مقارنة بين بعض طرائق الانحدار شبه المعلمي و اللامعلمي لنموذج البيانات الطولية العشوائي

الباحث احمد عبد الصمد حبيب

أ.د قتيبة نبيل نايف

كلية الإدارة والاقتصاد – جامعة بغداد

قسم الإحصاء

مقارنة بين بعض طرائق الانحدار شبه المعلمي و اللامعلمي لنموذج البيانات الطولية العشوائي

Comparison of some Non -parametric and Semi-parametric regression methods for balanced Random Effect Panel data (longitudinal) Model

Ahmed Abdulsamad Habeeb
Prof.Dr. Qutaiba N. Nayef Al-Kazaz
University of Baghdad /College of Administration&Economics/Statistics Department

Abstract:

This research aims to make a comparison between the non-parametric and semi-parametric method to estimate the random effects balanced panel data (longitudinal) model.

In order to achieve the purpose of the research, two methods: Local Linear Polynomial and the Nadaraya - Watson were used in the nonparametric estimation. For the semi-parametric estimation, the Profile Least Square method was used. It was concluded that the semi-parametric method gave better results than the non-parametric methods for estimating the random effect balanced Panel data (longitudinal) model and the sample sizes (n=50,100,200), and variances ($\sigma^2 = 0.5, 1, 1.5$).

Simulation experiments were formulated and the applying of the methods used in this research was verified using the average sum of squares error criterion for different sample sizes and levels of variance.

المستخلص

يهدف هذا البحث الى أجراء مقارنة بين الأسلوب اللامعلمي وشبه المعلمي لتقدير أنموذج البيانات الطولية المتزنة ذي التأثيرات العشوائية وفي سبيل تحقيق هدف البحث تم إستعمال طريقتي التقدير الموضعي الخطي الموضعي متعدد الحدود (Local Linear Polynomial) وطريقة نداريا− واتسون (Nadaraya &Watson) في التقدير اللامعلمي , وفي التقدير شبه المعلمي تم إستعمال طريقة بروفايل (Profile Least Square) , وتم التوصل الى ان الطريقة شبه المعلمية اعطت نتائج أفضل من الطرائق اللامعلمية في تقدير أنموذج البيانات الطولية المتزنة العشوائي ولأحجام العينات (n=50,100,200) وانواع التباينات ($\sigma^2 = 0.5, 1, 1.5$).

وتمت صياغة تجارب المحاكاة والتحقق من أداء الطرائق المستعملة في هذا البحث وذلك بإستعمال معيار معدل مجموع مربعات الخطأ .

المصطلحات الاساسية للبحث :-

طريقة التقدير الموضعي الخطي متعدد الحدود , طريقة نداريا –واتسون , طريقة بروفايل Profile Least Square , البيانات الطولية المتزنة , التقدير شبه المعلمي , التقدير اللامعلمي . أنموذج البيانات الطولية العشوائي .

1- المقدمة : (Introduction)

احدى الغايات الرئيسة لدراسة أي مشكلة أو ظاهرة سواء كانت (اقتصادية أم اجتماعية أم علمية) هو ايجاد الانموذج المناسب الذي يمثلها والذي من خلاله نتوصل الى فهم هذه الظاهرة وتحديد معالمها الأساسية وهو ما يعرف في علم الاحصاء بنمذجة الظواهر ، وحيث أن هذه العملية قد تطورت بمرور الزمن الا انها لمدة طويلة اعتمدت على نوع واحد هو الانحدار المعلمي (Parametric Regression) اذ تم تطوير كل ما يتعلق بهذا النوع من تقديرات وفرضيات واختبارات ، لكن لوحظ أن هذا النوع قد لايمثل الظاهرة قيد الدراسة تمثيلا ملائما في بعض الجوانب التطبيقية وذلك لاتخاذ بعض المتغيرات سلوكا معلميا والاخر لا معلميا ، وعدم الأخذ بنظر الاعتبار التأثيرات اللاخطية للمتغيرات التوضيحية على متغير الاستجابة , ومن ثم يمكن أن تكون التقديرات الناتجة لهذا النوع من النماذج مضللة لذا ظهر الى العلن فكر احصائي جديد والذي لايعتمد على القيود الصارمة المفروضة في الانحدار المعلمي وهذا الفكر اطلق عليه الانحدار اللامعلمي (nonparametric Regressioin) ، والذي يتميز بمرونته العالية ، ويعتمد على تمهيد البيانات بإستعمال الأوزان الثابتة ومن ثم تطور هذا النوع عن طريق إستعمال واحدة من اشهر الدوال الوزنية هي الدالة اللبية (kernel function) والتي بإستعمالها تم تمهيد البيانات على اساس الأوزان غير الثابتة ، وشكلت بواسطتها مجموعة من المقدرات الإحصائية والتي تعرف باسم مقدرات الانحدار المعلمي مثل مقدرات (نداريا– واتسون ،برستل جاو، الشرائح المقطعية، ممهد الانحدار الخطي الموضعي) . ولكن في الوقت نفسه فإن الانحدار اللامعلمي يعاني من ما يسمى بمشكلة تعدد الأبعاد (The course of dimensionality) التي تحدث عند تزايد عدد المتغيرات التوضيحية الداخلة في

التحليل مما يؤدي إلى تناقص ثقة المقدر اللامعلمي حتى بدأت تتشكل ملامح الجيل الثالث من نمذجة الظواهر عند منتصف الثمانينات، وهو جيل الانحدار شبه المعلمي إذ تأتي الميزة الكبيرة لهذا الانموذج في انه يحوي كل المميزات الايجابية التي يتضمنها الانموذجان السابقان مثل إحتواءه على القيود الصارمة في جزأه المعلمي ومرونته الكاملة في جزئه اللامعلمي ولوضوح التفاعل بين مكوناته المعلمية واللامعلمية والتي لاقت قبولا واسعا في الدراسات الاقتصادية والاجتماعية وغيرها من الدراسات الحديثة ، وذلك بسبب المرونة العالية التي يتمتع بها هذا النوع من الانحدار وحله لمشكلة السلوك غير المفهوم لبعض المتغيرات الداخلة في الدراسة إذ أن فكرته تأتي من فكرة النماذج التجميعية حيث يتم الدمج بين المكونين المعلمي واللامعلمي في هذا الانموذج .

2- مفهوم وتعريف البيانات الطولية (Panel Data):[1] [2] [4] [17]

وتعرف على انها البيانات التي يمكن الحصول عليها من خلال المشاهدات المكررة (Repeated Observations) لظاهرة ماحول (N) من المقاطع العرضية (Cross- Sections) خلال سلسلة زمنية (T) (Time series) .

ومن التعريف المذكور آنفا فأن الظاهرة المدروسة تتغير على المستوى الأفقي فتكون البيانات المقطعية (Cross-sections data)، و تتغير على المستوى العمودي فتكون البيانات السلسلة الزمنية (Time Series Data)، وممكن قراءة بيانات فترة من فترات السلسلة الزمنية لكل المقاطع العرضية ،أو قراءة بيانات مقطع من المقاطع العرضية لكل فترات السلسلة الزمنية , وبشكل عام يمكننا التعبير عن مجموعة من البيانات الطولية (panel data) كما في الترتيب الموضح بالجدول (1) الآتي :

الجدول (1) البيانات الطولية panel data

المتغيرات التوضيحية x_{ijr}	المتغير المعتمد y_{ij}	السلسلة الزمنية	المقطع العرضي
		1	1
		2	1
			1
		1	2
		2	2
			2
		1	
		2	

إذ يمثل :

y_{ij} : قيمة المتغير المعتمد في الوحدة المقطعية i عند المدة الزمنية j .

x_{ijr} : المتغيرات التوضيحية للوحدة المقطعية i عند المدة الزمنية j .

$i = 1,2, \dots\dots, n$

n : تمثل عدد المشاهدات في الوحدات المقطعية .

t_i : تمثل المدة الزمنية في الوحدات المقطعية .

q : عدد المتغيرات التوضيحية .

3-الانحدار اللامعلمي : (Non-Parametric Regression)[8] [13]

يعرف الانحدار اللامعلمي على انه توفيق منحنيات المتغير المعتمد من خلال وصف نمط المتغيرات التوضيحية قيد الدراسة دون اللجوء الى تقدير معلمات الانموذج المدروس , ومن ثم يهدف الانحدار اللامعلمي الى إيجاد افضل منحنى تمهيدي يطابق او يقترب من التطابق لمنحنى المتغير المعتمد , يوصف انموذج الانحدار اللامعلمي رياضيا بصورة عامة كمايلي:

$$Y =$$

إذ تمثل :

Y : قيمة المتغير المعتمد

$m(X)$: دالة تمهيد مجهولة

E : يمثل الخطأ العشوائي

وان الهدف الرئيس من الانحدار اللامعلمي هو تقدير الدالة المجهولة $m(X)$, إذ ان :

$$m(X) = E(Y$$

التمهيد اللامعلمي (Non-Parametric Smoothing)(NPS) هو المدخل الرئيس للانحدار اللامعلمي ويتضمن توظيف طرائق التقدير اللامعلمية المختلفة وأساليب تمهيد البيانات , ومن أهم تلك الأساليب وأكثرها إستعمالاً هو الدوال اللبية (Kernel Functions) والتي تعرف على إنها دوال وزنية واسعة الإستعمال تمتاز بالمرونة في عملية التقدير تم إقتراحها من (Rosenblatt) عام (1952) وتم تطويرها من (Parzen) و (Epanechnikove) , وتمتاز بكونها دوالاً

إحتمالية حقيقية مستمرة ومتماثلة حول الصفر , تعمل الدوال اللبية على إعادة تعديل البيانات للحصول على مقدرات متقاربة مع صفات المعلمات الحقيقية , وتعرف الدوال اللبية رياضياً من خلال الصيغة التالية :

$$p(a \leq u \leq b)$$

$$=$$

كما وتمتلك الدوال اللبية الخصائص العامة الآتية :

1− الاستمرارية (Continuity) أي أن $0 \leq k(u) \leq \infty$.

2− التماثل (Symmetric) , أي أن $k(u) = -k(u)$.

3− دالة الكثافة الإحتمالية (Probability Density Function) , $\int k(u)du = 1$.

4− لها تباين معلوم ومتوسط مساوٍ للصفر , أي :

$$E(u) = \int u \;\; k$$

$$E(u^2) = \int u^2$$

وإن الصيغة الرياضية لتقدير دالة الكثافة الإحتمالية $f(x)$ للمتغير العشوائي x بإستعمال الدوال اللبية تعرف بالشكل الآتي :

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n}$$

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n}$$

$$=$$

إذ تمثل :

$k(u)$: دالة التمهيد اللبية .

h : عرض الحزمة أو معلمة التمهيد .

هناك العديد من أنواع الدوال اللبية تختلف من حيث صيغتها وحدودها ودواعي إستعمالاتها والجدول (2) الآتي يتضمن بعض صيغ الدول اللبية (kernel) من ضمنها الصيغ شائعة الإستعمال في التمهيد اللامعلمي وهي (Gaussian) و (Epanechnikov) :

الجدول (2) الدوال اللبية المستعملة في البحث (المصدر(5))

Kernel Function		حدود الدالة
Gaussian		
Epanchnikove		

3−1−1 عرض الحزمة : (Bandwidth) [8]

من المسائل المهمة عند إستعمال الدوال اللبية هي ما يعرف بعرض الحزمة أو بمعلمة التمهيد (Smoothing Parameter) ويرمز له عادة بالرمز h اذ تملك معلمة التمهيد تأثيراً قوياً في نتائج الدوال اللبية ومن ثم فأن أي مقدر من مقدرات التمهيد اللامعلمي اللبي سيكون حساساً جداً تجاه التغيرات في قيمة (h) والذي سيؤدي بدوره الى تغيير شكل منحنى التمهيد، وقد بين (Flashier) بأنه يجب مراعاة حقيقتين في إختيار معلمة التمهيد:الحقيقة الأولى هي أنه عند إختيار قيمة عالية لمعلمة التمهيد فإن ذلك سيؤدي الى الحصول على تمهيد دقيق للبيانات المدروسة مع تحيز كبير وتباين قليل , والحقيقة الثانية هي أنه عند إختيار معلمة تمهيد ذات قيمة صغيرة سيؤدي الى تمهيد اقل دقة للبيانات المدروسة مع تحيز قليل وتباين كبير .

عادة تكون الخطوة الأولى لتقدير عرض الحزمة هي تحديد معيار الخطأ المناسب للمقدر $\hat{m}(x)$, فإذا كان معيار الخطأ هو مربعات الخطأ التكاملية Integral Square Error(ISE) فإن :

ISE

الهدف هنا هو جعل الخطأ المقدر (ISE) أقل ما يمكن , ولما كان (ISE) للدالة $m(x)$ يعتمد على العينة $(x_1, y_1), \dots, (x_n, y_n)$ لذلك يتم استعمال المعيار Mean Integral Square Error (MISE) في هذه الحالة وكما يلي :

$MISE\{\hat{m}$

$= \int$

$=$

عندها فإن عرض الحزمة هو المقدار الذي يجعل متوسط مربعات الخطأ التكاملية (MISE) أقل ما يمكن .

$h_{MISE} =$

وهناك عدة طرائق وأساليب لتقدير أو إختيار عرض الحزمة نذكر منها قاعدة التوزيع الطبيعي أو قاعدة الإبهام (Rule of Thumb), قاعدة الملء المباشر (Plug in Rule) , اسلوب (Bootstrap Technique) وطريقة العبور الشرعي (Cross Validation Method) .

2−1−3 طريقة العبور الشرعي : (Cross Validation Method)[8]

تعد هذه الطريقة من أدق الصيغ وأكثرها استعمالا في تقدير معلمة التمهيد , اقترحت هذه الطريقة من قبل (Rudemo) عام (1982م) وتم تطويرها من قبل (Scott) عام (1987م) , وتعتمد هذه

الطريقة في ايجاد معلمة التمهيد على عدة تقنيات وأساليب منها طريقة المربعات الصغرى , طريقة الامكان الأعظم وطريقة العبور الشرعي المتحيز . في هذا البحث سيتم توظيف طريقة المربعات الصغرى للعبور الشرعي (Least Square Cross Validation)(LSCV) التي تعمل على جعل متوسط مربعات الخطأ التكاملية (Mean Integrated Squared Error)(MISE) اقل ما يمكن , وتعتمد على إستبعاد قيمة واحدة من قيم المشاهدات في كل مرة (Leave – One –Out) , ويمكن تلخيص خطوات التقدير على وفق طريقة (LSCV) للبيانات الطولية بإفتراض قيم أولية لمعلمية التمهيد (h) , ومن ثم حذف المشاهدة x_i ويتم ذلك من خلال صيغة دالة المربعات الصغرى للعبور الشرعي الآتية :

$LSCV$(

إذ أن :

$\hat{m}_{(-ij)}(x_{ij})$: تمثل دالة التقدير لطريقة المربعات الصغرى

h : معلمة التمهيد

يتم تكرار حساب الصيغة (5) عندها يتم اختيار معلمة التمهيد التي تقابل اصغر $LSCV(h)$ أي :

$$\hat{h}_{LSCV} = argmin_h LSCV(h)$$

وفي حالة وجود أكثر من متغير توضيحي (x_1, x_2) فإن تقدير عرض الحزمة

على وفق طريقة المربعات الصغرى للعبور الشرعي يكون على النحو التالي :

$$LSCV(h) = \frac{1}{nt}\sum_{i}$$
$$=$$

وبتكرار الصيغة (6) فانه يتم اختيار معلمات التمهيد التي تقابل اصغر $LSCV(h)$, وكما يلي :

$$\hat{h}_{LSCV} = argmin_{h_1 \ldots}$$

3-2 إنموذج البيانات الطولية اللامعلمي ذو التأثيرات العشوائية : (9) (13) (14)

(Non-Parametric Panel data Random Effect Model)

يعرف إنموذج البيانات الطولية (Panel data) اللامعلمي ذو التأثيرات العشوائية على وفق الصيغة الرياضية الآتية:

$$\boldsymbol{y_{ij}} =$$

$$\boldsymbol{\varphi_{ij} = u_i + v_{ij}}$$

إذ أن :

$\boldsymbol{y_{ij}}$: تمثل متجه بالابعاد $(nt \times 1)$, يمثل قيمة المتغير المعتمد في الوحدة المقطعية i عند الفترة الزمنية j .

$m(.)$: دالة تمهيد مجهولة يتم تقديرها بالطرائق اللامعلمية , وأن $m(x) = E(y_{ij} \backslash x_{ij} = x)$

u_i : الخطأ في الوحدة المقطعية i

v_{ij} : متجه الخطأ العشوائي بالأبعاد $t \times 1$.

وفي ظل الشروط والافتراضات الآتية :

$$E\left(v_{ij} \backslash x_{ij}, u_i\right)$$
$$E\left(v_{ij} v'_{ij} \backslash x_{ij}, u\right.$$
$$E\left(u_i \backslash x_{ij}\right) = 0$$
$$E\left(u_i^2 \backslash x_{ij}\right) = \sigma$$

وإن مصفوفة التباين والتباين المشترك للخطأ المركب $\Sigma = E(\varphi\varphi')$ لجميع الفترات الزمنية (t) للمقطع (i) ستكون :

$$\Sigma = \sigma_u^2 \, I_t + \sigma_v^2$$
$$\Sigma = I_{nt} \otimes \left[\left(\sigma \right.\right.$$
$$\sigma_1^2 = t\sigma_u^2 + \sigma_v^2$$

3-2-1 طرائق تقدير إنموذج البيانات الطولية اللامعلمي ذي التأثيرات العشوائية :

(Non-Parametric Panel data Random Effect Model Estimation Methods)

تتمتع طرائق التقدير اللامعلمية التي تعتمد على الداول اللبية (kernel) بالدقة فضلاً عن إمكانية استعمالها مع التصميم الثابت والعشوائي , نذكر منها ط يقة نداريا واتسون (Nadaraya & Watson) وطريقة التقدير الخطي الموضعي متعدد الحدود (Local Linear Polynomial Estimation) .

3-2-1-1 طريقة التقدير ندارايا و واتسون [8]

(Nadaraya & Watson Estimation Method)

يعد مقدر ندارايا واتسون من المقدرات اللامعلمية ذات التطبيق الواسع

في الإنحدار اللامعلمي اقترحه كلا الباحثين (**Nadaraya & Watson**) عام (1994م) , إذ طور الباحثان طريقة تعتمد على فكرة الرسم البياني في التقدير قدمت مسبقا من (Tukey) في عام (1961م) تسمى (Regressogram).

يتم تقدير دالة التمهيد $m(x_{ij})$ للأنموذج العشوائي اللامعلمي للبيانات الطولية (Panel data) والمعرف في الصيغة (8) بطريقة (NW) على وفق الصيغة التالية :

$$\hat{m}(x_{ij})_{NW} = \frac{\sum}{}$$

2-1-2-3 طريقة التقدير الموضعي الخطي متعدد الحدود :[3] (13)(8)

(Local Linear Polynomial Estimation Method)

قدم كل من (Fan) و (Gijbels) في عام (1992م) اسلوباً موسعاً للتقدير اللبي يسمى بالإنحدار الموضعي الخطي متعدد الحدود (Local Linear Polynomial Regression)إن هذا الإسلوب يعمل على تصحيح بعض العيوب في المقدر اللبي مع إمكانية حالة التعميم على الإنحدار المتعدد , ويعتمد هذا الإسلوب على تعميم المقدر اللبي الى حالة مواءمة متعدد الحدود عند النقطة (x) , ويمتاز بالقابلية على التكيف مع طبيعة الإنموذج المدروس سواء كان الإنموذج ثابتاً بمعلمة تمهيد واحدة أو عشوائياً بمعلمة تمهيد متغيرة , كما يمكن أن تكون الممهدات الموضعية من أكثر الممهدات ملاءمة للأنواع المختلفة من الدوال اللبية وعرض الحزم المستعملة في التقدير, لذلك تصنف بأنها ممهدات ذات كفاءة عالية.

إن التقدير الموضعي الخطي متعدد الحدود لإنموذج البيانات الطولية (Panel data) ذي التأثير العشوائي المعرف في الصيغة (8) يعتمد على متسلسلة تايلور وكما هو موضح في الصيغة التالية :

$$m(x_{ij}) \approx m$$

وبإجراء $(p+1)$ من المشتقات المتسلسلة للدالة المراد تقديرها $m(x_{ij})$ نحصل على :

$$m^{(1)}(x_{ij}) = m$$
$$+$$
$$m^{(2)}(x_{ij}) = 2m_2$$
$$-$$
$$\vdots$$
$$m^{p}(x_{ij}) = p!\,$$

وبإفتراض أن :

$$m(x_{ij}) = m_0($$
$$m^{(1)}(x_{ij}) = m$$
$$m^{(2)}(x_{ij}) = 2$$
$$\vdots$$
$$m^{(p)}(x_{ij}) = p!$$

بتعويض $\delta_0, \delta_1, \delta_2, \dots, \delta_p$ بدلاً من $m_0(x), m_1(x), m_2(x), \dots, m_p(x)$ على الترتيب فإن الصيغة (10) تصبح بالشكل الآتي :

$$m(x_{ij}) \approx \delta_0 + \ldots$$
$$+ \ldots$$

أو :

$$m(x_{ij}) = \sum \ldots$$

وبالتعويض عن $m(x_{ij})$ في إنموذج البيانات الطولية (Panel data) المعرف في الصيغة (8) فإن :

$$y_{ij} = \sum_{s=0}^{p} \delta_s (x_{i} \ldots$$
$$= \ldots$$

ومن خلال تحديد قيمة (p) فإنه يمكن إيجاد مقدرات الإنحدار الموضعي متعدد الحدود المختلفة في ظل الشروط والإفتراضات التي يضعها إنموذج البيانات الطولية (panel data) ذو التأثيرات العشوائية وذلك بإستعمال طريقة المربعات الصغرى الموزونة (Wieghted Least Square)(WLS) .

ويمكن الحصول على المقدر الموضعي الخطي من الدرجة الأولى إذا كانت (p=1) عندها فإن إنموذج البيانات الطولية (Panel data) ذو التأثيرات العشوائية المعرف في الصيغة (12)سيكون على وفق الصيغة التالية :

$$m(x_{ij}) \approx \delta_0 + \delta_1(x_{i} \ldots$$
$$y_{ij} = \delta_0 + \delta_1(x_{ij} - \ldots$$
$$\varphi_{ij} = x_{ij} - \delta_0 \ldots$$
$$- \ldots$$
$$- \ldots$$

وبتطبيق طريقة المربعات الصغرى الموزونة فإن مجموع مربعات الخطأ للإنموذج العشوائي تعرف بالشكل الآتي :

$$\sum_{i=1}^{n}\sum_{j=1}^{t}\varphi_{ij}^{2}K_{h}\big(x_{i}$$

$$=$$

$$-$$

يمكن كتابة الصيغة (14) بدلالة المصفوفات وكما يلي :

$$\varphi' K\varphi = \{Y -$$

$$-$$

$$\varphi = \begin{bmatrix}\varphi_1' \\ \varphi_2' \\ \vdots \\ \varphi_n'\end{bmatrix}_{nt\times 1} , \quad Y$$

$$\delta = \begin{bmatrix}\delta_0 \\ \delta_1\end{bmatrix}_{(p+1)\times 1}$$

$$diag(K) = \big[K$$

لإيجاد تقدير للمتجه δ يتم تقليل مجموع مربعات الخطأ بإشتقاق الصيغة (15) بالنسبة الى δ ومساواة المشتقة بالصفر لنحصل على المقدر الموضعي الخطي لإنموذج بيانات البانل داتا (Panel data) ذي التأثيرات العشوائية من الدرجة الاولى الآتي:

$$\frac{\partial\varphi' K\varphi}{\partial\delta} = -DK$$

$$\widehat{m}(x_{ij})_{LLL}$$

$$= I'(D'K$$

I : تمثل متجه بالأبعاد $[(1+p)\times 1]$, تكون أول (p) من عناصر مساوية الى 1 وبقية عناصره أصفاراً .

من الجدير بالذكر أن (Nadaraya & Watson)(NW) ماهو إلا حالة

خاصة من المقدر الموضعي الخطي متعدد الحدود وذلك عندما $(p = 0)$ وبذلك تكون $m(x_{ij}) \approx \delta_0$, ويسمى عندها بالمقدر الثابت الموضعي (Local Constant Estimator).

إذ يمكن توضيح طريقة إيجاد المقدر الثابت الموضعي أو مقدر (NW) لإنموذج البيانات الطولية (Panel data) ذي التأثيرات العشوائية المعرف في الصيغة (13) من خلال توظيف طريقة المربعات الصغرى الموزونة (WLS) بالدالة اللبية K_h كما يلي :

$$y_{ij} = \delta_0 + \varphi_{ij}$$

$$\varphi_{ij} = y_{ij} - \delta_0$$

وبتربيع الصيغة (17) وضرب طرفيها بالدالة اللبية $K_h(x_{ij} - x_i)$, ومن ثم إدخال المجموع بالنسبة لكل من (i, j) نحصل على :

$$\sum_{i=1}^{n}\sum_{j=1}^{t}\varphi_{ij}^{2}$$

وبإشتقاق الصيغة (18) بالنسبة الى δ_0 ومساواة الناتج بالصفر نحصل على صيغة الحالة الخاصة لمقدر (NW) لإنموذج بيانات البانل داتا (Panel data) ذي التأثير العشوائي :

$$\hat{m}(x_{ij})_{NW} = \hat{\delta} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{t} y_{i}}{\sum_{i=1}^{n}\sum_{j=1}^{t}}$$

4- الانحدار شبه المعلمي :Semiparametric Regression [13]

(16)(14)

يعد أنموذج الانحدار الخطي الجزئي (Partially Linear Model PLM) أحد نماذج الانحدار شبه المعلمية ، وهو من النماذج التي تعتمد على متغيرات خطية (Linear) ومتغيرات غير خطية لا معلمية (nonparametric) وهي عادة ما تكون مستمرة ، وتؤثر هذه المتغيرات الخطية وغير الخطية في متغير الاستجابة Y ، وهو يعد أيضا حالة خاصة من النماذج التجميعية Additive modeles لذلك يكون افضل من النماذج المعلمية لأنه يتجنب مشكلة الأبعاد في النماذج المعلمية وهو اكثر مرونة من النماذج الخطية القياسية لأنها تقلل من الافتراضات الخطية المفروضة على هذه النماذج ، وله عدة تسميات منها ، أنموذج الانحدار شبه المعلمي (SPRM) او الانموذج الخطي الجزئي (PLM) سبب تسميته بالخطي الجزئي كونه يتضمن جزءاً لا معلمي غير خطي وجزءاً معلمياً خطياً وكما يلي :

$$y_i = \ldots$$

1-4 الأنموذج الخطي الجزئي ذو التأثيرات العشوائية للبيانات الطولية Partially Linear With Random Effects for Panel data Models (REPLM) (14)(13)

يمكن تعميم النموذج (20) ليكون ملائما للبيانات المزدوجة كما يلي :

$$y_{it} = x_{it}' \ldots$$

y_{it} : متجه متغير الاستجابة من الدرجة nx1

x_{it} : مصفوفة المتغيرات التوضيحية المعلمية من الدرجة n x p

β : متجه المعالم المجهولة ذو مرتبة px1

$x'B$: الجزء المعلمي للانموذج قيد الدراسة

z_{it}: متغير مستمر ويمثل المتغير اللامعلمي في البيانات من الدرجة n x1

$m(z_{it})$: الجزء اللامعلمي يمثل دالة تمهيدية من الدرجة nx1

ε_{it}: متجه الأخطاء العشوائية المشتركة للسلاسل الزمنية والمقاطع العرضية ذو مرتبة nx1 وتكون مستقلة بمتوسط صفر وتباين σ^2

α_i : الحد الثابت والتغير فيه يحدد نوع النموذج اذا كان ثابتاً او عشوائياً فاذا كان يتغير بأسلوب ثابت يكون النموذج ذات تأثيرات ثابتة و اذا كان يتغير بأسلوب عشوائي يكون النموذج ذات تأثيرات عشوائية .

ولذلك يسمى الانموذج العشوائي أنموذج الخطأ المركب فيمكن إعادة كتابة النموذج (21) كما يلي :

$$y_{it} = \alpha +$$

$$y_{it} = \alpha +$$

إذ أن :

μ_i : اذا كان التغير أما على مستوى السلاسل الزمنية او المقاطع العرضية

w_i :يمثل الخطأ العشوائي المركب .

2-4 طرائق تقدير انموذج البيانات الطولية الخطي الجزئي ذي التأثيرات العشوائية Partially Linear Panel data Model Estimatoin Methods

1-2-4 طريقة التقدير بروفايل (Profile Least Square Method) (6)(13)

تم اقتراح هذه الطريقة من قبل (Robinson) عام (1988) لتقدير الانموذج الخطي الجزئي وتعتمد على تحويل الانموذج شبه المعلمي الى انموذج لامعلمي وذلك بطرح الجزء المعلمي من الانموذج شبه المعلمي , ثم جرى تطوير هذه الطريقة لتلائم البيانات الطولية (panel data) وسميت من قبل (Fan) و (Huang) عام 2005 بـ (PLLLS) (Profile Least Square Method)
بالامكان تطبيق هذه الطريقة على النموذج الخطي الجزئي للبيانات المزدوجة (panel data) باتباع الخطوات الآتية :

$$Y_{it} =$$

$$Y_{it} - X_{it}^T\beta = m$$

$$Y_{it}^* = Y_{it} - X_{it}^T$$

$$Y_{it}^* = m(Z_{it}) + w_i \quad \dots (25)$$

وبالامكان تقدير دالة التمهيد المجهولة باستعمال التقدير الخطي الموضعي متعدد الحدود على وفق الطرائق المتبعة في التقدير اللامعلمي , وذلك بافتراض أن θ_0 و θ_1 تمثل الحلول الناتجة بعد تصغير مجموع مربعات الخطأ للانموذج شبه المعلمي وكما يلي :

$$\sum_{i=1}^{N}\sum_{t=1}^{T}[(y_{it} - $$

$$-$$

علما أن $\hat{\theta}' = (\hat{\theta}_0, \hat{\theta}_1)$ تمثل تقديرات لكل من $m_0(z_{it})$ و $m_1(z_{it})$ على التوالي بحيث أن :

$$\hat{\theta} = I_{NT}($$

I_{NT} : q من المتجهات التي أول p من عناصرها يساوي 1 والباقي اصفاراً

S : مصفوفة التمهيد $S = I_{NT}(H'K_zH)^{-1}H'K_z$

K : مصفوفة الاوزان لدالة اللب القطرية بالابعاد (nT x nT) ,

$diag(K$

Y: متجه المتغير المعتمد بالابعاد (NT x 1) , $\mathrm{Y} = [\mathrm{Y}_{11}, \dots, \mathrm{Y}_{NT})$

X : مصفوفة بالابعاد (NT) x (1+p) حيث أن H :

$$\mathrm{H} = \begin{bmatrix} 1 & \cdots \\ \vdots & \ddots \\ 1 & \cdots \end{bmatrix}$$

وبالتعويض عن $m(z_{it})$ في الصيغة (23) يمكن كتابة نموذج الانحدار للبيانات الطولية (المزدوجة) بصيغة المصفوفات والمتجهات بالشكل الآتي :

$$\hat{Y} = \hat{X}'\beta + w^*$$

إذ أن :

$$\hat{\mathrm{Y}} = (I_{NT} - S)$$

$$\hat{\mathrm{X}} = (I_{NT} - S)$$

$$w^* = (I_{NT} - S$$

$$m(z) = m(z_{11}$$

وعند تطبيق طريقة المربعات الصغرى يتم الحصول على التقدير شبه المعلمي لمتجه المعلمات (β) ودالة التمهيد $m(z)$ كما في الآتي:

$$\hat{\beta}_{PLLLS} = \left(\hat{\mathrm{X}}'\hat{\mathrm{X}}\right)^{-1}(\hat{\mathrm{X}}'$$

$$\hat{m}_{PLLLS}($$

5- المحاكاة :-تم استعمال برنامج لغة الآر (R Language) لتنفيذ تجارب المحاكاة بإستعمال عشرة مقاطع عرضية (n=10) وتمثل عدد القطاعات مع ثلاث فترات زمنية هي (t=20,t=10,t=5) وبذلك سيكون لدينا ثلاثة حجوم للعينات (NT=50) و (NT=100) و (NT=200) وللحصول على نتائج دقيقة ومستقرة تم تكرار كل تجربة (Replicates(R)=1000) .

1-5 توليد المتغيرات العشوائية (Generating Random Variables)

1- تم توليد المتغيرات التوضيحية (x1,x2) للجزء المعلمي على وفق التوزيع الطبيعي $N(0,\sigma^2)$

2- تم توليد المتغير التوضيحي اللامعلمي (z) حسب التوزيع المنتظم $U(0, 1)$.

3- علما أن عملية توليد المتغيرات تتم على وفق عدد المقاطع العرضية والفترات الزمنية وبذلك فإن الأنموذج اللامعلمي تألف من ثلاثة متغيرات (x1,x2,z) , أما الأنموذج شبه المعلمي فقد تألف من المتغيرين التوضيحيين (x1,x2) اللذين يتم توليدهما من التوزيع الطبيعي فضلاً عن المتغير التوضيحي اللامعلمي (z) الذي يتم توليده بحسب التوزيع المنتظم .

4- تم توليد متجه الاخطاء العشوائية e بحيث يتبع التوزيع الطبيعي $N(0, \sigma^2)$, اما متجه الأخطاء العشوائية للتأثيرات المنفردة u_i بحيث يتبع التوزيع الطبيعي القياسي , ثم يتم تحويله الى التوزيع الطبيعي بمتوسط صفر وتباين σ^2 كما يلي :

$$u_i = \sigma^2 * S$$

وقد تم اختيار ثلاث قيم لتباين الخطأ σ^2 وكالآتي :-

a) $\sigma^2=0.5$

b) $\sigma^2=1$

c) $\sigma^2=1.5$

اما متجه الخطأ vij فيتم توليده بشكل مباشر من خلال البيانات التي تم توليدها وبالإعتماد على الصيغة التالية المستوفاة من بحوث منشورة سابقاً [7]

$$v_{ij} = x_{ij}e \quad , e \sim N \quad - 5)$$

وبذلك يمكن الحصول على متغير الخطأ المركب (φ_{ij}) للأنموذج العشوائي :

$$\varphi_{ij} = u_i + v_{ij}$$

5- دالة التمهيد اللامعلمية تم اختيار الدالة التالية على وفق الصيغة التالية : [15]

$$m(z_{ij}) = exp \quad -$$

6- تم توليد المتغير المعتمد y_{it} باستعمال المتغيرات التوضيحية والأخطاء العشزائية ودوال التمهيد التي تم توليدها في الفقرات 1-3 المذكورة آنفاً وحسب الانموذج المطلوب .

7- كما تم تحديد قيم المعلمات للانموذج شبه المعلمي من خلال تقديرها بطريقة المربعات الصغرى وبشكل يتلاءم مع طبيعة البيانات القياسية المدروسة في الجانب التجريبي وكما يلي :

$$(\beta_0 = 0.5, \beta_1 = -0.1 , \beta_2 = 0.1)$$

8- تم استعمال دالتي اللب (Gaussian) و (Epanchnikove) فضلا عن إستعمال طريقة العبور الشرعي في تقدير عرض الحزمة h ولطرائق التقدير كافة وحسب الخطوات التي تم تناولها في الجانب النظري , والجدول (3) الآتي يمثل وصفاً للنماذج المقدرة حسب طرائق التقدير :

الجدول (3) وصف النماذج المقدرة

ت	طرائق التقدير	النماذج
1	**NW Estimation , Gaussian Kernel**	Model I
2	**Local Estimation, Gaussian Kernel**	Model II
3	**NW Estimation , Epanechnikov Kernel**	Model III
4	**Local Estimation, Epanechnikov Kernel**	Model IV
5	**Profile, NW Estimation, Gaussian Kernel**	Model IX
6	**Profile, Local Estimation, Gaussian Kernel**	Model X
7	**Profile, NW Estimation, Epanechnikov Kernel**	Model XI
8	**Profile, Local Estimation, Epanechnikov Kernel**	Model XII

2-5عرض وتحليل النتائج (Presentation and Analysis of Results)

تم الإعتماد على برنامج مكتوب بلغة (R) بالاصدار (4.0.1) وتم استخراج النتائج الخاصة بالجانب التجريبي والجانب التطبيقي ،وتعد لغة (R) من اللغات ذات الإمكانات المتقدمة في الجانب الإحصائي كونه يحتوي على عدد ضخم من المكتبات والبرامج الإحصائية الجاهزة ، وقد تم وضع نتائج معدل متوسط مربعات الخطأ للنماذج المختلفة ولمختلف مستويات التباين وللدالة اللامعلمية في الصيغة (3–5) ولحجوم العينة المختلفة أي عندما (t=5,t=10,t=20) في الجدول (4) الآتي :

الجدول (4) قيم متوسط مربعات الخطأ المقدرة للنماذج عند الفترات الزمنية t=5 و t=10 و t=20

	Model		AMSE		
		Time	t=5	t=10	t=20
	I-NW-Gaussian		1.2858097	1.4986920	1.3407787
σ	II-Loc-Gaussian		1.3319610	1.3318322	1.3256126

	III- NW- Epanechnikov	1.2983010	1.3470976	1.1925589
	IV Loc- Epanechnikov	1.3022477	1.3231794	1.1914576
	V- Profile, NW-Gaussian	0.2052279	0.2687551	0.2269983
	VI- Profile,Loc-Gaussian	0.2072771	0.2881808	0.2582405
	VII Profile, NW-Epanechnikov	0.1950638	0.2577413	0.2209237
	VIII Profile, Loc-Epanechnikov	0.1994209	0.2881808	0.2582405
	I-NW-Gaussian	1.4273528	1.3055301	1.2185321
	II-Loc-Gaussian	1.4321077	1.3271043	1.4224950
	III- NW- Epanechnikov	1.4232113	1.3064367	1.1328122
	IV Loc- Epanechnikov	1.4271961	1.3042334	1.1297584
	V- Profile, NW-Gaussian	0.8795318	0.9286425	1.0012010
	VI- Profile,Loc-Gaussian	0.8152605	0.8125857	1.0121186
	VII Profile, NW-Epanechnikov	0.5043395	0.8204961	0.9959044
	VIII Profile, Loc-Epanechnikov	0.8152605	0.6434255	0.9844894
	I-NW-Gaussian	2.074883	2.436126	2.12572022
	II-Loc-Gaussian	2.043568	2.365598	2.14293376
	III- NW- Epanechnikov	2.037250	2.413330	2.11745842
	IV Loc- Epanechnikov	2.043568	2.301966	2.08150292
	V- Profile, NW-Gaussian	1.555373	1.985199	0.68197144
σ	VI- Profile,Loc-Gaussian	1.831690	1.802119	0.50874800
	VII Profile, NW-Epanechnikov	1.144003	1.172775	0.58913833
	VIII Profile, Loc-Epanechnikov	1.315758	1.182727	0.00125282

وقد تبين من خلال الجدول المذكور آنفاً الذي وضح مقدار معدل متوسط مربعات

الخطأ لمستويات التباين المختلفة ولتجارب المحاكاة كافة مايلي :

1- النماذج المقدرة على وفق الطرائق شبه المعلمية أعطت قيم معدل متوسط مربعات الخطأ أقل من تلك القيم التي أعطتها الطرائق اللامعلمية .

2- سلوك النماذج المقدرة حسب تباينات الأخطاء المفترضة , كانت القيم المقدرة لمعدل متوسط مربعات الأخطاء تزداد بزيادة قيمة تباين الخطأ , وأن تأثير هذه الزيادة في الأنموذج اللامعلمي كان أكبر بكثير من النماذج المقدرة على وفق الطرائق شبه المعلمية .

3- فيما يخص زيادة عدد مشاهدات السلسلة الزمينة فأنه يؤثر بشكل طردي في قيم متوسط مربعات الخطأ المقدرة للنماذج كافة, إذ تزداد قيم معدل متوسط مربعات الخطأ عند $t=20$ بشكل ملحوظ عن $t=5, t=10$, وأن تلك الزيادة في قيم معدل متوسط مربعات الخطأ تكون كبيرة في النماذج اللامعلمية وتقل في النماذج شبه المعلمية .

4- وقد حلت الطرائق شبه المعلمية بالمرتبة الاولى عند التباينات 1.5 و 1 و $\sigma^2=0.5$ في حين حلت الطرائق اللامعلمية بالترتيب الثاني .

6-الاستنتاجات :(Conclusions)

بعد أن تم تنفيذ تجارب المحاكاة وما تم عرضه وتحليله من نتائج استنتج الباحث ما يلي :

1- إن استعمالنا لمفهوم البيانات الطولية (أو المزدوجة) سهل لنا دراسة شاملة عند استعمالنا البيانات المقطعية والسلاسل الزمنية في الوقت ذاته , إذ انتج أنموذجا رياضيا واحدا يمثل طبيعة البيانات المدروسة مقارنة مع توظيف عشرة نماذج انحدار اعتيادية (اي أنموذج واحد لكل سلسلة زمنية) أو فيما لو تم استعمال عشرة نماذج انحدار لسلسلة زمنية اي بمعنى (انموذج انحدار سلسلة زمنية لكل مقطع عرضي) .

2- الطرائق الاكثر كفاءة في تقدير انموذج البيانات الطولية العشوائي هي الطرائق شبه المعلمية , إذ اثبتت دقتها في الحصول على قيم معلمات وتمهيد لامعلمي في التباينات الواطئة .

3- الطرائق شبه المعلمية ثبتت كفاءتها في التيابنات العالية .

4- كلما ازدادت قيمة تباين الخطأ لأنموذج البيانات الطولية العشوائي ازدادت قيم معدل متوسط مربعات الخطأ وعليه يمكن القول ان العلاقة بين قيمة تباين الخطأ وكفاءة طرائق التقدير هي علاقة طردية .

7-التوصيات :(Recommendations)

بناءا على ما توصل اليه الباحث من استنتاجات وانسجاما مع فكرة وموضوع البحث واستكمالا له يمكن اجمال التوصيات التالية :

1- استعمال نماذج البيانات الطولية في تحليل ودراسة النماذج ذات الطبيعة مزدوجة التأثير للمقاطع والسلاسل الزمنية على وفق الطرائق شبه المعلمية .

2- الاعتماد على تنويع الدوال اللبية عند المقارنة بين طرائق التقدير اللامعلمية في تقدير الانموذج العشوائي للبيانات الطولية .

3- استعمال طرائق تمهيد لامعلمية وشبه معلمية اخرى ممكن أن تثبت كفاءتها اكثر من الطرائق المستعملة في هذا البحث مثل طريقة شرائح التمهيد التكعيبية (CSS) وطريقة شرائح الجزاء (Penalized) .

4- استعمال طرائق اخرى مثل التقدير البيزي .

المصادر العربية :

1- القيسي، باسم شليبة ، (2009) ،" التحليل البيزي لنماذج الانحدار الخاصة بالبيانات المزدوجة panel data "، اطروحة دكتوراه في الاحصاء كلية الادارة والاقتصاد جامعة بغداد ".

2- بدر ، دريد حسين ، (2016) ،" تشخيص وتقدير دالة الانحدار المعلمي للبيانات المزدوجة panel data في حالة عدم تحقق بعض فرضياته "، اطروحة دكتوراه في الاحصاء، كلية الادارة والاقتصاد ، جامعة بغداد .

3- حمود , مناف يوسف , عاشور,مروان عبد الحميد , 2012م"مقارنة بضعة مقدرات لأخطية لتقدير دالة الإنحدار ", مجلة العلوم الإدارية والإقتصادية المجلد 18, العدد 68 .

4- عبد الحافظ,علي سيف الدين,2012م "تقدير إنموذج لا معلمي للبيانات الطولية للقطاعات الإقتصادية في العراق " أطروحة دكتوراه في الإحصاء ,كلية الإدارة والإقتصاد ,جامعة بغداد.

المصادر الاجنبية :

5- Blanchet . J and Wadsworth. J . (2012) . " Applied Statistics" ,Institute of Mathematics, Analysis, and Applications EPF Lausanne , An MSc Course for Applied Mathematicians .

6- Fan, J., & Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli, 11(6), 1031-1057.

7- Evdokimov, K. (2010)." Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity",Journal of the American Statistical Association December 2010, Vol.196, No. 654.

8- Hardle,W.,Muller.,Sperlich,S.,H.,Werwatz,A.,(2004)."Non parametric and semiparametric Models an Introduction".Springer Edition.
9- Henderson,D.J. and Ullah, A.(2005)." A Nonparametric Random Effects Estimator". Economics Letters 88:403-407.
10- Macculloch, C. E. and Searle,S. R., (2001), "Generalized Linear and Mixed models", Wiley, New York.(37) .
11- Moffatt, M. (2010), "Definition of Longitudinal Data", http://economics.about.com/od/termsbeginningwith1/Longitudinal-data-htm.(38) .
12- Robinson, P. M. (1988). Root-N-consistent semiparametric regression. Econometrica: Journal of the Econometric Society, 931-954.
13- Rodriguez-Poo, J. M., & Soberon, A. (2017). Nonparametric and semiparametric panel data models: Recent developments. Journal of Economic Surveys, 31 (4), 923-960.
14- Su, L., & Ullah, A. (2011). Nonparametric and semiparametric panel econometric models: estimation and testing. Handbook of empirical economics and finance, 455-497.
15- Wang, Q., Linton, O., & Härdle, W. (2004). "Semiparametric regression analysis with missing response at random". Journal of the American Statistical Association, 99(466), 334-345 .
16- You, J., & Zhou, X. (2006). Statistical inference in a panel data semiparametric regression model with serially correlated errors. Journal of multivariate analysis, 97(4), 844-873.

17- Zeger, S. L., Diggle, P. J., Heagerty, P. and Liang, K. Y., (2002), "Analysis of longitudinal data", second edition, OXFORD university press, New York.(69)