)$, $\hat{\beta}(Dette)$ and $\hat{\beta}(DFD)$, and Using the Nonparametric Function *F (U)* with Four Different Weights (0, 2, 4, 6). The Results are Recorded for Difference Sequences of Order 1 ≤ M ≤ 10, Four Weights (0, 2, 4, 6) for the Nonparametric Function F (U), and Sample Size N = 200

Difference Order (M)	$\hat{\sigma}^2(diffl)$ Using $\hat{\beta}(Hall)$				$\hat{\sigma}^2(diffl)$ Using $\hat{\beta}(Dette)$				$\hat{\sigma}^2(diffl)$ Using $\hat{\beta}(DFD)$			
	f_0	f_2	f_4	f_6	f_0	f_2	f_4	f_6	f_0	f_2	f_4	f_6
1	0.897	1.021	0.958	0.897	0.897	1.021	0.958	0.897	0.897	1.021	0.958	0.897
2	1.76	1.63	1.50	0.38	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
3	0.31	0.31	0.31	0.31	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.039
4	0.65	0.52	0.40	0.30	0.911	0.911	0.911	0.911	0.045	0.045	0.045	0.045
5	2.36	2.05	1.76	0.49	1.818	1.818	1.818	1.818	0.156	0.175	0.193	0.212
6	1.12	0.87	0.66	0.49	1.333	0.917	0.584	0.334	0.046	0.046	0.046	0.046
7	2.47	2.05	1.66	1.34	4.051	4.051	4.05	4.051	0.021	0.024	0.028	0.031
8	5.44	4.71	4.08	3.59	4.485	4.485	4.485	4.485	0.009	0.009	0.009	0.009
9	2.58	2.03	1.57	1.19	4.260	4.260	4.260	4.260	0.003	0.005	0.007	0.009
10	0.48	0.25	0.10	0.04	3.613	3.613	3.613	3.613	0.001	0.001	0.001	0.001

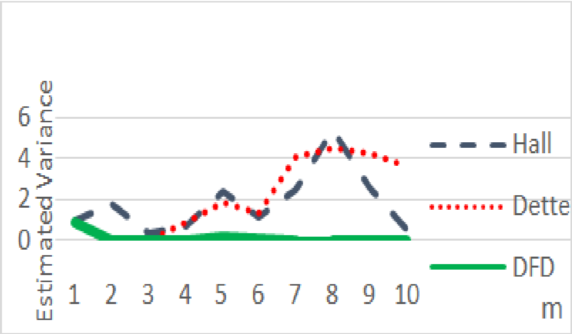

Figure (10–1): w = 0

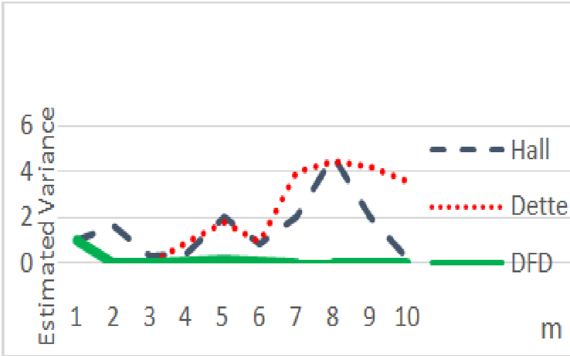

Figure (10–2): w = 2

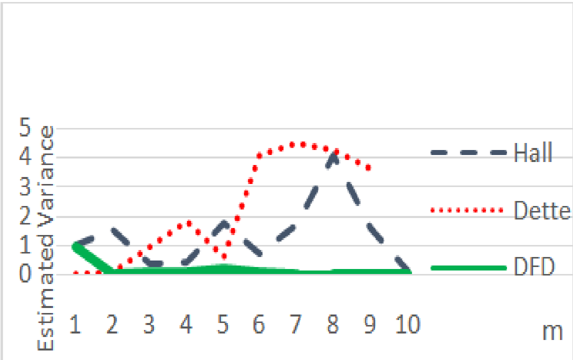

Figure (10–3) w = 4

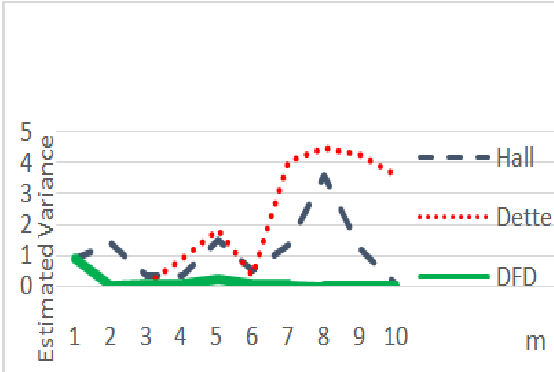

Figure (10–4): w = 6

Figure 10: The Estimated Variance $\hat{\sigma}^2$ for the three Fitted PLM using Different Parametric Estimation Methods $\hat{\beta}(Hall)$, $\hat{\beta}(Dette)$ and $\hat{\beta}(DFD)$, and using the Nonparametric Function *f (u)* with Four Different Weights (0, 2, 4, 6). The Results are Recorded for Difference Sequences of order $1 \leq m \leq 10$, Four Weights (0, 2, 4, 6) for the Nonparametric Function F (U), and Sample Size n = 200.

Simulation Results for the Residual Variance Estimator

The residual variance estimator $\hat{\sigma}^2(diff)$ is computed for each method, Hall, Dette, and DFD. By comparing $\hat{\sigma}^2(diff)$ for the three methods, it is found that best results are obtained, in the form of less values of $\hat{\sigma}^2(diff)$, for DFD method. Tables (6), and (7), and also Figures (9), and (10) show that DFD method has less values of $\hat{\sigma}^2(diff)$compared with the other two methods, Hall and Dette, for all sample sizes and all difference sequences ($1 \leq m \leq 10$).

CONCLUSIONS

This paper considers estimating the partially linear model PLM using the difference method. The proposed difference sequence which is called DFD method is applied when estimating both the parameter vector β and the residual variance of the fitted PLM. The results of DFD method are compared with those of Hall method (Hall et al., 1990), and Dette method (Dette et al., 1998). Best results are obtained in favor of the DFD, in the form of less values of the MSE of the estimated β, and less values of the residual variance of the fitted PLM. In the future, this method will be developed for other semiparametric regression models.

Appendix (A)

Proof of Theorem 1

$$\hat{\beta}(diff) = \left(\tilde{X}^T \tilde{X}\right)^{-1} \tilde{X}^T \tilde{Y} = \left(\tilde{X}^T \tilde{X}\right)^{-1} \tilde{X}^T \left[\tilde{X}^T \beta + \tilde{\varepsilon}\right]$$
$$= \beta + \left(\tilde{X}^T \tilde{X}\right)^{-1} \tilde{X}^T \tilde{\varepsilon}$$

$$\because E\left(\left(\tilde{X}^T \tilde{X}\right)^{-1} \tilde{X}^T \tilde{\varepsilon}\right) = 0$$

$$\therefore E\left(\hat{\beta}(diff)\right) = \beta$$

$Var\left(\hat{\beta}(diff)\right) = Var\left(\left(\tilde{X}^T \tilde{X}\right)^{-1} \tilde{X}^T \tilde{\varepsilon}\right) \sim \sigma^2\left(1 + 2\sum_{k=1}^{m} \delta_k^2\right)\left(X^T X\right)^{-1}$,(See remark 5 in Wang et al., (2011)), where i, for k=1,2,…,m.

$$\hat{\sigma}^2(DFD) = \frac{1}{n-m-p}\sum_{i=1}^{n-m}\left(\tilde{Y}_i - \tilde{X}_i^T \hat{\beta}\right)^2$$
$$= \frac{1}{n-m-p}\sum_{i=1}^{n-m}\left(\tilde{X}_i^T\left(\beta - \hat{\beta}\right) + \tilde{\varepsilon}_i\right)^2$$
$$= \frac{1}{n-m-p}\sum_{i=1}^{n-m}\tilde{\varepsilon}_i^2 + O_p\left(\frac{1}{n}\right)$$

by $\sqrt{n}$ -consistency of $\hat{\beta}$.

$$\therefore \hat{\sigma}^2(DFD) = \frac{1}{n-m-p}\sum_{i=1}^{n-m}\sum_{j=0}^{m} d_j^2 \tilde{\varepsilon}_i^2 + O_p\left(\frac{1}{n}\right)$$
$$= \frac{1}{n-m-p}\sum_{i=1}^{n-m}\sum_{j=0}^{m} d_j^2 \tilde{\varepsilon}_i^2 + O_p\left(\frac{1}{n}\right)$$
$$= \frac{n-m}{n-m-p}\sum_{j=0}^{m} d_j^2 \varepsilon^2 + O_p\left(\frac{1}{n}\right)$$
$$= \frac{n-m}{n-m-p}\varepsilon^2 + O_p\left(\frac{1}{n}\right)$$

$$E\left[\hat{\sigma}^2(DFD)\right] = \frac{n-m}{n-m-p}\sigma_\varepsilon^2 + O\left(\frac{1}{n}\right)$$

$$\hat{\sigma}^2(DFD) \xrightarrow{p} \sigma_\varepsilon^2$$

REFERENCES

1. *Buja, A., Hastie, T. J., and Tibshirani, R. j. (1989). "Linear smoother and additive models (with discussion), Annals of Statistics, 17, 453–555.*
2. *Wang, L., Brown, L. D., & Cai, T. T. (2011). A difference based approach to the semi parametric partial linear*

model. Electronic Journal of Statistics, 5, 619–641.

3. *Dette, H., Munk, A., and Wagner, T. (1998). "Estimating the variance in nonparametric regression-what is a reasonable choice?" Journal of the Royal Statistical Society,B60, 751–764.*
4. *Engle, R., Granger, C., Rice, J., and Weiss, A. (1986). "Nonparametric estimates of the relation between weather and electricity sales", Journal of the American Statistical Association, 81, 310–320.*
5. *Haggag, M. M., (2019). "A New Difference Method for Variance Estimation in Non-Parametric Regression Models", Advances and Application in Statistics, vol. 56.*
6. *Hall, P., Kay, J.W., and Titterington, D. M. (1990). "Asymptotically optimal difference-based estimation of variance in nonparametric regression", Biometrica, 77, 521–528.*
7. *Hardle, W., Liang, H., and Gao, J. (2000). Partially linear models, Springer Verlag.*
8. *Hardle, W., Muller, M., Sperlich, S., and Werwatz, A. (2004). Nonparametric and Semiparametric Modeling: An Introduction, Springer, New York.*
9. *Hastie, T. J., and Tibshirani, R. j. (1990). Generalized additive models, Vol. 43 of Monographs on statistics and applied probability, Chapman and Hall, London.*
10. *Heckman, N. E., (1986). "Spline smoothing in a partly linear model", Journal of the Royal Statistical Society, Series B., 48, 244–248.*
11. *Holland, A. (2017). "Penalized spline estimation in the partially linear model", Journal of Multivariate Analysis, 153, 211–235.*
12. *Liu, Q. (2010). "Asymptotic Theory for Difference-based Estimator of Partially Linear Models", Journal of the Japan Statistical Society, 39, 393–406.*
13. *Muller, M. (2001). "Estimation and testing in generalized partial linear models- a comparative study, Statistics and comuting, 11, 299–309.*
14. *Opsomer, and Ruppert(1999). "A root-n consistent backfitting estimator for semiparametric additive modelling".*
15. *Robinson, P. M. (1988). "Root n-consistent semiparametric regression", Econometrica, 56, 931–954.*
16. *Ruppert, D. Wand M. P., Carroll, R. J. (2003). Semiparametric regression, New York, Cambridge University press.*
17. *Severini, T.; and Wong, W (1992). "Generalized profile likelihood and condtional parametric models", Annals of Statistics, 20, 1768–1802.*
18. *Silverman, B.W. (1985). "Some Aspects of the Spline Smoothing Approach to Non-Parametric Regression Curve Fitting", Journal of the Royal Statistical Society. Series B Vol. 47, No. 1, 1–52.*
19. *Speckman, P. (1988). "Kernel smoothing in partial linear models", Journal of the Royal Statistical Society, Series B., 50, 413–436.*
20. *Wahba, G. (1990). Spline models for observational data, Society for Industrial and applied mathematics (Siam), Philadelphia, Pennsylvania.*
21. *Wang, L., Brown, L. D., and Cai, T. (2011). "A difference based approach to the semiparametric partial linear*

model", Electronic Journal of Statistics, 5, 619–641.

22. *Yatchew, A. (1997). "A elementary estimator of the partial linear model", Eeconomic Letters, 57, 135–143.*
23. *Yatchew, A. (1997). "An Elementary Estimator of the Partial Linear Model, Economics Letters, 57, 135–43.*
24. *Yatchew, A. (2003).Semiparametric regression for the applied econometrician, New York, Cambridge University press.*
25. *Zhou, S.; Chen, X.; Wolfe, D.A. (1998). "Local asymptotics for regression splines and confidence regions", Annals of Statistics, 26, 1760–1782.*
26. *Zhou, Y., Cheng, Y., Dai, W., and Tong, T. (2018). "Optimal difference-based estimation for partially linear models", Computational Statistics, 33, 863–885.*

Dr. Magda Mohammed Haggag Associate Professor in Applied Statistics, Department of Statistics, Mathematics, and Insurance, Facuty of commerce, Damanhour University, Head of the Department of Statistics, Mathematics, and Insurance – Department of Statistics, Mathematics, and Insurance, Faculty of commerce, Damanhour University, Egypt. Vice-Dean of Graduated Studies and Researches, Faculty of commerce, Damanhour University,

General skills in statistics and computing. Specific expertise and interests in

Statistics: linear regression, parametric and nonparametric regression, semiparametric regression, regularized regression models, multivariate statistical analysis, fuzzy regression models, and big data analytics. Skills in statistical analysis using Statistical Packages :SAS – SPSS – Minitab – S.Plus – R - XploRe- Gauss – Statgraphics , etc..

Mathematics: system of linear equations, analysis of fuzzy data, functional data analysis. Mathematical Analysis using Packages: Matlab – Mathcad.