

نماذج الانحدار شبه المعلمية دراسة مقارنة

علاء جابر طراد **

أ.م.د. رعد فاضل حسن *

المستخلص:

أحدى الغايات الرئيسية من دراسة اي مشكلة او ظاهرة (اقتصادية ، اجتماعية ، علمية.....) هو ايجاد المعادلة الرئيسية التي تمثلها والتي ستكون المدخل الرئيس لفهم هذه الظاهرة وتحديد معالمها الرئيسية ، وهو ما يعرف في علم الاحصاء بنمذجة العلاقة بين الظواهر ، وقد تم في الجانب النظري من هذا البحث توضيح مبسط لتقديرات بعض النماذج المعلمية مثل (أنموذج الانحدار الخطي (البسيط ، العام ، الموسع) ، وأنموذج الانحدار اللاخطي) ، وبعض مقدرات الانحدار اللامعلمي ، ومقدرات نماذج الانحدار شبه المعلمية مع التركيز اكثر على الانموذجين (الانحدار الخطي الجزئي PLM والانحدار الخطي الجزئي الموسع) ، وفي الجانب العملي تم اجراء تطبيق عملي على بيانات شركة خطوط الانابيب النفطية باجراء ثلاث مراحل حيث تم في المرحلة الاولى بناء اربع نماذج معلمية وثمانية وعشرين أنموذجاً شبه معلمياً ، وفي المرحلة الثانية تم اجراء المقارنة فيما بين هذه النماذج بثلاثة اساليب وعلى اساس ثلاثة معايير (AIC, BIC, R^2) لكل اسلوب مقارنة ، وفي المرحلة الثالثة تم تحسين درجة تمثيل نماذج الانحدار الخطي الجزئي بتغيير قيمة عرض الحزمة .

Abstract:

One of the major goals of studying any kind of problems or Phenomenon (Economics, Social , Scientific..) is to find there major function which represent it , it will be the main internet to understand it and specified the main parameter , in theoretical part of this research we represent simply for estimator of the parametric models like (linear regression(simple , general , generalized) and nonlinear regression) and some of the nonparametric regression estimator and semiparametric regression model estimator with focus on (partial linear model and generalized partial linear model's estimators) , in practical part we applied practically on (Oil Pipeline Company's data) by made three steps to analyze these data where we've done in a first step is construct four parametric and twenty eight semiparametric models , in second step we made a compare between these models by three methods and using three Standards (AIC, BIC, R^2) for each method ,in third step we increase the Coefficient of determination (R^2) and adjusted R^2 for the partial linear models by changing the value of the bandwidth.

1. المقدمة

أحدى الغايات الرئيسية من دراسة اي مشكلة او ظاهرة (اقتصادية ، اجتماعية ، علمية.....) هو ايجاد النموذج الرئيسي الذي يمثلها والذي سيكون المدخل الرئيسي لفهم هذه الظاهرة وتحديد معالمها الرئيسية ، وهو ما يعرف في علم الاحصاء بنمذجة الظواهر ، ان هذه العملية قد تطورت بتطور علم الاحصاء نفسه وان

* الجامعة المستنصرية / كلية الادارة والاقتصاد .

** باحث .

مقبول للنشر بتاريخ 2013/12/3

مستل من رسالة ماجستير

كانت لفترة طويلة تعتمد على نوع واحد هو الانحدار المعلمي PR بشقيه الخطي ($Linear$) واللاخطي ($nonlinear$) حيث تم تطوير كل ما يتعلق به من تقديرات وفرضيات واختبارات منذ ايام العالم فيشر وحتى الان علما بان جميع الحالات التي تمت نمذجتها كان يتم افتراض توزيعها توزيعا طبيعيا وان تكون حالات قابلة للقياس ، حتى جاء الباحثان $Nelder - Wderburn$ ⁽¹⁹⁾ ليتم تحديد الحالة الموسعة من الانحدار العام وهي الانحدار الخطي الموسع GLM والتي فتحت الباب على مصراعيه لنمذجة كافة الظواهر والمشاكل الغير القابلة للقياس او الوصفية ، ان الميزة الايجابية التي يتمتع بها الاسلوب المعلمي هي بحد ذاتها ايضا ميزة سلبية من حيث تحديد معالم دالة الانحدار والتي ستكون معادلة واحدة فقط لا يمكن تغييرها الا بعد (اضافة أو حذف) متغير منها او اضافة بيانات جديدة وبالتالي فان تحسين مقياس درجة تمثيل النموذج لبيانات اي ظاهرة وهو معيار (معامل التحديد R^2) قد يكون مستحيلا دون اتخاذ تلك الاجراءات ، وبالتالي فان الباحثين يكونون مجبرين على قبول هذه المعادلة عندما لا يكون هناك امكانية لتوفير الحلول المناسبة كالتي ذكرناها ، وفي مرحلة ثانية من نمذجة الظواهر ظهر الى العلن فكر احصائي جديد والذي لا يعتمد على القيود الصارمة المفروضة في الانحدار المعلمي والذي تمثل بتمهيد البيانات باستخدام الاوزان الثابتة ومن ثم تطور باستخدام واحدة من اشهر الدوال الوزنية وهي الدالة اللبية ($kernel function$) والتي باستخدامها تم تمهيد البيانات على اساس الاوزان غير الثابتة والتي تحدد وزنا مختلفا على اعتبار القرب والبعد بين قيم المتغير المستقل عن نقطة محددة سلفا ، وباستخدام الدالة الاخيرة تشكلت مجموعة من المقدرات الاحصائية اللامعلمية مكونة ما يعرف باسم مقدرات الانحدار اللامعلمي NPR من قبيل المقدرات (ناداريا - واتسون ، برستل جاو ، المقدر اللبي ، الشرائح المقطعية ، مهده الانحدار الخطي الموضوعي.....) ، ان التطور الهائل في مجال الحاسبات الالكترونية جعل الباحثين يطبقون طرقا جديدة للوصول الى النموذج الافضل المعبر عن الظاهرة المدروسة ايا كانت منطلقين من مبدأ ان المعادلة المستخرجة تحاول تحقيق تقارب بين قيم منحني y التقديرية و منحني Y الحقيقية فبدأ التوجه نحو مبدأ: تقدير منحني البيانات بدون الحاجة الى نموذج او معالم $Curve Fitting of data without model$ وباستخدام الدوال اللبية وتمهيد البيانات ($Smoothing Data$) وتحديد قيم المسافات البينية امكن تقريبا من الحصول على مالاتهاية من المنحنيات التي تقترب او تكاد من منحني بيانات المتغير المعتمد ، وهو ما اقره الباحثين على اللجوء الى هذا الاسلوب لدراسة الظواهر المختلفة ، وعلى الرغم من ذلك فان السلبيات التي تصاحب عملية التمهيد ليست بالهينة فان طرق التمهيد المختلفة تصبح طرقا صعبة جدا في حالة المتغيرات الكثيرة لوجود مشكلة البعدية ($curs of dimension$) من ناحية ومن ناحية اخرى فان فقدان ميزة الحصول على دالة معبرة عن الظاهرة ولو بالحدود الضيقة تعتبر من سلبيات هذه الطريقة ، وفي المرحلة الثالثة عند منتصف الثمانينات بدأ يتشكل الجيل الثالث من نمذجة الظواهر وهو الانحدار شبه المعلمي SPR حيث انه يمثل من الناحية العملية الحالة الوسيطة بين أنموذج الانحدار المعلمي PR والانحدار اللامعلمي NPR ، وتأتي الميزة الكبيرة لهذا الأنموذج في كونه يحوي على كل مميزات الايجابية التي يتضمنها الأنموذجين السابقين ولوضوح التفاعل بين مكوناته العملية واللامعلمية والتي لاقت قبولا واسعا في الدراسات الاقتصادية والاجتماعية والعلمية الحديثة وذلك بسبب حله لمشكلة السلوك الغير مفهوم لبعض المتغيرات الداخلة في الدراسة من جهة و للمرونة العالية التي يتمتع بها هذا النوع من الانحدار من جهة اخرى ، وقد تم اقتراحه من منطلق فكرة النماذج التجميعية حيث تم الدمج بين المكونين اللامعلمي والمعلمي بأنموذج شبه المعلمي الذي احتوى على القيود الصارمة في جزئه المعلمي ومرونته الكاملة في جزئه اللامعلمي وتطرق البحث الى ثلاثة مفاصل رئيسية حيث احتوى الاول التطرق الى النماذج المعلمية واللامعلمية وشبه المعلمية مع التطرق الى اساليب تقديرها والثاني هو اجراء تطبيق عملي على بيانات شركة خطوط الانابيب النفطية والثالث هو تحديد اهم الاستنتاجات والتوصيات .

2- الانحدار المعلمي

من أهم مجالات دراسة العلوم هو ايجاد وصف للعلاقة التي تجمع ما بين مجموعة من المتغيرات المختلفة التي تمثل ظاهرة من ظواهر الحياة المتعددة وهذا الوصف يتمثل بأنموذج معين يربط بين متغير عشوائي معتمد ومتغيرات عشوائية مستقلة ترتبط باوزان $Weights$ او معالم $Parameters$ وهذا النوع من النماذج يسمى بأنموذج الانحدار المعلمي $Parametric Regression Model$ ، يمكن التعريف بنوعين من نماذج الانحدار المعلمي :-

(1-2) الانحدار المعلمي الخطي

وهو من الاساليب الاحصائية التي تستخدم في قياس العلاقة بين متغيرين أو عدة متغيرات حيث سيكون أحدها هو المتغير المعتمد Y والمتغير الاخر أو المتغيرات الاخرى ستكون المتغيرات المستقلة ، وبصورة عامة يوصف الأنموذج بالصيغة :-

$$Y = f(X_i, \beta) + e \dots\dots\dots(1)$$

أذ أن Y, X تمثل متغيرات عشوائية و $f(X_i, \beta)$ دالة خطية للمتغيرات المستقلة X_i ولمعالم غير معروفة (β) ، ويدعى النموذج (1) بأنموذج الانحدار الخطي العام GLM والذي عند كتابته بدلالة المصفوفات سيكون :-

$$\underline{Y} = \underline{X}\underline{\beta} + \underline{e} \dots\dots\dots(2)$$

والذي يسمى بالانحدار الخطي العام GLM ، وبافتراض أن $E\underline{e} = 0$ و $Var(\underline{e}) = \sigma^2 I$ والذي يشير الى تجانس التباين وأن $\underline{Y}$ يمثل متجه المتغير المعتمد ، ان تقدير معالم GLM (أنموذج الانحدار العام) يكون بأسلوب الامكان الاعظم أو بأسلوب المربعات الصغرى والتي كلاهما تتوصل الى صيغة واحدة لتقدير المعالم

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

$$\hat{\sigma}^2 = \frac{1}{n-p} (Y^T Y - \hat{\beta}^T X^T Y) \dots\dots\dots(3)$$

أن الحالة الموسعة من النماذج الخطية والتي تمكن من نمذجة البيانات والتي لا تحقق شرط التوزيع الطبيعي والبيانات غير الكمية تسمى بالانحدار الخطي الموسع GLM_Z ، وقد اورد ⁽⁹⁾ Lindsey عدة امثلة مشتركة بكونها حالات تكون نمذجتها بأسلوب GLM_Z ، وان الاساس في هذه الامثلة هو اشتراكها في عائلة التوزيعات الاسية *Exponential Distribution Family* أو عائلة التشتت الاسية *Exponential Dispersion Family* ، ومن مميزات الأنموذج الخطي الموسع (GLM_Z) انه يحرر البيانات من شرط الخطية في علاقة Y مع X_s وكذلك يحرر الباحثين من قيود التحليل المفروضة على دالة الانحدار وبهذا يمكن نمذجة العديد من التوزيعات {التوزيع الطبيعي ، توزيع ثنائي الحدين ، توزيع بواسون ، معكوس الطبيعي ، معكوس ثنائي الحدين ، توزيع كاما.....} ، ان الصيغة العامة لدوال العائلة الاسية تكون :-

$$f(y_i, \theta_i) = \exp\{a(y)b(\theta) - c(\theta) + d(y)\} \dots\dots\dots(4)$$

وعندما $a(y) = y$ فان الدالة (4) يمكن كتابتها بالصيغة

$$f(y_i, \theta_i) = \exp\{yb(\theta) - c(\theta) + d(y)\} \dots\dots\dots(5)$$

وعندها سيطلق على هذه الصيغة بالصيغة القانونية *(Canonical Form)* والدالة $b(\theta)$ تعرف بالمعلمة الطبيعية *(Natural Parameter)* ، والحالة العامة من الصيغة القانونية تكون باضافة الدالة $\{r(\varphi)\}$ ، والتي تكون بدلالة معلمة التشتت الى الصيغة اعلاه لتكون الصيغة كالآتي:-

$$f(y, \theta, \varphi) = \exp[(y\theta - b(\theta))/r(\varphi) + d(y, \varphi)] \dots\dots\dots(6)$$

ان اي دالة استجابة قابلة للتحويل الى العائلة الاسية يمكن من خلالها استخراج الدوال الخاصة بها ومنها يمكن ان نجد ان الوسط الحسابي والتباين توصف بالصيغ :-

$$E(Y) = \mu = b'(\theta) \dots\dots\dots(7)$$

$$Var(Y) = \sigma_y^2 = b''(\theta) r(\varphi)$$

ان الصيغة العامة لأنموذج الانحدار الخطي الموسع (GLM_Z) تكون

$$E(Y) = G(X^T \beta) \dots\dots\dots(8)$$

ويتكون هذا الأنموذج من ثلاثة اقسام رئيسية :-

- المركبة العشوائية *(Random component)* والذي يتمثل بالمتغير العشوائي y الذي يتوزع بتوزيع ينتمي الى عائلة التوزيعات الاسية.
- المركبة النظامية *(Systematic component)* والذي يتمثل بصيغة الوسط الحسابي الذي يكون على شكل تركيبة خطية تعرف بالتنبؤ الخطي *linear predictor* بدلالة المتغيرات التوضيحية
- دالة الربط *The Link Function G(.)* وهي الدالة التي تربط بين القسمين الاول والثاني وهي دالة رتيبة (تصاعدية أو تنازلية) خطية او لاخطية وصيغتها تكون :-

$$\eta_i = G(\mu) = G(X^T \beta)$$

وهناك عدة أنواع من دوال الربط لها علاقة باستجابة المتغير وتوزيعه والجدول (2-1) يوضح بعض توزيعات العائلة الاسية والمعالم الطبيعية ودوال الربط الخاصة بها :-

ولاجل تقدير معالم أنموذج الانحدار الخطي الموسع (GLM_z) فقد حدد كل من الباحثين $Nelder - Wedrburn^{(19)}$ بأن غالبية الحالات التي يمكن في نمذجتها بطريقة GLM_z ستتبع نفس الاسلوب في تقدير معالمها بطريقة الامكان الاعظم ، أن تقدير معالم الأنموذج ستعتمد على التوسيع الخطي

الجدول رقم (1) توزيعات العائلة الاسية - دالة الربط - النوع		
النوع	دالة الربط	التوزيع
Identity	$\eta_i = G(\mu) = \mu$	الطبيعي
Log	$\eta_i = G(\mu) = \text{Log}(\mu)$	بواسون
Logit	$\eta_i = G(\mu) = \log \frac{1}{1 - \mu}$	ثنائي الحدين

(Linearization) لدالة الربط $\eta = G(X^T \beta)$ حيث ستكون محل الاهتمام بدلا عن β ، وان الصيغة التي حددها $Nelder - McCullagh^{(11)}$ ضمن اسلوب المربعات الصغرى الموزونة التكرارية $IWLS$ ستكون :-

$$\hat{\beta}^{new} = \hat{\beta}^{old} + (X^T W X)^{-1} X^T (y - \underline{\mu}) \dots\dots\dots(9)$$

(2-2) الانحدار المعلمي اللاخطي:-

في الحياة العملية لا يمكن الحصول على علاقة خطية بين المتغيرات وفي احيان كثيرة تكون العلاقة الحقيقية بينها لاخطية ولغرض تحديد هذه العلاقة يتم اللجوء الى اسلوب الانحدار اللاخطي .وقد حدد كل من $Draper , Smith^{(3)}$ بأن اي صيغة تخالف شكل دالة الانحدار: $Y = f(X_i, \beta) + e$ ولا يمكن الحصول عليها باي تحويل لوغاريتمي أو غيره ستكون أنموذج الانحدار اللاخطي ، وعلى هذا يمكن افتراض ان الأنموذج يأخذ الصيغة التالية:

$$\underline{Y} = f(X_i, \theta_i) + \underline{e}_i \dots\dots\dots(10)$$

أذ أن :-

θ يمثل معالم الانموذج

$$\theta = [\theta_1 \theta_2 \dots \dots \dots \theta_p]^T$$

فان دالة تايلور ستكون بالوصف :

$$f(X_i, \theta) = f(X_i, \theta_0) + \sum_{i=1}^p \left[\frac{\partial f(x, \theta)}{\partial \theta} \right]_{\theta=\theta_0} (\theta_i - \theta_{i0}) \dots\dots\dots(11)$$

$$f^0 = f(x_i, \theta_0), \underline{\beta}_i^0 = \theta_i - \theta_i^0 \dots\dots\dots \text{وبتعويض}$$

$$\underline{Z}^0 = \frac{\partial f(X_i, \theta)}{\partial \theta} \dots\dots\dots \text{سيكون}$$

$$\underline{Y} = \underline{f}^0 + \sum_{i=1}^p \beta_{i0} Z_{i0} + \underline{e}_i \dots\dots\dots(12)$$

وهي تمثل دالة انحدار خطية ، وعند تحويل المقدار $\underline{f}^0$ الى يمين المعادلة (12) يكون :

$$\underline{Y} - \underline{f}^0 = \sum_{i=1}^p \beta_{i0} Z_{i0} + \underline{e}_i \dots\dots\dots$$

$$\dots\dots\dots \text{لتكون صيغة المعالم الاولى بالوصف :-}$$

$$\underline{\hat{\beta}}^0 = (Z^{0T} Z^0)^{-1} Z^{0T} (\underline{Y} - \underline{f}^0) \dots\dots\dots(13)$$

ان الصيغة التكرارية للحصول على قيم المعالم

$$\theta_{m+1} = \theta_m + \hat{\beta}_m = \theta_m + (Z_m^{0T} Z_m^0)^{-1} Z_m^{0T} (\underline{Y} - \underline{f}^0) \dots\dots\dots(14)$$

أذ أن m تشير الى مرحلة التقدير .

3- الانحدار اللامعلمي

وهو اسلوب احصائي مغاير لمفهوم الانحدار المعلمي ويتفق معه بالهدف النهائي وهو الحصول على افضل تقدير لمنحنى بيانات تتطابق او تقترب من التطابق مع منحنى المتغير المعتمد (Y) ، فلو كانت لدينا ظاهرة من الظواهر المختلفة ولا توجد هناك اي فرضيات تحكم العلاقة بين متغير الاستجابة (Y) والمتغير

التوضيحي (Z_i) ولا يمكن تحديد اي علاقة سواء كانت علاقة خطية او علاقة لا خطية بينهما عندها ستكون العلاقة المقترحة بين (Y, Z) تسمى بالانحدار اللامعلمي (*Nonparametric Regression*)
 أنموذج الانحدار اللامعلمي البسيط ويعد من أكثر النماذج اللامعلمية شيوعا وصيغته

$$Y = m(Z_i) + e \quad \dots\dots\dots(15)$$

$m(.)$ تمثل دالة التمهيد اللامعلمي الواجب تقديرها بالطرق اللامعلمية وتعتمد اساليب تقديره في مجملها على التمهيد اللامعلمي *Nonparametric Smoothing* ، وهو توفيق المنحنيات من دون الحاجة الى أنموذج لوصف نمط متغير الاستجابة (Y) ، ويعد المدخل الرئيس للانحدار اللامعلمي الذي يهدف بالاساس الى ايجاد افضل منحني ممهد يطابق او يقترب من التطابق مع منحني متغير الاستجابة (Y) وهناك عدة اساليب لتوفيق المنحنيات ومنها الدالة اللبية (*Kernel Function*) في صيغها الرياضية المختلفة ، وهي من أشهر الدوال الوزنية التي تستخدم في تمهيد بيانات المتغير المعتمد (Y) والتي تملك الخواص الاستمرارية التماثل ، الكثافة احتمالية ، تملك تباين معلوم ومتوسط = صفر ، وللدالة اللبية عدة صيغ مختلفة منها ما موضح في الجدول ادناه:-

الجدول رقم (2) بعض انواع الدوال اللبية لمتغير واحد		
ت	الدالة اللبية	صيغة الدالة
1	Uniform Kernel	$K(u) = \frac{1}{2} \quad u < 1$
2	Triangular Kernel	$K(u) = 1 - u \quad \sin(u) < 1$
3	Gaussian Kernel	$K(u) = (2\pi)^{-\frac{1}{2}} e^{-\frac{u^2}{2}} \quad 0 < u < \infty$
4	Epanchnikov kernel	$K(u) = \frac{3}{4} (1 - u^2) \quad u < 1$

أذ أن الصيغة 4 تقع ضمن عائلة واحدة من الدوال باسم *Epanchnikov* والتي تملك صيغة موحدة هي :

$$K(u) = C_a (1 - u^2)^a$$

$$C_a = \frac{\Gamma(2a+2)}{\Gamma(a+1)^2} 2^{-2a-1}$$

أذ أن :-

تمثل درجة الدالة المطلوبة ($a = 1, 2, 3 \dots$)

$$u = \frac{Z_i - Z_0}{h}$$

أذ أن :-

Z_i يمثل قياسات المتغير التوضيحي

Z_0 يمثل مشاهدة من مشاهدات المتغير التوضيحي

h يمثل عرض الحزمة (*The bandwidth*) ويشار اليها في مصادر عديدة باسم معلمة التمهيد (*Smoothing Parameter*) ، وهي الجزء الاهم في الدوال اللبية ، حيث اوضح ⁽⁶⁾ *Hardle* بان اختيار h هو الجزء الاهم في عملية التمهيد ، ويسبق اختيار اي دالة من الدوال اللبية وبين ايضا بأن h تمثل المعلمة الحرة التي تملك تأثيرا قويا على نتائج الدوال اللبية وبالتالي فإن أي مقدار من مقدرات الانحدار اللامعلمي التي تعتمد على الدالة اللبية فضلا عن الدالة نفسها سيكون حساسا جدا للتغيرات في قيمة h ، والذي سيؤدي بدوره الى تغيير شكل المنحنى الممهد للمتغير المعتمد (Y) ، والتي يتم احتسابها بالصيغ التي حددها ⁽¹³⁾ *Scott* للمتغير الواحد ولعدة متغيرات

$$\left. \begin{aligned} h &= 1.06 \hat{\sigma} n^{-0.2} \\ h &= 1.06 R n^{-0.2} \end{aligned} \right\} \dots\dots\dots(16)$$

أذ أن R تمثل الفرق بين الربع الثالث والربع الاول و $\hat{\sigma}$ تمثل الانحراف المعياري لمشاهدات العينة

وقد بين $Emmanuel^{(17)}$ ضرورة مراعاة حقيقتين رئيسيتين في اختبار معلمة التمهيد h الاولى عند اختيار معلمة التمهيد (h) كبيرة سيؤدي الى تمهيد كبير للبيانات مع تحيز كبير وتباين قليل والثانية عند اختيار معلمة التمهيد (h) صغيرة سيؤدي الى تمهيد صغير للبيانات وتحيز قليل وتباين كبير .
(1-3) اساليب تقدير الاحدار اللامعلمي أن المنطلق الاساسي في تقدير الانموذج اللامعلمي والذي صيغته $Y = m(Z_i) + e$ هو بتمهيد مشاهدات المتغير المعتمد (Y) عن طريق تحديد الاوزان لهذه المشاهدات حيث سيتم تخصيص الوزن الاكبر للمشاهدات الاقرب (الاكثر تأثيرا) والوزن الاقل او الصفر للمشاهدات البعيدة (الاقل تأثيرا) ، ان المفهوم الرياضي للاوزان يمكن توضيحه عن طريق

$$E(Y/Z) = \int y f(y/z) dy = \int \left[\frac{f(y,z)}{\hat{f}(z)} \right] y dy$$

أذ أن الجزء $\int \left[\frac{f(y,z)}{\hat{f}(z)} \right]$ يمثل كدالة وزنية لمشاهدات المتغير المعتمد (y) وهناك اساليب متعددة لتحديد المقدّر اللامعلمي $m(Z_i)$ سنوضح منها ما يأتي :-

(1-1-3) المدرج التكراري

وهو من اكثر المقدرات اللامعلمية شهرة والذي حدد ملامحه الكثير من الباحثين ، والصيغة الرياضية الخاصة بالمدرج التكراري هي:

$$\hat{f}(z) = \frac{1}{nh} \sum_{i=1}^n \sum_{j=1}^n I(z_i \in B_j) I(z \in B_j) \dots \dots \dots (17)$$

وبين $Hardle^{(6)}$ انه في حالة تطوير هذا المقدّر فيتم التعويض كالآتي :

$$Z_i = m_j$$

$$B_j = [m_j - \frac{h}{2}, m_j + \frac{h}{2}] \quad \text{فان الفترة } B_j \text{ ستكون}$$

$$\hat{f}_h(m_j) = \frac{1}{nh} \# \{X_i \in B_j\} \quad \text{والمقدّر سيكون} \quad \dots \dots \dots (18)$$

(2-1-3) المقدّر الساذج Naïve Estimator

وهو من المقدرات البسيطة الشائعة الاستخدام وصيغته الرياضية :-

$$\hat{m}(z) = \frac{1}{nh} \sum_{i=1}^n W_i(u) Y \quad \dots \dots \dots (19)$$

$$u_i = \frac{Z_i - Z_0}{h} \quad \text{أذ أن :-}$$

$$W_i(u) = \begin{cases} 1 & |u| < \frac{1}{2} \\ 0 & o.w \end{cases} \quad W_i(u) \text{ يمثل مؤشرا وزنيا يعتمد على العلاقة الآتية :-}$$

واورد $Hansen^{(8)}$ صيغة اخرى للمقدّر هي:

$$\hat{m}(z) = \frac{\sum_{i=1}^n 1(|Z_i - Z_0| \leq h) Y_i}{\sum_{i=1}^n 1(|Z_i - Z_0| \leq h)} \quad \dots \dots \dots (20)$$

وهو يمثل الوسط الحسابي الموزون لقيم متغير الاستجابة (Y) المراد تمهيدها ، وبالتالي فان الصيغة النهائية للمقدّر بدلالة المؤشرات الوزنية ستكون :

$$\hat{m}(z) = \frac{\sum_{i=1}^n W_i(u) Y_i}{\sum_{i=1}^n W_i(u)} \quad \dots \dots \dots (21)$$

(3-1-3) مقدّر برستلي جاو

ويكون بالصيغة

$$\hat{m}(z) = \frac{1}{h} \sum_{i=1}^n (Z_i - Z_{i-1}) K\left(\frac{Z_i - Z}{h}\right) Y \quad \dots \dots \dots (22)$$

وعندما تكون المشتقات المستمرة لدالة الجزء اللامعلمي $\hat{m}(z)$ تنتمي للفترة $[0,1]$ فان صيغة المقدّر ستكون

$$\hat{m}(z) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{Z_i - Z}{h}\right) Y \quad \dots \dots \dots (23)$$

وهي صيغة المقدّر اللبي ($Kernel Estimator$) والذي يعد من اهم المقدرات الشائعة الاستخدام من المقدرات اللبية ، وفي حالة متعدد المتغيرات فان صيغة المقدّر اللبي ستكون بالوصف :-

$$\hat{m}(z) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\det(H)} K(H^{-1}(z_i - z)) Y \quad \dots \dots \dots (24)$$

(3-1-4) مقدر ناداريا - واتسون

تم اقتراح هذا المقدر من قبل الباحثين Nadareya - Watson وذلك بإبدال المؤشر الوزني في صيغة المقدر الساذج (19) بالدالة اللبية ليصبح وصف المقدر بالصيغة الآتية :-

$$\hat{m}_h(z) = \frac{\sum k\left(\frac{z_i - z}{h}\right)y}{\sum k\left(\frac{z_i - z}{h}\right)} \dots\dots\dots(25)$$

وبالتالي تم الحصول على مقدر أكثر كفاءة من المقدر الساذج مع بقاء حساسيته العالية لمعلمة التمهيد h ، وأورد كل من ⁽¹⁵⁾ Demir, Toktamis صيغة محدثة للمقدر بالاعتماد على الترتيب نفسه في الصيغة المحدثة للمقدر اللبي (25) فإن الصيغة ستكون بالوصف الآتي :-

$$\hat{m}_h(z_i) = \frac{\sum_{i=1}^n \frac{1}{h\lambda_i} K\left(\frac{z_i - z}{h\lambda_i}\right)y}{\sum_{i=1}^n \frac{1}{h\lambda_i} K\left(\frac{z_i - z}{h\lambda_i}\right)} \dots\dots\dots(26)$$

مع الإشارة الى صيغة ناداريا - واتسون (مع ترك مشاهدة واحدة) *leave one out N Estimator* والتي ستكون بالوصف الآتي:

$$\hat{m}_{-1}(Z_i) = \frac{\sum_{i \neq j} K\left(\frac{Z_i - Z_j}{h}\right)Y_j}{\sum_{i \neq j} K\left(\frac{Z_i - Z_j}{h}\right)} \dots\dots\dots(27)$$

(3-1-5) مقدر الانحدار الموضوعي لكثيرات الحدود

عند افتراض أن مشتقات دالة الانحدار موجودة ومستمرة وبدرجة $(P+1)$ في أي نقطة من نقاط (Z_i) أن أن :-

$$\underline{\beta}(Z) = [\beta_0(Z), \beta_1(Z), \dots \dots \dots \beta_p(Z)]^T$$

$$\beta_j(Z) = \frac{m^j(Z)}{j!}$$

أن أن :-

وعند تصغير الدالة والتي هي :

$$\psi(\underline{\beta}_{(z)}) = \sum_{i=1}^n [Y_i - \sum_{j=0}^p \beta_j(z)(Z_i - Z)]^2 w_{n,i} \dots\dots\dots(28)$$

$$w_{n,i} = \frac{1}{n} K\left(\frac{Z_i - Z}{h}\right)$$

أن أن :-

أن تقدير $\underline{\beta}_{(z)}$ يكون بأسلوب المربعات الصغرى الموزونة أن أن :-

$$\hat{\underline{\beta}} = [Z^T w Z]^{-1} Z^T w Y \dots\dots\dots(29)$$

عندها سيكون المقدر اللبي لمجموعة المتغيرات يمثل

$$\hat{m}_{p,h}(Z_i) = \hat{\beta}_0(Z) \dots\dots\dots(30)$$

وقيم المتغير المعتمد الموزونة يكون كالآتي :

$$T_{h,j} = \sum_{i=1}^n K_h(Z_i - Z) (Z - Z_i)^j Y_i \dots\dots\dots(31)$$

أن أن الدالة الوزنية ستكون

$$S_{h,j} = \sum_{i=1}^n K_h(Z_i - Z) (Z - Z_i)^j \dots\dots\dots(32)$$

$$\hat{m}_{0,h} = \frac{\sum_{i=1}^n K_h(Z_i - Z) Y_i}{\sum_{i=1}^n K_h(Z_i - Z)} = \frac{T_{0,h}}{S_{0,h}}$$

وعندما $P = 0$ فإن :-

وهو يمثل بالضبط مقدر ناداريا - واتسون ، وعندما $P = 1$ فإن $\hat{\beta}_{(z)}$ ستكون بالوصف الآتي :-

$$\hat{\underline{\beta}}_{(Z)} = \begin{bmatrix} S_{h,0} & S_{h,1} \\ S_{h,1} & S_{h,2} \end{bmatrix}^{-1} \begin{bmatrix} T_{h,0} \\ T_{h,1} \end{bmatrix} \dots\dots\dots(33)$$

4- الانحدار شبه المعلمي Semiparametric Regression Model⁽⁴⁾⁽⁵⁾⁽⁶⁾⁽¹⁰⁾

وهو أسلوب احصائي يحقق الخصائص العامة للانحدار المعلمي واللامعلمي ويتفق معهما في الغاية نفسها وهي الحصول على أفضل منحنى للبيانات يطابق أو يقترب من التطابق مع منحنى المتغير المعتمد (Y) بالدمج بين أساليب التقدير المعلمية واللامعلمية و حدد ⁽⁶⁾ Hardle تعريف الانحدار شبه المعلمي بأنه

أنموذج يحتوي على قسمين أحدهما معلمي بأبعاد نهائية والآخر لامعلمي بأبعاد لا نهائية ، ان الصيغة العامة للأنموذج شبه المعلمي هي :-

$$Y = f(X, \beta) + m(Z) + e \quad \dots\dots\dots(34)$$

وتختلف طرق تقدير معالم أنموذج الانحدار شبه المعلمي باختلاف صيغته الرياضية ولكنها بصورة عامة تكون بأسلوبين رئيسيين الأول وهو الأكثر استخداماً من قبل الباحثين ويعني بتقدير الجزء المعلمي في المرحلة الأولى (بأي أسلوب من الأساليب التقدير المعلمي) وبعدها في المرحلة الثانية يتم تقدير الجزء اللامعلمي (بالأساليب التمهيد اللامعلمي) بالاعتماد على تقديرات المرحلة الأولى والثاني وهو معاكس للأسلوب الأول حيث يتم تقدير الجزء اللامعلمي في المرحلة الأولى ، وفي الثانية يتم تقدير الجزء المعلمي بالاعتماد على تقديرات المرحلة الأولى .

ان اختيار الأسلوب الأول (معلمي لامعلمي) يوفر للباحثين من الناحية المبداية إمكانية تقدير عدد كبير من المنحنيات المتوافقة مع منحنى متغير الاستجابة (Y) واختيار الأسلوب الثاني (لامعلمي معلمي) فإنه سيوفر عدد كبير من النماذج الموفقة وبالتالي عدداً كبيراً من المنحنيات التقديرية لهذه النماذج.

(1-4) أنموذج الانحدار الخطي الجزئي Partial Linear Model (PLM)⁽⁹⁾⁽¹⁰⁾

ويعد من أشهر النماذج شبه المعلمية ويرد عند الكثير من الباحثين باسم الأنموذج شبه المعلمي البسيط Simple Semiparametric Model وتم اقتراحه من قبل الباحثين Engle , Granger, Rice and Weiss⁽²⁸⁾ وصيغته الرياضية :-

$$Y = X^T \beta + m(z) + e \quad \dots\dots\dots(35)$$

ان الأنموذج PLM يعد الأنموذج الكلاسيكي المعبر عن مفهوم الانحدار شبه المعلمي وذلك لوضوح حالة التفاعل بين المكون المعلمي والمكون اللامعلمي ولهذا السبب فإنه كان محط اهتمام للكثير من الباحثين والذي انعكس على الأساليب المتعددة لتقدير معالمه .

(1-1-4) طريقة التقدير بالبواقي Residuals Estimation Method⁽⁶⁾⁽¹⁷⁾

عند اخذ التوقع لأنموذج الانحدار الخطي الجزئي

$$E(Y/Z) = E(X^T \beta / Z) + E(m(Z)/Z) + E(e/Z)$$

وبطرح المعادلة الأخيرة من معادلة الأنموذج PLM يكون :

$$Y - E(Y/Z) = (X^T - E(X^T/Z))\beta + (m(Z) - E(m(Z)/Z)) + (e - E(e/Z))$$

فان الأنموذج سيكون بالوصف الآتي :

$$Y - E(Y/Z) = (X^T - E(X^T/Z))\beta + e \quad \dots\dots\dots(36)$$

$$D = Y - E(Y/Z)$$

وبتعويض

$$R = X^T - E(X^T/Z)$$

فان المعادلة (36) ستكون بالوصف :

$$D = R\beta + e \quad \dots\dots\dots(37)$$

والأنموذج الأخير يمثل معادلة انحدار خطية تقليدية يمكن تقدير معالمها بالأساليب المعلمية إذ أن :-

$$\hat{\beta} = (R^T R)^{-1} R^T D$$

ويمكن كتابة الأنموذج (37) بصيغة الفروق الآتية :

$$e_1 = e_2 \beta + e \quad \dots\dots\dots(38)$$

وستكون صيغة التقدير لمعالم الدالة المعلمية بالوصف الآتي :

$$\hat{\beta}_{res} = (e_2^T e_2)^{-1} e_2^T e_1 \quad \dots\dots\dots(39)$$

(2-1-4) طريقة التقدير بالفروقات Differences Estimation Method¹⁴

حدد Yatchow⁽¹⁴⁾ أسلوباً آخر لتقدير أنموذج PLM حيث يتم ترتيب المتغيرين Y, X على اساس ترتيب المتغير Z إذ أن :-

$$Z_1 \leq Z_2 \leq Z_3 \leq \dots \leq Z_i \quad i = 1, 2, 3 \dots \dots \dots n$$

يتم اخذ الفرق الأول $Y_i - Y_{i-1} = (X_i^T - X_{i-1}^T)\beta + (m(Z_i) - m(Z_{i-1})) + (e_i - e_{i-1})$ من الناحية التقريبية فان

$$m(Z_{i-1}) = 0$$

فيكون

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

$$\Delta_e = e_i - e_{i-1} \quad , \quad \Delta_{X^T} = X_i^T - X_{i-1}^T \quad , \quad \Delta_Y = Y_i - Y_{i-1}$$
 فان النموذج سيكون بالوصف الاتي :

$$\Delta_Y = \Delta_{X^T}\beta + \Delta_e$$

وهو يمثل دالة انحدار خطية تقليدية وبالتالي فان تقدير معالم الجزء المعلمي سيكون بالوصف :-

$$\hat{\beta}_{diff} = [\Delta_{X^T}^T \Delta_{X^T}]^{-1} \Delta_{X^T}^T \Delta_Y \quad \dots\dots\dots(40)$$

(3-1-4) طريقة Speckman للتقدير باسلوب تطوير الجزء اللامعلمي

طور $Speckman^{(21)}$ طريقة التقدير باقتراحه وصفا معلمي للجزء اللامعلمي $m(\cdot)$ والتعويض عنه بالمصفوفة $\mathcal{W}_Y$ اذ ان $\mathcal{W}$ تمثل مصفوفة من درجة $n \times q$ ومن الرتبة التامة والدرجة q تكون غير معلومة ليكون النموذج PLM بالمصفوفات الاتية :

$$Y = X^T \beta + \mathcal{W}_Y + e \quad \dots\dots\dots(41)$$

ان المعادلات الطبيعية التي وصفها $Speckman^{(21)}$ تكون :

$$X^T X \beta = X^T (Y - \mathcal{W}_Y)$$

والتي تمثل مصفوفة الاسقاطات (Projection Matrix) ، واستند $Speckman^{(21)}$ الى اسلوب $Green, Jenson, Seheult$ لتحديد صيغة $\hat{\beta}$ وذلك بابدال مصفوفة P_w بمصفوفة الممهدات اللامعلمية S والتي تم تعريفها في اسلوب التقدير اللامعلمي (Locall Linear Regression) بالصيغة (2-45) حيث سيكون المقدر المعلمي واللامعلمي بالوصف الاتي :

$$\hat{\beta}_{GJS} = (X^T (I - S) X)^{-1} X^T (I - S) Y \quad \dots\dots\dots(42)$$

$$\hat{m}_{GJS} = S(Y - X^T \hat{\beta}) \quad \dots\dots\dots(43)$$

(2-4) نموذج الانحدار الخطي الجزئي الموسع $Generalized Partial linear Model (GPLM)^{(6)(10)}$

يعد النموذج $GPLM$ هو الامتداد لمفهوم الانحدار الخطي الموسع GLM_z ومن علاقته بالنموذج الشبه معلمي وان اساليب تقديره تعتمد بالدرجة الاساس على تقدير GLM_z . ان الصيغة الرياضية للنموذج

$$E(Y/X, Z) = G \{X^T \beta + m(Z)\} \quad \dots\dots\dots(44)$$

ان مبدأ تقدير هذا النموذج يعتمد على أساسين :

تقدير $\hat{\beta}$ مع اعتبار $m(\cdot)$ معلومة تقدير و $m(\cdot)$ على اساس ان β معلومة ، ولقد بينت $Muller^{(14)}$ بانه عند التقدير سيتم افتراض أن أول عزمين شرطين للمتغير المعتمد Y موجودين اذ أن :-

$$E(Y/X, Z) = \mu = G(\eta) = G(X^T \beta + m(Z))$$

$$Var(Y/X, Z) = \sigma^2 Var(\mu)$$

(1-2-4) طريقة Backfitting

وهي الطريقة التي تم اقتراحها من قبل $Hastie \& Tibshirani^{(11)}$ كخوارزمية متسلسلة لتقدير النموذج التجميعي والفكرة الاساس تكون بتقدير المكونات التجميعية بصورة منفصلة، ولان النموذج شبه المعلمي يعد الحالة الخاصة من النموذج التجميعي وبدالة ربط (Identity) وبفرض P كمصفوفة اسقاطات (Projection Matrix) اذ أن :-

$$P = X(X^T X)^{-1} X^T$$

S و تمثل مصفوفة الممهدات اللامعلمية

وبالعودة الى النموذج $GPLM$ فان $Hastie \& Tibshirani$ اقترحا استخدام هذا النظام لاجاد تقدير موزون ومتسلسل ، ان الخوارزمية الخاصة بالتقدير للنموذج $GPLM$ باسلوب Backfitting هي :-

$$\left. \begin{aligned} \beta^{new} &= (X^T \mathcal{W} \tilde{X})^{-1} X^T \mathcal{W} \tilde{Z} \\ m^{new} &= S(Z - X^T \beta) \end{aligned} \right\} \quad \dots\dots\dots(45)$$

أذ أن :- $\tilde{X} = (I - S)X$, $v = \ell'_i(\eta)$

$$w = \ell''_i(\eta) , \tilde{Z} = \tilde{X}\beta - w^{-1}\underline{v}$$

و $\ell'_i(\eta)$, $\ell''_i(\eta)$ تمثلان المشتقة الاولى والثانية لدالة الامكان الاعظم $\ell_i(\eta)$ ولكل خطوة تسلسلية يمكن أن تشكل المعادلة الخطية الآتية :

$$X\beta^{new} + m^{new} = R^{\beta}Z \dots\dots\dots(46)$$

$$(R^{\beta} = \tilde{X}(X^T\tilde{w}\tilde{X})^{-1}X^T\tilde{w}(I - S) + S$$

أذ أن :-

5- الجانب العملي

النفط هو أهم دعائم الاقتصاد حيث يشكل نسبة كبيرة من الدخل القومي للبلد المنتج له ، أن الصناعات النفطية لاتعتمد على إنتاج النفط فقط ، بل هي وحدة متكاملة تتضمن عدة عمليات مترابطة من استخراجه وتكريره وتسويقه ونقله في الداخل والخارج للمنتوجين الخام والمكرر ، ويعد النقل الداخلي للمنتوجات النفطية البيضاء (بنزين ، نفط ابيض ، زيت الغاز (الكاز)) من العمليات الاساس المؤثرة في تطور القطاع النفطي ، حيث يتم نقل هذه المنتجات من المصافي المنتشرة في ارجاء العراق او من مستودعات الموانئ الى مستودعات شركة خطوط الانابيب النفطية تمهيدا لنقلها لمستودعات التخزين في المحافظات العراقية لغرض توزيعها على محطات التوزيع المختلفة ، وتعد شركة خطوط الانابيب النفطية هي المسؤول الرئيس والوحيد عن هذه العملية المعقدة منذ بدايات الانتاج النفطي في العراق وحتى الان، ان هدفنا من التطبيق هو تحديد افضل نموذج يمثل كمية النقل بواسطة الانابيب مع بقية المتغيرات الاخرى ، و قد تم الحصول على بيانات نقل المنتجات النفطية بواسطة الانابيب حيث تضمنت خمسة متغيرات رئيسية وهي Y التي تمثل كميات المنتجات النفطية البيضاء (النفط الابيض ، البنزين ، الكاز) و X_1 تمثل كمية الانتاج النفطي من هذه المنتجات من المصافي الجنوبية و X_2 يمثل كميات المنتجات المستوردة و X_3 ستمثل كمية المنتجات المنقولة برا و X_4 ستمثل فرق العدادات حيث تم اجراء تحليل هذه البيانات بثلاث مراحل متتابعة ففي المرحلة الاولى تم بناء مختلف النماذج المعلمية وشبه المعلمية وتحليل مبسط لنتائجها ، وفي المرحلة الثانية تم اجراء المقارنة بين النماذج المعلمية وشبه المعلمية لتحديد النموذج او النماذج الافضل وبثلاث اساليب ، وفي المرحلة الثالثة تم تحسين درجة تمثيل البيانات لمختلف النماذج شبه المعلمية واجراء المقارنات بينها لتحديد النموذج الافضل فيما بينها .

(1-5) بناء النماذج المعلمية وشبه المعلمية

باستخدام البرنامج التطبيقي ($i - xplore$) تم بناء النماذج المعلمية وشبه المعلمية ، حيث يوفر هذا النظام الامكانية لبناء النماذج المعلمية كالانحدار الخطي العام والموسع، وكذلك يوفر صفاً واسعاً من النماذج شبه المعلمية والتي سنكتفي منها بالنموذجين الخطي الجزئي PLM والخطي الجزئي الموسع $GPLM$ وحسب الآتي:-

(1-1-5) النماذج المعلمية

عند افتراض ان سلوك المتغيرات المستقلة يكون خطياً فقد تم بناء اربعة نماذج معلمية (نموذج GLM + ثلاثة نماذج GLM_z) وكما يأتي :-

$$\hat{Y} = 1147.37 + 0.485 X_1 + 0.460 X_2 + 0.101 X_3 - 2.08 X_4$$

$$\hat{Y} = G \{8.38 + 4.71E^{-5} X_1 + 4.16E^{-5} X_2 + 7.88E^{-6} X_3 - 0.17E^{-3} X_4\}$$

$$\hat{Y} = G \{0.17E^{-3} - 4.4E^{-9} X_1 - 3.7E^{-9} X_2 - 5.4E^{-10} X_3 - 1.3E^{-8} X_4\}$$

$$\hat{Y} = G \{- 0.1E^{-3} + 4.4E^{-9} X_1 + 3.7E^{-9} X_2 + 5.4E^{-10} X_3 - 1.3E^{-8} X_4\}$$

ودوال الربط (G) المستخدمة هي : (Log(μ), Reciprocal, Log logit)

جدول رقم (3) النماذج المعلمية					
Deviance	Dispersion	R ²	BIC	AIC	النموذج
8653531791	7524810	0.4615	21595.94	21570.69	M1
883877	-	0.4265	896720	896695	M2
103.98	0.0904	0.0002	27571	27546	M3
103.96	-	0.0025	23819	23793	M4

حيث يمكن تحديد الامور الآتية :-

- ان النماذج الخطية الموسعة GLM_Z غير كفوءة تماما لان تقديرات معالمها $(b_0, b_1, b_2, b_3, b_4)$ تكاد تقترب للصفر في معظمها، في حين ان الانموذج الخطي GLM كان اكثر كفاءة منها.
- ان التشنت $Dispersion$ والانحرافات $Deviance$ تكون عالية جدا في الانموذج الخطي GLM وقليلة في النماذج الموسعة GLM_Z في حين ان معايير AIC, BIC كانت في الانموذج الخطي اقل منها في النماذج الموسعة .
- نسبة تفسير معادلة الانحدار للبيانات R^2 بالنسبة للانموذجين الخطي العام والموسعة (بدالة الربط $G(.) = \log(\mu)$) هي افضل بكثير جدا من الانموذجين الموسعين الاخرين (انموذج بدالة الربط $Reciprocal$ والانموذج بدالة الربط $\logit$) ، مع بقاء كفاءة الانموذج الخطي العام GLM لكون مقدراته اقرب للواقع العملي .
- ومن ملاحظة الانموذج الخطي العام (رقم 1) فان الكمية الثابتة لضخ المنتجات النفطية البيضاء تكون بالمقدار (1147.37) م³ ، وان تأثير المتغير X_1 (الانتاج النفطي) هو تأثير موجب بمعنى ان الزيادة التي تحصل فيه وبمقدار وحدة واحدة تؤدي الى زيادة الكميات المنقولة في الانابيب بنسبة 0.485 من هذه الوحدة وكذلك الامر بالنسبة للمتغير X_2 (الكميات المستوردة) اذ ان زيادة وحدة واحدة منها تؤدي الى ارتفاع في الكميات المنقولة بنسبة 0.460 من هذه الوحدة ويمكن ملاحظة التأثير القليل نسبيا للمتغير X_3 (النقل البري) حيث ان زيادة وحدة واحدة من كمية المنتجات البيضاء المنقولة برا يؤدي الى زيادة طفيفة في المنتجات المنقولة داخل الانابيب تكون بحدود 0.101 من هذه الوحدة في حين ان تأثير المتغير الرابع X_4 (فرق العدادات) كان كبيرا جدا ولكن بالاتجاه السالب حيث ان زيادة وحدة واحدة من هذا المتغير تؤدي الى تقليل كمية المنتجات البيضاء المنقولة في الانابيب بمقدار الضعفين وبحدود 2.08 من هذه الوحدة وفي ادناه الاشكال الخاصة بالنماذج المعلمية الاربع.

(2-1-5) بناء النماذج شبه المعلمية

ان المرونة التي يتمتع ، الانموذج شبه المعلمي توافر الامكانية لتوزيع المتغيرات المستقلة على المكونين المعلمي واللامعلمي للانموذج ، وعلى هذا الاساس تم بناء 28 انموذجا شبه معلميا تمثل البيانات وكالاتي :-

14 انموذج PLM وحسب الصيغة (35) و 14 انموذج $GPLM$ وحسب الصيغة (44) وباختيار دالة الربط اللوغاريتمية فقط لكون دالتي الربط (كاما ومعكوس ثنائي الحدين) تحققان $adj R^2$ ضئيلة جدا تصل الى 0.0002 فقط مما تعذر اختيارها، وتم تحديد المسافات البينية $b.w$ بالاعتماد على الصيغة (16) حيث كانت

$$h_1 = 3077.9, h_2 = 2577.6, h_3 = 8684.5, h_4 = 160.32$$

(1-2-1-5) النماذج $PLM, GPLM$ ذات المتغير الواحد في المكون المعلمي

تم اعتبار ان هناك متغيرا واحدا يسلك سلوكا خطيا او معلميا (سلوكا مستقرا) مشكلا المكون المعلمي وثلاثة متغيرات تسلك السلوك اللاخطي او اللامعلمي (سلوكا غير مستقر) لتكون بدورها المكون اللامعلمي من الانموذج PLM وحسب الصيغة (35) أو الانموذج $GPLM$ وحسب الصيغة (44) وعلى هذا الاساس تم بناء النماذج الاتية:-

1. اربعة نماذج PLM تتكون من متغير واحد في المكون المعلمي وثلاثة متغيرات في المكون اللامعلمي والصيغ الرياضية التالية

$$\hat{Y} = 0.352X_1 + m(X_2, X_3, X_4) \quad \text{المكون المعلمي يمثل } (X_1) \text{ واللامعلمي يمثل } (X_2, X_3, X_4) \text{ سيكون}$$

$$\hat{Y} = 0.346X_2 + m(X_1, X_3, X_4) \quad \text{المكون المعلمي يمثل } (X_2) \text{ واللامعلمي يمثل } (X_1, X_3, X_4) \text{ سيكون}$$

$$\hat{Y} = 0.108X_3 + m(X_1, X_2, X_4) \quad \text{المكون المعلمي يمثل } (X_3) \text{ واللامعلمي يمثل } (X_1, X_2, X_4) \text{ سيكون}$$

$$\hat{Y} = -2.666X_4 + m(X_1, X_2, X_3) \quad \text{المكون المعلمي يمثل } (X_4) \text{ واللامعلمي يمثل } (X_1, X_2, X_3) \text{ سيكون}$$

2. اربعة نماذج $GPLM$ تتكون من متغير واحد في المكون المعلمي وثلاثة متغيرات في المكون اللامعلمي ، مع اختيار دالة بواسون فقط لكون بقية الدوال (كاما ، معكوس ثنائي الحدين) لا تحقق الا 0.0002 كقيمة تفسيرية للبيانات (R^2) وحسب الصيغ الرياضية التالية:-

$$\hat{Y} = G\{3.31E^{-5}X_1 + m(X_2, X_3, X_4)\} \quad \text{المكون المعلمي يمثل } (X_1) \text{ واللامعلمي يمثل } (X_2, X_3, X_4) \text{ سيكون}$$

- المكون المعلمي يمثل (X_2) واللامعلمي يمثل (X_1, X_3, X_4) سيكون $\hat{Y} = G\{2.96E^{-5}X_2 + m(X_1, X_3, X_4)\}$
 - المكون المعلمي يمثل (X_3) واللامعلمي يمثل (X_1, X_2, X_4) سيكون $\hat{Y} = G\{8.41E^{-6}X_3 + m(X_1, X_2, X_4)\}$
 - المكون المعلمي يمثل (X_4) واللامعلمي يمثل (X_1, X_2, X_3) سيكون $\hat{Y} = G\{-0.0002X_4 + m(X_1, X_2, X_3)\}$
- ومن ملاحظة النماذج اعلاه يمكن تحديد الملاحظات الآتية :-
- ان نماذج $GPLM$ تكون غير كفؤة لكون اوزان المكون المعلمي تقترب للصفر.
 - ان اوزان المتغير المعلمي في النماذج الاربعة الاولى (PLM) تكون قريبة من اوزان نفس المتغيرات في النموذج المعلمي الخطي ومن الواضح ان اوزان المتغيرين (الانتاج النفطي ، المستورد) في النموذج PLM هي اقل من مناظرتها في النموذج (GLM) في حين ان وزن المتغيرين الآخرين قد ارتفع عن مثيلاتها في النموذج المعلمي .
 - ان درجة تمثيل البيانات وحسب معاملي التحديد R^2 والتحديد المعدل $adjR^2$ في النماذج PLM تكون اعلى من نماذج $GPLM$ ، ولكل انموذج مقابل وكذلك اعلى من مقاييس الانموذجين GLM ، GLM_Z وكما في الجدول الآتي :

الجدول رقم (4) مقارنة بين مقاييس تمثيل البيانات			
ت	$R^2, adjR^2$ للنموذج PLM	$R^2, adjR^2$ للنموذج $GPLM$	$R^2, adjR^2$ للنموذج GLM, GLM_Z
1	$R^2 = 0.5658$ $adjR^2 = 0.5408$	$R^2 = 0.5461$ $adjR^2 = 0.5198$	$GLM \quad R^2 = 0.4615$ $adjR^2 = 0.4597$
2	$R^2 = 0.5591$ $adjR^2 = 0.5331$	$R^2 = 0.5354$ $adjR^2 = 0.5081$	$GLM_Z \quad R^2 = 0.4265$ $adjR^2 = 0.4245$
3	$R^2 = 0.5275$ $adjR^2 = 0.4929$	$R^2 = 0.5018$ $adjR^2 = 0.4654$	
4	$R^2 = 0.5604$ $adjR^2 = 0.5400$	$R^2 = 0.5346$ $adjR^2 = 0.5131$	

- ان مقاييس AIC, BIC في النموذج PLM تكون اقل بكثير جدا من مثيلاتها في النموذج $GPLM$
- ان التشتت $Dispersion$ والانحرافات $Deviance$ تكون عالية جدا في النموذج PLM وقليلة في النماذج $GPLM$

(2-2-1-5) النماذج $PLM, GPLM$ ذات المتغيرين في المكون المعلمي

حيث تم اعتبار ان هناك متغيرين يسلكان سلوكا خطيا او معلميا (سلوكا مستقرا) مشكلتا المكون المعلمي والمتغيرين الآخرين يسلكان سلوكا لخطيا او لامعلميا (سلوكا غير مستقر) لتكون بدورها المكون اللامعلمي من النموذج PLM ، وحسب الصيغة (35) أو النموذج $GPLM$ وحسب الصيغة (44) وعلى هذا الاساس تم بناء عدد 6 نماذج لكل نوع وحسب الآتي :-

1. ستة نماذج PLM تتكون من متغيرين في المكون المعلمي ومتغيرين في المكون اللامعلمي وحسب والصيغ الرياضية الآتية:-

- المكون المعلمي يمثل المتغيرين (X_1, X_2) والمكون اللامعلمي يمثل المتغيرين X_3, X_4 وبالصيغة الرياضية $\hat{Y} = 0.37X_1 + 0.33X_2 + m(X_3, X_4)$
- المكون المعلمي يمثل المتغيرين (X_1, X_3) والمكون اللامعلمي يمثل المتغيرين X_2, X_4 وبالصيغة الرياضية $\hat{Y} = 0.44X_1 + 0.11X_3 + m(X_2, X_4)$
- المكون المعلمي يمثل المتغيرين (X_1, X_4) والمكون اللامعلمي يمثل المتغيرين X_2, X_3 وبالصيغة الرياضية $\hat{Y} = 0.36X_1 - 2.56X_4 + m(X_2, X_3)$
- المكون المعلمي يمثل المتغيرين (X_2, X_3) والمكون اللامعلمي يمثل المتغيرين X_1, X_4 وبالصيغة الرياضية $\hat{Y} = 0.45X_2 + 0.101X_3 + m(X_1, X_4)$
- المكون المعلمي يمثل المتغيرين (X_2, X_4) والمكون اللامعلمي يمثل المتغيرين X_1, X_3 وبالصيغة الرياضية $\hat{Y} = 0.353X_2 - 2.46X_4 + m(X_1, X_3)$
- المكون المعلمي يمثل المتغيرين (X_3, X_4) والمكون اللامعلمي يمثل المتغيرين X_1, X_2 وبالصيغة الرياضية $\hat{Y} = 0.106X_3 - 2.17X_4 + m(X_1, X_2)$

2. ستة نماذج $GPLM$ تتكون من متغيرين في المكون المعلمي ومتغيرين في المكون اللامعلمي والصيغ الرياضية الآتية:-

- المكون المعلمي يمثل المتغيرين (X_1, X_2) والمكون اللامعلمي يمثل المتغيرين X_3, X_4 وبالصيغة الرياضية الآتية:-

$$\hat{Y} = G\{3.46E^{-5}X_1 + 2.83E^{-5}X_2 + m(X_3, X_4)\}$$

- المكون المعلمي يمثل المتغيرين (X_1, X_3) والمكون اللامعلمي يمثل المتغيرين X_2, X_4 وبالصيغة الرياضية الآتية:-

$$\hat{Y} = G\{4.26E^{-5}X_1 + 8.49E^{-6}X_3 + m(X_2, X_4)\}$$

- المكون المعلمي يمثل المتغيرين (X_1, X_4) والمكون اللامعلمي يمثل المتغيرين X_2, X_3 وبالصيغة الرياضية الآتية:-

$$\hat{Y} = G\{3.36E^{-5}X_1 - 0.2E^{-3}X_4 + m(X_2, X_3)\}$$

- المكون المعلمي يمثل المتغيرين (X_2, X_3) والمكون اللامعلمي يمثل المتغيرين X_1, X_4 وبالصيغة الرياضية الآتية:-

$$\hat{Y} = G\{4.07E^{-5}X_2 + 7.83E^{-6}X_3 +$$

$$m(X_1, X_4)\}$$

X_1, X_3 وبالصيغة الرياضية الآتية:-

$$\hat{Y} = G\{3.03E^{-5}X_2 - 0.2E^{-3}X_4 + m(X_1, X_3)\}$$

- المكون المعلمي يمثل المتغيرين (X_3, X_4) والمكون اللامعلمي يمثل المتغيرين X_1, X_2 وبالصيغة الرياضية الآتية:-

$$\hat{Y} = G\{8.24E^{-6}X_3 - 0.17E^{-3}X_4 + m(X_1, X_2)\}$$

ومن النماذج اعلاه يمكن تحديد الملاحظات التالية:-

- ان النماذج $GPLM$ غير كفوة لكون اوزان (معالم) المكون المعلمي قريبة للصفر في حين ان النماذج PLM تكون اكثر واقعية منها.
- ان معايير AIC, BIC تكون كبيرة جدا في النماذج $GPLM$ مقارنة بمثيلاتها في النماذج PLM .
- ان التشتت $Dispersion$ والانحرافات $Deviance$ تكون عالية جدا في نماذج PLM وقليلة في النماذج $GPLM$.
- وبملاحظة اوزان النماذج PLM مع اوزان الانموذج GLM ولكل انموذج على حدة وحسب الجدول (3-3).

حيث يمكن ملاحظة ان وزن المتغير X_1 (الانتاج النفطي) كان اقل من وزنه في الانموذج GLM مع ملاحظة ارتفاع قيمته في الانموذج الثاني ، وكذلك الامر بالنسبة لوزن المتغير X_2 (المستورد) حيث ان وزنه ينخفض لكل النماذج التي يشترك في مكونها المعلمي وارتفاع وزنه في الانموذج الرابع مع ملاحظة ان كلا المتغيرين يرتفع وزنه في الانموذج عندما يتشاركان في المكون المعلمي مع المتغير X_3 (النقل البري) والذي وزنه يتقارب بصورة كبيرة مع نظيره في الانموذج المعلمي ارتفاعا او انخفاضاً ولكافة النماذج ، اما المتغير الرابع X_4 فان وزنه يكون ولكافة النماذج اعلى من وزنه في الانموذج GLM مما يوضح ان تأثير هذا المتغير يكون بصورة اكبر في النماذج شبه المعلمية .

- ان درجة تمثيل البيانات $R^2, adjR^2$ في النماذج $PLM, GPLM$ وكما في الجدول الآتي:-

الجدول (5) مقارنة بين مقاييس تمثيل البيانات لنماذج المجموعة الثانية			
لنموذج $R^2, adjR^2$ GLM, GLM_z	لنموذج $R^2, adjR^2$ $GPLM$	لنموذج $R^2, adjR^2$ PLM	
$R^2 = 0.4615$ GLM $adjR^2 = 0.4597$ GLM_z $R^2 = 0.4265$ $adjR^2 = 0.4245$	$R^2 = 0.5085$ $adjR^2 = 0.4969$	$R^2 = 0.5332$ $adjR^2 = 0.5223$	
	$R^2 = 0.4641$ $adjR^2 = 0.4501$	$R^2 = 0.4952$ $adjR^2 = 0.4821$	
	$R^2 = 0.5058$ $adjR^2 = 0.4977$	$R^2 = 0.5299$ $adjR^2 = 0.5222$	
	$R^2 = 0.4537$ $adjR^2 = 0.4406$	$R^2 = 0.4831$ $adjR^2 = 0.4707$	
	$R^2 = 0.4994$ $adjR^2 = 0.4899$	$R^2 = 0.5288$ $adjR^2 = 0.5199$	
	$R^2 = 0.4575$ $adjR^2 = 0.4476$	$R^2 = 0.4878$ $adjR^2 = 0.4784$	

حيث يمكن وبصورة واضحة ملاحظة ان مقاييس تمثيل البيانات لنماذج PLM تكون بمجملها اعلى من مثيلاتها في النماذج GLM ، GLM_z ، $GPLM$ ، في حين ان نماذج $GPLM$ تتفوق مقاييس تمثيل البيانات فيها على مثيلاتها في النموذج GLM_z ولا تتفوق كلها على مقاييس النموذج GLM الا في النماذج رقم (5,3,1)، اما في النموذج (2) فان مقياس R^2 يكون اعلى بصورة طفيفة جدا ومقياس $adjR^2$ يكون ادنى بصورة طفيفة ايضا، اما في النموذج رقم (6) فان مقاييسه ادنى من مقاييس النموذج GLM ، والذي بمجمله يوضح تفوق النماذج شبه المعلمية على النماذج المعلمية في هذه المقاييس.

(3-2-1-5) النماذج $PLM, GPLM$ ذات الثلاث متغيرات في المكون المعلمي

حيث تم اعتبار ان هناك ثلاثة متغيرات تسلك سلوكا خطيا او معلميا (سلوكا مستقرا) مشكلتا المكون المعلمي ومتغير واحد يسلك سلوكا لاخطيا او لامعلميا (سلوكا غير مستقر) ليكون بدوره المكون اللامعلمي من النموذج PLM وحسب الصيغة (35) أو النموذج $GPLM$ وحسب الصيغة (44) وعلى هذا الاساس تم بناء عدد 4 نماذج لكل نوع وحسب الاتي:

1. اربعة نماذج PLM تتكون من ثلاثة متغيرات المكون المعلمي ومتغير واحد في المكون اللامعلمي وحسب الصيغ الرياضية التالية:-

- المكون المعلمي يمثل (X_1, X_2, X_3) واللامعلمي يمثل (X_4) وصيغته الرياضية الاتية:-

$$\hat{Y} = 0.47X_1 + 0.446X_2 + 0.101X_3 + m(X_4)$$

- المكون المعلمي يمثل (X_1, X_2, X_4) والمكون اللامعلمي يمثل (X_3) وصيغته الرياضية الاتية:-

$$\hat{Y} = 0.372X_1 + 0.344X_2 - 2.4X_4 + m(X_3)$$

- المكون المعلمي يمثل (X_1, X_3, X_4) والمكون اللامعلمي يمثل (X_2) وصيغته الرياضية الاتية:-

$$\hat{Y} = 0.455X_1 + 0.106X_3 - 2.15X_4 + m(X_2)$$

- المكون المعلمي يمثل (X_2, X_3, X_4) والمكون اللامعلمي يمثل (X_1) وصيغته الرياضية الاتية:-

$$\hat{Y} = 0.458X_2 + 0.099X_3 - 2.16X_4 + m(X_1)$$

2. اربعة نماذج $GPLM$ تتكون من ثلاثة متغيرات المكون المعلمي ومتغير واحد في المكون اللامعلمي وحسب الصيغ الرياضية الاتية:-

- المكون المعلمي يمثل (X_1, X_2, X_3) والمكون اللامعلمي يمثل (X_4) وصيغته الرياضية الاتية:-

$$\hat{Y} = G\{4.58E^{-5}X_1 + 4.05E^{-5}X_2 + 7.9E^{-6}X_3 + m(X_4)\}$$

- المكون المعلمي يمثل (X_1, X_2, X_4) والمكون اللامعلمي يمثل (X_3) وصيغته الرياضية الاتية:-

$$\hat{Y} = G\{3.49E^{-5}X_1 + 2.95E^{-5}X_2 - 0.19E^{-3}X_4 + m(X_3)\}$$

- المكون المعلمي يمثل (X_1, X_3, X_4) والمكون اللامعلمي يمثل (X_2) وصيغته الرياضية الاتية:-

$$\hat{Y} = G\{4.36E^{-5}X_1 + 8.37E^{-6}X_3 - 0.16E^{-3}X_4 + m(X_2)\}$$

- المكون المعلمي يمثل (X_2, X_3, X_4) والمكون اللامعلمي يمثل (X_1) وصيغته الرياضية الاتية:-

$$\hat{Y} = G\{4.14E^{-5}X_2 + 7.64E^{-6}X_3 - 0.18E^{-3}X_4 + m(X_1)\}$$

ومن النماذج اعلاه يمكن تحديد الملاحظات الاتية:-

- ان النماذج $GPLM$ غير كفوة لكون اوزان (معالم) المكون المعلمي قريبة للصفر في حين ان النماذج PLM تكون اكثر واقعية منها.

- ان معايير AIC, BIC تكون كبيرة جدا في النماذج $GPLM$ مقارنة بمثيلاتها في النماذج PLM .
- ان التشتت $Dispersion$ والانحرافات $Deviance$ تكون عالية جدا في نماذج PLM وقليلة في النماذج $GPLM$.

الجدول رقم (6) مقارنة بين مقاييس تمثيل البيانات لنماذج المجموعة الثالثة			
لنموذج	$R^2, adjR^2$ GLM, GLM_z	$R^2, adjR^2$ للنموذج $GPLM$	$R^2, adjR^2$ للنموذج PLM
GLM $R^2 = 0.4615$ $adjR^2 = 0.4597$		$R^2 = 0.4351$ $adjR^2 = 0.4299$	$R^2 = 0.4706$ $adjR^2 = 0.4658$
GLM_z $R^2 = 0.4265$ $adjR^2 = 0.4245$		$R^2 = 0.4847$ $adjR^2 = 0.4807$	$R^2 = 0.5135$ $adjR^2 = 0.5098$
		$R^2 = 0.4418$ $adjR^2 = 0.4387$	$R^2 = 0.4754$ $adjR^2 = 0.4725$
		$R^2 = 0.4335$ $adjR^2 = 0.4302$	$R^2 = 0.4635$ $adjR^2 = 0.4604$

• بملاحظة اوزان النماذج PLM مع اوزان الانموذج GLM ولكل انموذج على حدة يمكن بوضوح ملاحظة ان اوزان المتغير X_1 (الانتاج النفطي) قد انخفضت بصورة طفيفة في الانموذج الاول وبصورة أكبر في الانموذج الثاني ، وبصورة قليلة في الانموذج الثالث عن وزنه في الانموذج GLM ، اما المتغير X_2 (المستورد) فانه قد انخفض بصورة طفيفة في الانموذج الاول وبصورة أكبر في الانموذج الثاني ويكاد يكون متطابقا في الانموذج الثالث مع وزنه في الانموذج المعلمي ، اما المتغير الثالث X_3 (النقل البري) فان اوزانه للنماذج الثلاث قد بقيت قريبة جدا ارتفاعا او انخفاضاً من وزنه في الانموذج GLM ، اما المتغير الرابع X_4 (فرق العدادات) فقد بقي وزنه في كافة النماذج PLM اعلى من وزنه في الانموذج المعلمي.

• ان درجة تمثيل البيانات $R^2, adjR^2$ في النماذج PLM ، $GPLM$ وكما في الجدول الآتي :- وهو يوضح بقاء مقاييس تمثيل البيانات عند الانموذج PLM ولكافة النماذج أكبر من الانموذج المعلمي بشقيه الخطي والموسع ، اما الانموذج $GPLM$ فانه يتفوق فقط على الانموذج الموسع GLM_z في كافة المقاييس اما من جهة الانموذج الخطي GLM فانه يتفوق عليه عند الانموذج الثاني فقط .

(2-5) اختيار الانموذج الافضل

أن الغاية الرئيسية من اجراء المقارنة بين النماذج المعلمية وشبه المعلمية هي لتحديد الانموذج او النماذج الافضل التي تعبر عن بيانات نقل المنتجات النفطية داخل العراق ، فعند اجراء المقارنات بين كافة النماذج على اساس المعايير AIC, BIC, R^2 فان الانموذج حيث ستكون المقارنات بعدة اساليب Δ سيكون بمرحلتين حيث سيتم اجراء المقارنات بين النماذج المعلمية فيما بينها لاختيار الانموذج المعلمي الافضل ونفس الامر سيتم بالنسبة للنماذج شبه المعلمية فيما بينها كمرحلة اولى وفي المرحلة الثانية سيتم اجراء المقارنة بين الانموذجين الافضل لاختيار الافضل بينهما والذي سيكون $\hat{Y} = -2.666X_4 + m(X_1, X_2, X_3)$ والثاني هو الاسلوب الذي يعتمد على اختيار عدة نماذج شبه معلمية تحقق شرطا اساسيا وهو انها تملك اقل AIC, BIC من الانموذج المعلمي واختيار الافضل منها والذي سيكون $\hat{Y} = 0.36X_1 - 2.56X_4 + m(X_2, X_3)$ والثالث هو الاسلوب الذي سيعتمد على طريقة بناء النماذج حيث سيتم اختيار افضل انموذج من مجموعة النماذج التي مكوها المعلمي يتكون من متغير واحد ، ثم افضل انموذج من المجموعة الثانية ثم الثالثة وبالتالي سيتم الحصول على افضل ثلاثة نماذج والتي سيتم مقارنتها مع الانموذج المعلمي لاختيار الانموذج الافضل والذي سيكون :-

$$\hat{Y} = -2.666X_4 + m(X_1, X_2, X_3)$$

(3-5) تحسين درجة تمثيل البيانات للنماذج شبه المعلمية

تمت الاشارة المستمرة في ثنايا الموضوع الى المرونة التي يتمتع بها الانموذج شبه المعلمي من ناحية الامكانية في تحسين درجة تمثيله للبيانات ، وفي الحقيقة ان عملية تغيير هذه المقاييس تنتج نماذج جديدة ولكن بنفس المكونات المعلمية واللامعلمية اي ان خليط المتغيرات على اساس المكونين يبقى ثابتا وتغيير معالم الجزء المعلمي وممهدات الجزء اللامعلمي ، وان مفتاح هذا التغيير والعامل الاساس فيه هو تغيير قيمة المسافات البينية او عرض الحزمة $b.w$ حتى الوصول الى قيمة نهائية لا يمكن عندها زيادة درجة تمثيل البيانات ، وهذه النقطة ستمثل بالوصول الى اقصى درجة ممكنة من $adj R^2$ وعندها سنتمكن من اجراء مقارنة جديدة بين النماذج المحسنة لمعرفة افضل انموذج للبيانات والتعرف على مقاييسه الجديدة ، وفي الجدول التالي نماذج الانحدار الخطي الجزئي التي تم تحسين درجة تمثيلها للبيانات.

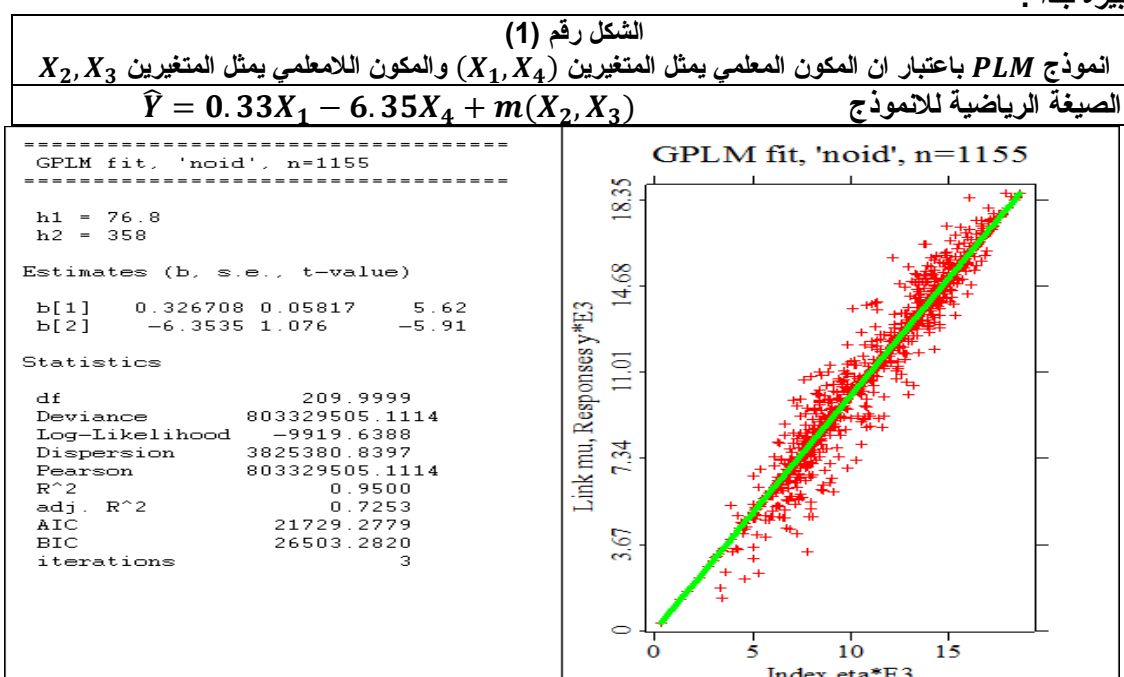
نماذج PLM التي تم تحسين درجة تمثيلها للبيانات (جدول رقم 7)					
تسلسل النماذج	R^2	$adjR^2$	AIC	BIC	المسافات البينية الجديدة
$M1$	0.9229	0.6847	21816.52	26226.22	$h_2 = 19, h_3 = 6890, h_4 = 233$
$M2$	0.8344	0.5844	21957.78	25470.11	$h_1 = 1500, h_3 = 190, h_4 = 300$
$M3$	0.8833	0.6121	21990.66	26071.91	$h_1 = 296, h_2 = 478, h_4 = 146$
$M4$	0.9213	0.6908	21782.21	26132.64	$h_1 = 4000, h_2 = 80, h_3 = 1370$
$M5$	0.8598	0.5910	22003.43	25839.62	$h_3 = 655, h_4 = 9$
$M6$	0.9447	0.6539	22022	26925.77	$h_2 = 9.6, h_4 = 90$
$M7$	0.9500	0.7253	21729.28	26503.28	$h_2 = 76.8, h_3 = 358$
$M8$	0.8373	0.5113	22220.61	26115.07	$h_1 = 700, h_4 = 5$
$M9$	0.9191	0.6087	22109.23	26738.75	$h_1 = 200, h_3 = 300$
$M10$	8500	0.5647	22073.36	25899.23	$h_1 = 145, h_2 = 300$
$M11$	0.6652	0.5023	21849.60	23763.14	$h_4 = 1.32$
$M12$	0.9034	0.5836	22152.07	26634.07	$h_3 = 7.1$
$M13$	0.9517	0.6605	22019.86	27025.44	$h_2 = 1.3$
$M14$	0.9032	0.5498	22261.77	26843.51	$h_1 = 3.61$

عند النظر للجدول (3-13) يمكن ملاحظة ما يأتي :-

- ان مقاييس R^2 ، $adj R^2$ قد ارتفعت في كافة النماذج شبه المعلمية عن مثيلاتها في النماذج نفسها التي تحوي نفس المكونين المعلمي واللامعلمي
- ان مقاييس AIC ، BIC قد ارتفعت بصورة واضحة لكافة النماذج عن مثيلاتها في النماذج نفسها التي تحوي المكونين المعلمي واللامعلمي نفسها وهذا الارتفاع يعد من الجزاءات التي يجب تحملها نتيجة تغيير عرض الحزمة في حين ان كل من مقاييس التشتت $Dispersion$ والانحرافات $Deviance$ قد انخفضت وبصورة ملحوظة.
- عند اجراء المقارنة بين النماذج الجديدة للانحدار الخطي الجزئي يظهر لنا أن النموذج M7 (الشكل 1) يمثل افضل نموذج معبر عن البيانات والذي سيكون بالصيغة:

$$\hat{Y} = 0.326X_1 - 6.353X_4 + m(X_2, X_3)$$

حيث تم سابقا شرح مكوناته المعلمية واللامعلمية مع ملاحظة ان وزن الانتاج النفطي بقي تقريبا على نفس المستوى وبانخفاض طفيف (من 0.359 الى 0.326) في حين ان وزن فرق العدادات قد ارتفع بصورة كبيرة مع التأثير السالب (من -2.564 الى -6.353) مما يعني ازدياد اهميته في النموذج بصورة كبيرة جدا .



6- الاستنتاجات

يمكن تحديد الاستنتاجات المستخلصة من البحث بالنقاط الآتية :-

1. ان المرونة التي يوفرها الانموذج شبه المعلمي في توصيف البيانات بصورة عامة تكون كبيرة جدا مقارنة بالانموذج الخطي ، ويزداد الامر وضوحا عندما تكون المشكلة او الظاهرة المدروسة تحوي على متغيرات كثيرة .
2. ان أنموذج الانحدار الخطي الجزئي الموسع $GPLM$ وبكافة تراكيبه من المتغيرات التي تسلك سلوكا خطيا مستقرا (المكون المعلمي) والتي لا تسلك السلوك الخطي (المكون اللامعلمي) على بيانات شركة خطوط الانابيب النفطية تكون غير كفوءة لكون اوزان المكون المعلمي ولكافة النماذج الرياضية تقترب للصفر ، ولعدة دوال ربط مختلفة من جهة ومن جهة اخرى ان درجة تمثيل البيانات بمعامل التحديد R^2 ، او معامل التحديد المصحح $adj R^2$ تكون قليلة جدا بالنسبة لدوال الربط الاخرى ، وعلى السياق نفسه فان الانموذج الخطي الموسع GLM_z ولمختلف دوال الربط يكون انموذجا غير كفوء ايضا وللاسباب نفسها .
3. ان أنموذج الانحدار الخطي الجزئي PLM وبكافة تراكيبه من المتغيرات التي تسلك سلوكا خطيا مستقرا (المكون المعلمي) ، والتي لا تسلك السلوك الخطي (المكون اللامعلمي) على بيانات شركة خطوط الانابيب النفطية كانت نماذج كفوءة بصورة مقبولة حسب المقاييس والاختبارات الاحصائية وعلى السياق نفسه فان الانموذج الخطي العام GLM كان انموذجا مقبولا ايضا .

4. عند اتباع اي اسلوب من اساليب المقارنة بين نماذج PLM فان الامودج الافضل او النماذج الافضل فيها تتفوق كلها على الامودج المعلمي GLM بمعايير على الاقل من معايير المقارنة الثلاث (AIC, BIC, R^2) مما يؤكد افضلية الامودج شبه المعلمي .
5. تم اثبات القدرة التي يتمتع بها الامودج شبه المعلمي PLM في تغيير درجة تمثيله للبيانات ($adj R^2, R^2$) وبصورة كبيرة جدا مع تقليل واضح وكبير لمقياس الانحرافات والتشتت وارتفاع بسيط في المعيارين (AIC, BIC)
6. كانت الخلاصة النهائية للنموذج الافضل الامودج السابع بالنسبة للنماذج شبه المعلمية بعد التحسين فان الامودج قد اصبح بالصيغة الرياضية:-

$$\hat{Y} = 0.33X_1 - 6.35X_4 + m(X_2, X_3)$$

وحسب (الشكل 1) ويتغير عرض الحزمة bw لتصبح ($h_2 = 76.8, h_3 = 358$) بعدما كانت ($h_2 = 2577.6, h_3 = 8684.5$) وكلا الصيغتين توضح عدم الاستقرار للمتغيرين (المستورد ، النقل البري) ، مما يمكن تفسيره بعدم وجود سياسة مستقرة وواضحة لعملية استيراد المنتجات البيضاء او لعملية نقلها برا لنقاط التوزيع او التخزين ، اما بالنسبة للمتغيرين الاخرين (الانتاج النفطي ، فرق العدادات) فان الاستقرار بالنسبة للمتغير الاول (الانتاج النفطي) قد تكون مؤشرا ايجابيا من جهة ومن جهة اخرى مؤشرا سلبيا فالمؤشر الايجابي يوضح استمرارية المنتج من قبل المصافي الجنوبية بوتيرة منتظمة او مقبولة في ظل الظروف الاستثنائية التي يمر به وطننا العراق ، اما المؤشر السلبي فهو يوضح وعلى اقل تقدير في السنوات الثلاث الماضية لم تجر محاولة جادة لتطوير كفاءة هذه المصافي ودراسة اسباب القصور فيها ، اما متغير فرق العدادات فمن الواضح ان الامودج شبه المعلمي والمعدل قد كشف عن مقدار خطورة هذا المتغير على العملية بمجملها ويمكن تحديد عدة تفسيرات لخطورة هذا المتغير منها ما هو قصور في الاداء الوظيفي عند عدادى دفع المنتج واستقباله او الاعطال الفنية في هذه العدادات او خلل في طبيعة المنتج المدفوع في الانابيب من حيث لزوجته او وجود الغازات والابخرة فيه مما يؤدي الى قراءات خاطئة او السرقات المنتظمة من الانابيب وبكميات منتظمة وقليلة نسبيا قياسا بضخامة المنتج .

المصادر

1. خمو،خلود يوسف(2004)، "مقارنة اساليب بيز مع طرائق اخرى لتقدير منحني الانحدار اللامعلمي"، اطروحة دكتوراه في الاحصاء، كلية الادارة والاقتصاد/ جامعة بغداد
2. كاظم ،أموري هادي والدليمي ،محمد مناجد عيفان ،(1988)،"مقدمة في تحليل الانحدار الخطي"،كلية الادارة والاقتصاد / جامعة بغداد.
3. Draper, N. R., & Smith, H. (1981), "Applied Regression Analysis" John Wiley and sons, 2nd. Ed., New York.
4. Hardle, W " Applied Nonparametric Regression"(1990). Cambridge University Press. Cambridge
5. Hardle, W & Liang, H & Gao, J "Partilly Linear Models"(2000) Heidelberg: Physica-Verlag
6. Hardle, W, Müller. M, Sperlich, S & Werwatz, A "Nonparametric and Semi parametric Models an Introduction " (2004) , Springer Edition
7. Hastie, Trevor & Tibshirani, Robert "Generalized Aditive Model"(1990), Chapman and Hall, London.
8. Hansen, Bruce E, "Econometrics"(2012) University of Wisconsin
9. Lindsey, James K, "Applying Generalized Linear Model" (1997), Springer.
10. Muller, Marlene, "Semiparametric Extension GLM_Z" (2000), der Humboldt – Universitat zu Berlin.
11. McCullagh, P., and Nelder, J. A" Generalized Linear Models", (1989). 2nd ed. London: Chapman& Hall.
12. Ruppert, D., Wand, M. and Carroll, R. (2003). "Semiparametric Regression". Cambridge University Press. Cambridge.
13. Scott, D. W, " Multivariate Density Estimation", (1992), John Wiley & Sons, New York
14. Yatchew, Adonis "SPR for The Applied Econometric"(2003), Cambridge University Press.
15. Demir . S , Toktamis.O," ON THE ADAPTIVE NADARAYA-WATSON KERNEL REGRESSION ESTIMATORS" ,Journal of Mathematics and Statistics Volume 39 (3) (2010), 429 – 437

16. ENGLE, R. F., GRANGER, C. W. J., RICE, J. and WEISS, A. (1986). "Semiparametric estimates of the relation between weather and electricity" sales. Journal of the American Statistical Association, 81:394, 310-320
17. Flashier, Emmanuel., "Non-Parametric Econometrics" www.vcharite.univ-mrs.fr/.../nonparametric-chap3.pdf.
18. Muller.M," Estimation and testing in generalized partial linear models—A comparative study", (2001), Statistics and Computing 11: 299–309
19. NELDER .J.A, WEDDERBURN.W.M," Generalized Linear Models", Journal of the Royal Statistical Society. Series A (General), Vol. 135, No. 3 (1972), pp.370-384.
20. Robinson. P. M , "Root-N-Consistent Semiparametric Regression", Econometrica, Vol. 56, No. 4 (Jul., 1988), pp. 931-954.
21. Speckman .P ,"Kernel Smoothing in Partial Linear Models " Journal of the Royal Statistical Society. Series B (1988), Vol.50, No.3, pp.413-436.
22. Van Kerm, Philippe "Adaptive kernel density estimation", The Stata Journal (2003) 3, Number 2, pp. 148–156.

.....
.....
.....