استعمال بعض طرائق التقدير اللامعلمي لأنموذج البيانات المزدوجة في حالة وجود مشكلة عدم التجانس لتقدير الموازنة العامة للدول العربية للمدة (2002-2014)

بحث مستل من أطروحة دكتوراه مقدمة إلى مجلس كلية الإدارة والاقتصاد في جامعة بغداد وهي جزء من متطلبات نيل درجة دكتوراه فلسفة في الاحصاء والموسومة " تقدير أنموذج الأنحدار اللامعلمي للبيانات المزدوجة في حالة عدم تحقق بعض فرضياته "

أ.د. ظافـــــر حسين رشيـــد

جامعة بغداد - كلية الادارة والاقتصاد - قسم الاحصاء

م.م. دريد حسين بدر

جامعة البصرة - كلية الادارة والاقتصاد - قسم الاحصاء

The use of some nonparametric methods to model of nonparametric panel data in case there is a problem of heterogeneity to estimate of the general budget for the Arab States for the period(2002-2014)

Prof. Dr. Dhafer Hussein Rashid

Lecturer. Dr. Duraid Hussein Badr

استعمال بعض طرائق التقدير اللامعلمي لأنموذج البيانات المزدوجة في حالة وجود مشكلة عدم التجانس لتقدير الموازنة العامة للدول العربية للمدة (2002-2014)

أ.د. ظافر حسين رشيد

م.م. دريد حسين بدر

المستخلص

في هذا البحث تمت التطرق إلى دراسة بعض المقدرات اللامعلمية لأنموذج التأثيرات العشوائية اللامعلمية للبيانات المزدوجة (Nonparametric Panel Data Models With Random Effects) في حالة وجود مشكلة عدم التجانس، إذ أن هناك مشكلة خطيرة هي في اختيار عرض المعلمة التمهيدية (h)، وبالتالي نحن بحاجة لإيجاد طريقة سريعة لتحديد دقيق لمعلمة التمهيد من خلال تناول معيار تقاطع الشرعية (العبور الشرعي)، فتم تناول كلاً من مقدر الثابت الموضعي اللامعلمي باستعمال توسيع تايلور (Local Constant Nonparametric Estimator by a Taylor expansion) (LCNE)، والمقدر الخطي الموضعي اللامعلمي باستعمال توسيع تايلور(Local Linear Nonparametric Estimator by a Taylor expansion) (LLNE)، ومقدر طريقة التقدير ذات المرحلتين (Nonparametric Two – Step Estimation Method in panel data) (Two – Step)، من خلال تطبيق تلك الطرق على بيانات واقعية تشير إلى الموازنة العامة للدول العربية للمدة (**2002–2014**)، وقد أثبتت النتائج أفضلية مقدر المربعات الصغرى الخطية الموضعية الموزونة للأنموذج وتفوقه على المقدرين الأخرين من خلال حصوله على أقل قيمة لمعايير (RAMSE) (AMAPE) .

الكلمات الأفتتاحية : البيانات المزدوجة، مقدر الثابت الموضعي (LLNE)، المقدر الخطي الموضعي (LLNE)،مقدر طريقة التقدير ذات المرحلتين (Two–step)، عرض الحزمة، معيار (RAMSE) ،معيار (AMAPE)، أنموذج التأثيرات العشوائية اللامعلمية للبيانات المزدوجة .

ABSTRACT

In this research has addressed to the study of some estimators nonparametric to model random effects nonparametric double data (Nonparametric Panel Data Models With Random Effects) in case of a problem of heterogeneity, as there is a serious problem is in the selection of display primaries parameter, and therefore we need to find a A quick way to determine the accurate boot by eating standard intersection of legitimacy to the parameter (legitimate transit), and was eating both destined hard topical nonparametric using expansion Taylor (Local Constant Nonparametric Estimator by a Taylor expansion) (LCNE), and the estimated linear topical nonparametric using expansion Taylor (Local Linear Nonparametric Estimator by a Taylor expansion) (LLNE), and an estimated two-stage estimation method (Nonparametric Two - Step Estimation Method in panel data) (Two - Step), by applying these methods on realistic data indicate the general budget for the Arab States for the period (2014-2002) , and the results proved destined least squares linear positional weighted to model and superiority over the estimated others preference by receiving a lower value to standards (RAMSE), (AMAPE) .

Keyword Panel data: LCNE ,LLNE , Two-step , Bandwidth , RAMSE , AMAPE Random effects .

المقدمة :

تكتسب الطرق الإحصائية اللامعلمية جاذبية خاصة لدى العاملين في مجال البحث العلمي، وهي من الطرائق الإحصائية الاستدلالية التي يمكن أستعمالها للتوصل إلى استنتاجات بشان المجتمع في ضوء العينة بغض النظر عن نوع التوزيع النظري لذلك المجتمع وتكون أكثر مرونة في التعامل مع البيانات وتتعامل مع طرائق لا تحتاج إلى فرضيات أو على أقل تقدير تحتاج إلى فرضيات أقل من الأسلوب المعلمي خاصة فيما يتعلق بعدم الحصول على معلمات أو أن بيانات العينة تكون وصفية أو رتبوية أو نسبية وهي موجودة في الكثير من حالات التطبيق، وعلى الرغم من أن نماذج الانحدار اللامعلمي هي اضعف وصفا من نماذج الانحدار ألمعلمي إلا أنها في الوقت نفسه تحتاج إلى قيود أو شروط اقل على شكل الدالة التي تجعل أنموذج الانحدار خطي أو لاخطي، إحتمالي أو لوجستي ... الخ، وان التطبيقات اللامعلمية لها القابلية على كشف تركيبة البيانات خصوصاً عندما يصغر حجم العينة من النماذج المعلمية للاستدلال على البيانات، وهذا الأمر تحديداً هو الذي جعل من نماذج الانحدار اللامعلمية أداة مرغوبة جدا لدى الباحثين لكون إن البيانات الفعلية ليست دوما لديها مواصفات مثالية، ولكن في حالة وجود بيانات غير مثالية فيكون الاهتمام عادة بالاعتماد على إعطاء وصف عام للعلاقة وليس دراسة تفاصيل العلاقة الدقيقة، مما أدى إلى انتعاش طريقة التفكير بنماذج الانحدار اللامعلمي .

مشكلة البحث : (Objective of research)

إن عملية تحليل البيانات تستوجب تقدير معلمات الأنموذج الممثل لها و لغرض تقدير أنموذج الانحدار اللامعلمي للبيانات المزدوجة لابد إن يتمتع هذا الأنموذج بخصائص معينة تعتمد على افتراضات تتعلق بأنموذج الانحدار، إذ انه في حالة عدم توفر تلك الافتراضات فأن الأنموذج سيعاني جملة من المشكلات المهمة بشكل عام تواجه نماذج الانحدار بأشكالها المختلفة وهي (مشكلة الارتباط الذاتي ومشكلة التعدد الخطي ومشكلة عدم تجانس التباينات ومشكلة البيانات المفقودة ومشكلة خطا القياس ...الخ). والتي تجعل عملية التقدير خاطئة أو في بعض الأحيان غير ممكنه، ونتيجةً لذلك سنتطرق إلى طرائق تقدير أنموذج البيانات المزدوجة اللامعلمية في حالة عدم توفر بعض الافتراضات، إذ اننا بصدد الحديث عن عدم تحقق على الأقل واحد من فرض ثبات تجانس التباين ($\sigma_1^2 = \sigma_2^2 = ... = \sigma_n^2$) بمعنى عندما يكون تباين المجتمع الإحصائي غير ثابت لكل المشاهدات .

هدف البحث :

يتمثل في تسليط الضوء على بعض الطرائق اللامعلمية التي تستعمل لتقدير دالة الانحدار اللامعلمي الخطي للبيانات المزدوجة (بوجود بيانات تامة)، وذلك باستعمال كلاً من مقدر الثابت الموضعي اللامعلمي باستعمال توسيع تايلور (Local Constant Nonparametric Estimator by a Taylor expansion) (LCNE)، والمقدر الخطي الموضعي اللامعلمي باستعمال توسيع تايلور (Local Linear Nonparametric Estimator by a Taylor expansion) (LLNE) ومقدر طريقة التقدير ذات المرجلتين – (Nonparametric Two Step Estimation Method in panel data) (Two – Step) ومن ثم اختيار معلمة التمهيد وذلك باستعمال معيار العبور الشرعي العام (Generalized Cross Validation (GCV)) ولعل أهم ما يتمركز إليه البحث هو اعتماد الطرائق في الأنموذج اللامعلمي و أختيار الأفضل ملائمة لهذه المجموعة من البيانات، وذلك باستعمال المعايير (RAMSE)،(AMAPE)، لاختيار الطريقة الأفضل والأكثر ملائمة لانموذج التأثيرات

العشوائية اللامعلمية للبيانات المزدوجة، وذلك من خلال دراسة الموازنة العامة للدول العربية مقاسة بملايين الدولارات كدالة بدلالة عدد من المتغيرات التفسيرية وحسب وجهة نظر أهل الاختصاص للمدة (2002 – 2014) مصاغة لانموذج التأثيرات العشوائية اللامعلمية للبيانات المزدوجة وذلك من خلال إيجاد المقدرات اللامعلمية لأنموذج التأثيرات العشوائية اللامعلمية .

البيانات المزدوجة (Panel Data) : (2)، (3)

لقد اكتسبت هذه النماذج في الاونة الاخيرة اهتماما بالغا وخصوصا في الدراسات الاقتصادية والطبية والوبائية والمالية لأنها تأخذ في الاعتبار اثر التغير في الزمن وكذلك اثر التغير في المشاهدات المقطعية على حد سواء الكامن في بيانات عينة الدراسة، تعرف البيانات المزدوجة على أنها مشاهدات مقطعية مقاسه في فترات زمنية معينة أذ تشمل المشاهدات المقطعية الدول، الأسر, السلع، المحافظات،المؤسسات...الخ، وتبرز أهمية استعمال البيانات المزدوجة في أنها تأخذ في الاعتبار مايوصف "بعدم التجانس أو الاختلاف غير الملحوظ " (Unobserved heterogeneity) الخاص بمفردات العينة سواء المقطعية أو الزمنية، وبالتالي فان نماذج البيانات المزدوجة في ثلاثة إشكال رئيسية هي أنموذج الانحدار التجميعي (Pooled Regression Model) وأنموذج التأثيرات الثابتة (Fixed Effects Model) وأنموذج التأثيرات العشوائية (Random Effects Model) الذي سيتم تناوله في حالة وجود مشكلة عدم التجانس .

الجانب النظري :

نماذج الانحدار اللامعلمية Non–Parametric Regression Models (1)، (2)، (12)، (13)

أذ أن الشكل العام :

$$Y_i = m(x_i) + \varepsilon_i \qquad ,i = 1,2,.....,n \qquad \dots (2\text{–}1)$$

إذ أن :

$m(.)$: دالة التمهيد غير المعروفة .

Y_i : تمثل قيمة متغير الاستجابة .

x_i :تمثل قيمة المشاهدات للمتغير التوضيحي (التفسيري) وهو متجه من الدرجة p× 1.

ε_i : قيمة الأخطاء العشوائية ذو التوزيع الطبيعي مستقلة ومتماثلة التوزيع $\varepsilon_i \sim N(0, \sigma^2)$.

ومن الأنموذج (1–2) فانه ليس هنالك حاجة لتقدير معالم هي عبارة عن نقاط , وبالتالي نحتاج إلى تمهيد(Smoothing) وهو أداة تستخدم لوصف اتجاه وتأثير المتغير التوضيحي في متغير الاستجابة, وهذه الأداة يمكن تمثيلها كدالة بدلالة واحد أو أكثر من المتغيرات التوضيحية لمنحني الأنموذج $m(x)$ ، وهو أسلوب بياني لامعلمي يمكن استخدامه لتقدير أي دالة أحصائية، إذ أن هدف أساليب التمهيد هو تقريب دالة الانحدار اللامعلمي التقريبية إلى دالة الانحدار اللامعلمية الحقيقة، أذ إن تحليل الانحدار اللامعلمي هو الأفضل في حالة غياب أي معلومات عن الدالة الحقيقية، أي أن تقنيات الانحدار اللامعلمي تعول كثيراً على البيانات في تحديد تلك الدالة أكثر من اعتماده على الافتراضات، للحصول على معلومات عن $m(.)$ بدلا من الصيغة الدالية وفي بعض الحالات ينتهي تحليل الانحدار اللامعلمي بتوفيق منحنى للبيانات دون تحديد شكل معين للدالة، أذا ماقورنت بنظيراتها ألمعلمية، إذ إن الانحدار اللامعلمي معروف جيداً بأنه يجعل البيانات أكثر تكيفاً، وأقل تقييداً من المنهج المعلمي،لانه يتعامل مع طرائق لاتحتاج إلى فرضيات أو في اقل تقدير تحتاج إلى فرضيات أقل وخاصة فيما

يتعلق بعدم الحصول على المعلمات أو أن بيانات العينة تكون رتبية أو نسبوية أو وصفية، لذلك برز الانحدار اللامعلمي كبديل واعد لمعالجة البيانات المزدوجة (الطولية) .

نماذج البيانات المزدوجة اللامعلمية (Models of Nonparametric Panel Data)

أنموذج الانحدار التجميعي اللامعلمي للبيانات المزدوجة (8)، (9)، (10)

(Pooled Regression Model of Nonparametric Panel Data) (NPRM)

يعتبر هذا الأنموذج من ابسط نماذج البيانات المزدوجة حيث تكون فيه جميع المعاملات ثابتة لجميع الفترات الزمنية (إي يهمل تأثير للزمن)، ويفترض في هذا النموذج (نتعامل مع حالة التجانس في التباين بين المجاميع) وان

$$cor\,(\varepsilon_{it},\varepsilon_{jt})=0 \quad ; i\neq j \quad ; t=1,2,3,\ldots\ldots\ldots\ldots,\ n.$$

ويكون الأنموذج بالصيغة الأتية:

$$y_{it}=m(x_{it})+u_{it} \qquad ,i=1,2,\ldots\ldots\ldots,n \quad ,t=1,2,\ldots\ldots,T \qquad \ldots (2-2)$$

إذ إن :

y_{it} : تمثل قيمة متغير الاستجابة في المشاهدة i عند الفترة الزمنية t.

x_{it} : تمثل قيمة المتغير التوضيحي (التفسيري) في المشاهدة i عند الفترة الزمنية t، وهو متجه صفي من الدرجة P ×1 ، $x_{it}\in R^{q}$.

$m(.)$: دالة التمهيد غير المعرفة ويمكن تقديرها بطرق التقدير اللامعلمية، $\mathrm{E}\left[y_{it}/x_{it}=x\right]=m(x)$.

u_{it} : تمثل قيمة الخطأ في المشاهدة i عند الفترة الزمنية t وتكون متغير عشوائيا مستقلاً ومتماثل التوزيع (i.i.d) (0 ، σ_u^2)، لكل قيم x_{jt} ، i,j =1,2,……n ، t=1,2,….T ويحقق الشرط $\mathrm{E}(u_{it}/x_{it})=0$ ، $Var\,(u_{it}^2/x_{it})=\sigma^2(x_{it})$.

أنموذج التأثيرات الثابتة اللامعلمية للبيانات المزدوجة (8)، (9)، (10)

(Nonparametric Panel Data Models With fixed Effects) (NFEM)

تكون معلمات أنموذج التأثيرات الثابتة صحيحة وغير متحيزة عادة مايفرض بان تباين الخطأ غير ثابت (غير متجانس) لجميع المشاهدات المقطعية وهناك ارتباط ذاتي خلال الزمن بين كل مجموعة من مجاميع المشاهدات المقطعية في فترة زمنية محددة , ويكون الأنموذج يكون بالصيغة الآتية:

$$y_{it}=m(x_{it})+\alpha_i+u_{it} \qquad ,i=1,2,\ldots\ldots\ldots,n \quad ,t=1,2,\ldots\ldots,T \qquad \ldots (2-3)$$

إذ إن :

α_i : معالم التأثيرات الثابتة الفردية غير المتجانسة وتكون مستقلة ومتماثلة التوزيع (0 ، σ_u^2).

وان α تحتوي على معالم غير متجانسة ومرتبطة مع الخطأ u_{it} وانحدار x_{it} ، ويقصد بمصطلح التأثيرات الثابتة بان المعلمة α_i لكل مجموعة بيانات مقطعية لاتتغير خلال الزمن (Time invariant)، وإنما يكون التغير فقط في مجاميع البيانات المقطعية وتم وضع الفرضية ($\sum_{i=1}^{n}\alpha_i$ =0)، ويمكن تقريب الأنموذج في الصيغة (2-3) وكالاتي :

$$Y \approx X(x)\delta(x) + D\alpha + U \qquad \dots (2-4)$$

إذ أن :

$$\alpha = (\alpha_1, \alpha_2, \ldots\ldots, \alpha_n)' \ , \ D = (I_n \otimes l_T)\, d_n \ , d_n = \left[-l_{n-1} \ I_{n-1} \right]'$$

I_n : متجه الوحدة ذا سعة n ×1 .

U : موجه ذو مرتبة (T×1) من الأخطاء العشوائية.

وأن $U \sim N\left(0, \sigma^2 I_n\right)$

$\otimes$: تعني الضرب المباشر (Kronecker Product).

$$Y = (y_{11}, y_{12}, \ldots\ldots, y_{1T}, y_{n1}, \quad , y_{nT})'$$

$$X(x) = \left((x_{11}(x), x_{12}(x), \ldots\ldots, x_{1T}(x), \ldots\ldots, x_{n1}(x), \ldots\ldots\ldots\ldots\ldots , x_{nT}(x) \right)'$$

$$B(x) = \frac{\partial m(x)}{\partial (x)} \qquad , \delta(x) = \left(m(x) B(x)' \right)'$$

ويطلق على الأنموذج في الصيغة (2-4) تسمية أنموذج المربعات الصغرى للمتغيرات الوهمية (Least Squares Dummy Variable Model) للأنموذج، بالإضافة لظهور تسميات أخرى للأنموذج وهي أنموذج المتغيرات الوهمية (Dummy variables model)، أنموذج تحليل التغاير (Analysis of covariance model) .

أنموذج التأثيرات العشوائية اللامعلمية للبيانات المزدوجة (Nonparametric Panel Data Models With Random Effects) (NREM) (8)، (9)، (10)

تكون معلمات أنموذج التأثيرات العشوائية صحيحة وغير متحيزة عادة مايفرض بان تباين الخطأ غير ثابت (غير متجانس) لجميع المشاهدات المقطعية وهناك ارتباط ذاتي خلال الزمن بين كل مجموعة من مجاميع المشاهدات المقطعية في فترة زمنية محددة , وان حد الخطأ $u_{it} \sim (0, \sigma_u^2)$ ، ويعتبر أنموذج التأثيرات العشوائية أنموذج ملائم في حالة وجود خلل في احد الفروض المذكورة في أنموذج التأثيرات الثابتة .

الأنموذج يكون بالصيغة الآتية :

$$y_{it} = m(x_{it}) + v_i + u_{it} \qquad , i = 1,2,\ldots\ldots\ldots, n \quad , t = 1,2,\ldots\ldots, T \qquad \dots (2-5)$$

إذ إن :

v_i : التأثيرات العشوائية المفردة ويمثل متغير عشوائيا مستقلاً ومتماثل التوزيع (i.i.d)

(σ_v^2 ، 0)، وهو حد الخطأ في مجموعة البيانات المقطعية .

u_{jt} ، v_i : غير مرتبطة لجميع قيم i,j =1,2,……n ، t=1,2,….T متغير عشوائيا مستقلاً ومتماثل التوزيع (i.i.d) (0 ، σ_u^2) ، $E\left(v_i + u_{it} / x_{it}\right) = 0$.

وتم وضع بعض الفرضيات لوصف الاختبارات الممكنة:

$$\varepsilon_{it} = v_i + u_{it} \ , \ \varepsilon_i = \left(\varepsilon_{i1}, \ldots\ldots, \varepsilon_{iT} \right)' \ and \ \ \varepsilon_i = \left(\varepsilon_1, \ldots\ldots, \varepsilon_n \right)'$$

$$\sum = E\left(\varepsilon_i \, \varepsilon_i' \right) = \sigma_u^2 I_T + \sigma_v^2 I_T \, I_T'$$

$$\Omega = \mathrm{E}\left(\varepsilon_i \varepsilon_i'\right) = \begin{pmatrix} \Sigma & \cdots & \cdots & 0 \\ 0 & \Sigma & \cdots & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \Sigma \end{pmatrix} = (In \otimes \Sigma)$$

إذ أن :

ε_{it} : تمثل قيمة الخطأ في المشاهدة i عند الفترة الزمنية t .

$\otimes$: تعني الضرب كرونكر المباشر (Kronecker Product).

I_n : متجه الوحدة من الدرجة $n \times 1$.

Ω : مصفوفة التباين والتباين المشترك .

ويطلق على أنموذج التأثيرات العشوائية أحيانا بأنموذج مكونات الخطأ (Error Components Model) بسبب أن النموذج في الصيغة (2-5) يحتوي مركبتين للخطأ في الأنموذج (يكون حد الخطأ $\varepsilon_{it} = v_i + u_{it}$, $i = 1,2,....,n$, $t = 1,2,.....,T$)

وبالتالي يكون الأنموذج بالصيغة الأتية :

$$y_{it} = m(x_{it}) + \varepsilon_{it} \qquad ,i = 1,2,.....,n \quad ,t = 1,2,.....,T \qquad \ldots (2-6)$$

ويمتلك أنموذج التأثيرات العشوائية خواص رياضية ($\mathrm{E}(\varepsilon_{it}) = 0$ ، $\mathrm{var}(\varepsilon_{it}) = \sigma_\varepsilon^2$) ، وكذلك يسمى أنموذج مركبات التباين (Variance components model) .

الأختبارات اللامعلمية المستعملة (**Test of Nonparametric**) :هناك العديد من الاختبارات اللامعلمية منها :

الأختبار اللامعلمي لمشكلة عدم التجانس [5] **Test of Heteroscedasticity of Nonparametric**

المشكلة تكمن في أيجاد اختبار لامعلمي يكتشف مشكلة عدم التجانس ولأختبار الفرضية $H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_n^2$ ضد الفرضية البديلة على الأقل واحد من فرض ثبات تجانس التباين ($H_1 : \sigma_1^2 \neq \sigma_2^2 \neq ... \neq \sigma_n^2$)، وذلك من خلال استعمال مقدر دالة التباين المتكامل المعطي من قبل (Rice)(1984) معتبراً مجموع مربعات البواقي

$$\frac{1}{n-1\sum_{i=1}^{n} R_{i,n}^2} \qquad \ldots (2-7)$$

إذ أن :

$R_{i,n}^2 = \frac{1}{2}\left(Y_{i,n} - Y_{i-1,n}\right)^2$ ، تحت فرضية تحقق $E\left[R_{i,n}^2\right] \approx \sigma^2(x_i, n)$ ، ولاختبار الفرضية ($\mathrm{E}(uu') = \sigma_u^2 I_n$) فقد قام العديد من العلماء بتطوير الأختبار للوصول للصيغة النهائية التي قدمها (Zheng)(1996) وتكون بالشكل :

$$T_n = \frac{1}{(n-1)(n-2)h} \sum_{|i-j| \geq 2} K\left(\frac{x_{i,n} - x_{j,n}}{h}\right)\left(R_{i,n}^2 - \overline{R_n^2}\right)\left(R_{j,n}^2 - \overline{R_n^2}\right) \qquad 2-8$$

إذ أن :

$K(\,)$: الدوال اللبية (Kernel) الموزونة .

h : عرض الحزمة (المعلمة التمهيدية) .

$$\overline{R_n^2}=\frac{1}{n-1}\sum_{i=2}^{2}R_{i,n}^2 \quad .$$

ويعتمد القرار النهائي على أيجاد قيمة $p-value$ فإذا كانت أقل من أو تساوي 0.05 نقبل H_1 وهذا يعني أن الأنموذج يعاني من مشكلة عدم التجانس وبعكسه نقبل الفرضية H_0 .

اختبار هاسمان اللامعلمي (A nonparametric Hausman test) [8]، [10]

وهو اختبار وضع لاختيار نوع الأنموذج النهائي الملائم لبيانات البحث الذي يجب أستعماله للبيانات المزدوجة (Panel data)، حسب الفرضية الآتية :

$$H_0 : E(\frac{v_i}{x_{i1},x_{i2},.....,x_{iT}}) =0$$

أنموذج التأثيرات العشوائية هو الأنموذج الملائم لبيانات الدراسة ضد الفرضية

$$H_1 : E(\frac{v_i}{x_{i1},x_{i2},.....,x_{iT}}) \neq 0$$

أنموذج التأثيرات الثابتة هو الأنموذج الملائم لبيانات الدراسة وبوضع الأفتراض $E(\frac{u_{ti}}{x_{i1},x_{i2},.....,x_{iT}}) = 0$ ، وبالتالي فان فرضية العدم تصبح بالصورة الأتية

$$H_0 : E(\frac{\varepsilon_{ti}}{x_{i1},x_{i2},.....,x_{iT}}) =0 \quad \text{ضد البديلة} \quad H_1 : E(\frac{\varepsilon_{ti}}{x_{i1},x_{i2},.....,x_{iT}}) \neq 0$$

، $\varepsilon_{it}=v_i+u_{it}$ ، ونتيجة لذلك تم أقتراح أختبار على أساس العينة $J=E\left\{\varepsilon_{it}\,E(\frac{\varepsilon_{it}}{x_{it}})m(x_{it})\right\}$ ، تحت فرضية العدم $J=0$ ، $J=E\left\{E(\frac{\varepsilon_{it}}{x_{it}})^2\,m(x_{it})\right\}>0$ تحت الفرضية البديلة، ويمكن اختبار الفرضية أعلاه بأستعمال احصاءة الاختبار وكالاتي :

$$J_n=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}_{it}\,\hat{E}_{-it}\,E(\frac{\hat{\varepsilon}_{it}}{x_{it}})\,\hat{m}_{-it}(x_{it}) \qquad (2-9)...$$

إذ أن:

$\hat{\varepsilon}_{it}$: تقدير البواقي في أنموذج التأثيرات العشوائية .

$\hat{m}_{-it}(x_{it})$ ، $\hat{E}_{-it}(\frac{\hat{\varepsilon}_{it}}{x_{it}})$: تقديرات كل من دالة $m(x_{it})$ ، $E(\frac{\varepsilon_{it}}{x_{it}})$ باستعمال مشاهدات $\{x_{it},\hat{\varepsilon}_{it}\}$.

وباستعمال احصاءة الأختبار من قبل (Henderson,Carroll,and Li) (2008) وتقارن قيمتها مع قيمة χ^2 المجدولة بدرجة حرية k (عدد المتغيرات التوضيحية) فاذا كانت المحتسبة اكبر من المجدولة نقبل H_1 وبعكسه نقبل الفرضية H_0 (أو إذا كانت $p-value$ أقل من أو تساوي 0.05)، وهذا يدل على أن أنموذج التأثيرات الثابتة هو الملائم لبيانات البحث، وعلى العكس فان أنموذج التأثيرات العشوائية هو الملائم لبيانات البحث، وان الأنموذج النهائي الملائم لبيانات البحث هو أنموذج التأثيرات العشوائية كما مبين في الجانب التطبيقي وبالتالي سيتم تناول المقدرات اللامعلمية لأنموذج التأثيرات العشوائية اللامعلمية للبيانات المزدوجة في حالة وجود مشكلة عدم التجانس .

طرائق تقدير أنموذج التأثيرات العشوائية اللامعلمية للبيانات المزدوجة

.مقدر الثابت الموضعي اللامعلمي باستعمال توسيع تايلور[4]، (6) ، (9)، (11) (Local Constant Nonparametric Estimator by aTaylor expansion) (LCNE) لتوضيح عمل هذا المقدر وصيغه لإيجاد تقدير أنموذج البيانات المزدوجة اللامعلمية للتأثيرات العشوائية (6-2) باستعمال توسيع تايلور نتبع ما يأتي :

$$y_{it} = m(x) + \varepsilon_{it} + O\left(\left| x_{it} - x \right|\right)$$

$$y_{it} = m(x) + \varepsilon_{it} + O\left(\left| h \right|\right). \quad ، \quad \lim_{n \to \infty} h = 0$$

$$y_{it} = m(x) + \varepsilon_{it}$$

تم أضافه $O\left(\left| h \right|\right)$ للـ (ε_{it}) لتكبير حجم العينة ،وتكتب بشكل مصفوفات وكالاتي :

$$Y = L_{nt}\, m(x) + \varepsilon$$

إذ أن :

l_{nt} : متجه بسعة nT×1.

$$Y = (y_{11}, \ldots\ldots, y_{1T}, y_{21}, \ldots\ldots, y_{2T}, \ldots\ldots\ldots, y_{n1}, \ldots\ldots, y_{nT})' = (y'_1, y'_2, \ldots\ldots, y'_n)'$$

$$U = (u_{11}, \ldots\ldots, u_{1T}, u_{21}, \ldots\ldots, u_{2T}, \ldots\ldots\ldots, u_{n1}, \ldots\ldots, u_{nT})' = (u'_1, u'_2, \ldots\ldots, u'_n)'$$

وبتصغير مجموع مربعات الخطأ نحصل على :

$$\sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it} - a \right)^2 K\left(\frac{x_{it} - x}{h} \right) = \left(Y - a l_{nT} \right)' K(x) \left(Y - a l_{nT} \right)$$

وبالتالي نحصل على مقدر الثابت الموضعي اللامعلمي لـ $m(x)$ وكالاتي :

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} y_{it}\, K\left(\frac{x_{it} - x}{h} \right)}{\sum_{i=1}^{n} \sum_{t=1}^{T} K\left(\frac{x_{it} - x}{h} \right)} = \left(l'_{nT} K(x) l_{nT} \right)^{-1} l'_{nT} K(x) Y \quad \ldots(2\text{-}10)$$

إذ أن :

$K(x)$: مصفوفة قطرية $(nT \times nT)$ تضم العناصر $K_{it} = K\left(\frac{(x_{it} - x)}{h} \right)$ لدالة (kernel) اللبية أو الدالة الموزونة بأخذ أقل قيمة لـ x_{it} بعيدة عن x ، وأعلى قيمة لـ x_{it} عندما تكون قريبة x ، وتستخرج من الجدول (1) التالي، وتم أستعمال الدالة اللبية المقترحة (1) .

$\hat{m}(x)$: مقدر الثابت الموضعي اللامعلمي وهي معدل الممهدات لقيم y لـ x_{it} من القيم في الفترة الصغيرة لكي نجعل $\left(\left| x_{it} - x \right| \right) = O\left(\left| h \right|\right)$.

الجدول (1)

بعض دوال (Kernel) اللبية المستعملة . [9]

رقم الدالة	Kernel function	الشكل الظاهري K (u)	حدود الدالة
1	Epanechnikov(or quadratic)	$K(u) = \frac{3}{4}(1-u^2)$	[-1 , 1]
2	Gaussian(or normal)	$K(u)= (2\pi)^{-1/2} \exp(-u^2/2)$	(-∞ , ∞)I
3	Triangular (or Triangle)	K(u)=(1-|u|)	[-1 , 1]
4	uniform (or rectangular or box)	K(u)=1/2	[-1 , 1]
5	Caushy	$(\pi(1+u^2))^{-1}$	[-1 , 1]
6	Double Exp. Or Laplacian	1/2exp(-|u|)	(|u|<∞) I
7	Biweight(or Quartic)	$\frac{15}{16}(1-u^2)^2$	[-1 , 1]
8	Triweight	$\frac{35}{32}(1-u^2)^3$	[-1 , 1]
9	Cosine	$\frac{\pi}{4}\cos(\frac{\pi}{2}u)$	[-1 , 1]
10	Tricubic	$\frac{70}{81}(1-\|u\|^3)^3$	[-1 , 1]
11	Suggest	$\frac{519}{460}(1-u^2)^5$	[-1 , 1]

Blanchet . J and Wadsworth. J . (2012)

أن الدالة المقترحة (11) تعتمد على الصيغة العامة للدالة اللبية (Kernel) $K(x) = c(1-|x|^m)^n, I_{\{|x|\le 1\}}$ ، بإعطاء قيم لـ n=2، m=3 للدالة المقترحة (11) ثم بعد ذلك نوجد قيمة C، وتم كتابة برنامج بلغة الحزم الجاهزة ((R.Version.3.1.2(31/10/2014) بالصيغة العامة والمبين في ملحق البرامج .

وان دالة (Kernel) اللبية يجب إن تحقق الشروط الآتية :

1. $K(u)$ دالة مستمرة وذات قيم حقيقية موجبة .
2. $\int_{-\infty}^{\infty} K(u)du = 1$.
3. $K(u)$ دالة متماثلة حول الصفر، $\mu_2(K) = \int u^2 K(u)du \neq 0$.
4. $\int K^2(u)du < \infty$.

وان مقدرات بعض دوال (Kernel) اللبية المستعملة هي أحدى طرق تمهيد البيانات وهي دالة وزن تضع أوزان في نقاط مختلفة وأعلى وزن للمشاهدات التي فيها قيم (x) قريبة الى النقطة المراد تقديرها (x_0)، وتقل كلما ابتعدت (x) من قيم (x_0)، وان الوزن يحدد من قبل هذه الدوال، وهي تفضل للتأكد من المقدرات لدالة التمهيد $\hat{m}(.)$ تكون ممهدة (على نحو سلس)، وتضمن أن البيانات فقط إلى النقطة يتم تقدير m، وان الخيار الأمثل

الذي أشار له معظم الباحثين في ضل الفرضيات الأساسية هي دالة (Epanechnikov) والتي تعد كونها من أفضل دوال (Kernel) التي تلاؤم اغلب حالات التمهيد، وان اختيار أي دالة تمهيد من دوال (Kernel) تكون مقبولة النتائج لأنها ليست في غاية الأهمية، لان دقة التقدير تعتمد على عرض الحزمة (Bandwith)، وان عرض الحزمة الأمثل لتقليل RAMSE لكل من $\hat{m}(x)$ $h = O\left(n^{-1/(q+4)} \right)$.

خوارزمية أيجاد مقدر الثابت الموضعي اللامعلمي باستعمال توسيع تايلور

(Local Constant Nonparametric Estimator by aTaylor expansion) (LCNE)

يمكن تلخيص خطوات أجراء خوارزمية أيجاد مقدر الثابت الموضعي اللامعلمي باستعمال توسيع تايلور كالأتي :

1 . وصف عرض حزمة ملائم (h) .

2 . أيجاد المقدر اللبي ناداريا واتسون ناداريا واتسون بالصيغة (2-10) من خلال أستعمال الدالة اللبية المقترحة (11) .

المقدر الخطي الموضعي اللامعلمي باستعمال توسيع تايلور [4]، [6]، [9]، [11] (Local Linear Nonparametric Estimator by aTaylor expansion) (LLNE)

لإيجاد تقدير أنموذج البيانات المزدوجة اللامعلمية للتأثيرات العشوائية (2-5)، وبتوسيع m حول ($x_{it} = x$) تستخدم مركبات الخطأ u_{it}

$$y_{it} = m(x) + (x_{it} - x) B(x) + v_i + u_{it}$$

$$y_{it} = \left(1, (x_{it} - x)\right)\left(m(x), B'(x)\right)' + v_i + u_{it}$$

$$y_{it} = z_{it}\ \delta(x) + \varepsilon_{it} \qquad \ldots(2\text{-}11)$$

إذ أن :

$$\varepsilon_{it} = v_i + u_{it}$$

$$z_{it} = \left(1, (x_{it} - x)\right)$$

$$\delta(x) = \left(m(x), B'(x)\right)'$$

$$y_i = \left(y_{i1}, y_{i2}, \ldots\ldots\ldots\ldots, y_{iT}\right)'$$

$$z_i = \left(z'_{i1}, z'_{i2}, \ldots\ldots\ldots\ldots, z'_{iT}\right)'$$

$$\varepsilon_i = \left(\varepsilon_{i1}, \varepsilon_{i2}, \ldots\ldots\ldots\ldots, \varepsilon_{iT}\right)'$$

$$Y = \left(y'_1, y'_2, \ldots\ldots\ldots\ldots, y'_n\right)'$$

$$Z(x) = \left(Z'_1, Z'_2, \ldots\ldots\ldots\ldots, Z'_n\right)'$$

$$\varepsilon = \left(\varepsilon'_1, \varepsilon'_2, \ldots\ldots\ldots\ldots, \varepsilon'_n\right)'$$

ويمكن كتابة الصبغة (2-11) بصيغة المصفوفات وكالاتي :

$$Y=Z(x)\,\delta(x)+\varepsilon \qquad \ldots(2\text{-}12)$$

ولكون :

$$V=\mathrm{E}(\varepsilon_i\,\varepsilon'_i)=\begin{pmatrix}\sigma_v^2+\sigma_u^2 & \sigma_v^2 & \cdots & \sigma_v^2\\ \sigma_v^2 & \sigma_v^2+\sigma_u^2 & \cdots & \sigma_v^2\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_v^2 & \sigma_v^2 & \cdots & \sigma_v^2+\sigma_u^2\end{pmatrix}=\sigma_v^2 J_T+\sigma_u^2 I_T$$

$$\mathrm{E}(\varepsilon_i\,\varepsilon'_j)=O_{T\times T} \qquad i\neq j$$

فان مصفوفة التباين والتباين المشترك (ε) تكون :

$$\Omega=\mathrm{E}(\varepsilon\,\varepsilon')=\begin{pmatrix}\sigma_v^2 J_T+\sigma_u^2 I_T & 0 & \cdots & 0\\ 0 & \sigma_v^2 J_T+\sigma_u^2 I_T & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \sigma_v^2 J_T+\sigma_u^2 I_T\end{pmatrix}$$

$$\Omega=I_n\otimes V$$

إذ أن :

$$l_T\,l'_T=J_T$$

وبالتالي

$$\Omega^{-1}=I_n\otimes V^{-1}$$

إذ أن :

$$V^{-1}=\frac{1}{\sigma_u^2}\left[I_T+(1-\lambda)\,l_T\,l'_T/T\right]$$

$$\lambda=\sigma_u^2/\left(\sigma_u^2+T\sigma_v^2\right)$$

ولكون

$$V^{-\frac{1}{2}}=\left[I_T-(1-\lambda^{-\frac{1}{2}})\,l_T\,l'_T/T\right]$$

نحصل على :

$$\Omega^{-\frac{1}{2}}=I_n\otimes V^{-\frac{1}{2}}=I_n\otimes\left[I_T-(1-\lambda^{-\frac{1}{2}})\,l_T\,l'_T/T\right]$$

$$=\left[I_{nT}-(1-\lambda^{-\frac{1}{2}})\left[I_n\otimes(l_T\,l'_T)\right]/T\right]$$

$$=\left[I_{nT}-(1-\lambda^{\frac{1}{2}})\,(I_n\otimes l_T)(I_n\otimes l'_T)/T\right]$$

$$=\left[I_{nT}-(1-\lambda^{\frac{1}{2}})\,DD'/T\right]$$

إذ أن :

$$D=I_n \otimes l_T \ , \Omega^{-\frac{1}{2}} \Omega^{-\frac{1}{2}} = \Omega^{-1}$$

ومن الصيغة (2-12) نحصل على :

$$\Omega^{-\frac{1}{2}} Y = \Omega^{-\frac{1}{2}} Z(x)\, \delta(x) + \Omega^{-\frac{1}{2}} \varepsilon$$

$$Y^* = Z^*(x)\, \delta(x) + \varepsilon^* \qquad \ldots(2\text{-}13)$$

وبتحقق الشرط :

$$\mathrm{E}(\varepsilon^* \varepsilon^{*\prime}) = \mathrm{E}\left[\Omega^{-\frac{1}{2}} \varepsilon \, (\Omega^{-\frac{1}{2}} \varepsilon)' \right] = \Omega^{-\frac{1}{2}} \mathrm{E}(\varepsilon \varepsilon')\, \Omega^{-\frac{1}{2}} = I_{nT}$$

وتكتب الصيغة (2-13) بدلالة المصفوفات :

$$y^*_{it} = z^*_{it}\, \delta(x) + \varepsilon^*_{it} \qquad \ldots(2\text{-}14)$$

فان مقدر التأثيرات العشوائية اللامعلمية $m(x)$ ، $B(x)$ للصيغة (2-13) وبتصغير مجموع مربعات الخطأ نحصل على :

$$\sum_{i=1}^{n} \sum_{t=1}^{T} \left(y^*_{it} - z^*_{it}\, \delta(x) \right)^2 K\left(\frac{x_{it} - x}{h} \right)$$

$$= \left(Y^* - Z^*(x)\, \delta(x) \right)' K(x) \left(Y^* - Z^*(x)\, \delta(x) \right)$$

$$= \left(Y - Z(x)\, \delta(x) \right)' \Omega^{-\frac{1}{2}} K(x)\, \Omega^{-\frac{1}{2}} \left(Y - Z(x)\, \delta(x) \right)$$

إذ أن :

$$Y^* = \Omega^{-\frac{1}{2}} Y \quad ، \quad Z^*(x) = \Omega^{-\frac{1}{2}} Z(x)$$

$$y^*_{it} = y_{it} - (1 - \lambda^{\frac{1}{2}})\, \bar{y} i. \quad ، \quad z^*_{it} = z_{it} - (1 - \lambda^{\frac{1}{2}})\, \bar{z} i.$$

$$\lambda = \frac{\sigma_u^2}{\left(\sigma_u^2 + T \sigma_v^2 \right)}$$

وبالتالي :

$$\hat{\delta}(x) = \left[\sum_{i=1}^{n} \sum_{t=1}^{T} z^*_{it}{}' \, z^*_{it} \, K\left(\frac{x_{it} - x}{h} \right) \right]^{-1} \left[\sum_{i=1}^{n} \sum_{t=1}^{T} z^*_{it}{}' \, y^*_{it} \, K\left(\frac{x_{it} - x}{h} \right) \right]$$

$$\hat{\delta}(x) = \left(Z^{*\prime}(x)\, K(x)\, Z^*(x) \right)^{-1} Z^{*\prime}(x)\, K(x)\, Y^*$$

$$\hat{\delta}(x) = \left(Z'(x)\, \Omega^{-\frac{1}{2}} K(x)\, \Omega^{-\frac{1}{2}} Z(x) \right)^{-1} \left(Z'(x)\, \Omega^{-\frac{1}{2}} K(x)\, \Omega^{-\frac{1}{2}}\, Y \right) \ldots(2\text{-}15)$$

وللحصول على تقدير $\delta(x)$ نقدر قيمة λ :

$$\hat{\lambda} = \frac{\hat{\sigma}_u^2}{\left(\hat{\sigma}_u^2 + T\, \hat{\sigma}_v^2 \right)}$$

إذ أن :

$$\hat{\sigma}_u^2 = \frac{1}{n(T-1)} \sum_{i=1}^{n} \sum_{t=1}^{T} \left[y_{it} - \bar{y}i. - (x_{it} - \bar{x}i.)\hat{B}_{FE}(x) \right]^2$$

ومن أنموذج الانحدار:

$$y_{it} - \bar{y}i. = (x_{it} - \bar{x}i.)B(x) + u_{it} - \bar{u}i.$$

إذ أن :

$$\hat{\sigma}_v^2 = \left(\hat{\sigma}_\eta^2 - (\hat{\sigma}_u^2 / T) \right)$$

$$\hat{\sigma}_\eta^2 = \left(\sum_{i=1}^{n} (\bar{y}i. - \hat{m}(\bar{x}i.))^2 / n \right)$$

فان $\tilde{m}(x)$ نحصل عليها بأجراء تقدير المربعات الصغرى الموضعية للأنموذج :

$$\bar{y}i. = m(x) + (\bar{x}i. - x)B(x) + v_i + \bar{u}i.$$

مشروطة بـ $x_{it.}$ لكل x معطى ∋ $Supp(x)$ ،

$$V(\hat{\delta}(x))$$
$$= (Z^{*\prime}(x)\, K(x)\, Z^{*}(x))^{-1} Z^{*\prime}(x)\, K^2(x) Z^{*}(x) (Z^{*\prime}(x)\, K(x)\, Z^{*}(x))^{-1}$$

إذ أن مقدرات التأثيرات العشوائية اللامعلمية للـ $m(x)$، $B(x)$ تكون كالأتي :

$$\hat{m}_{RE}(x) = (1, 0_q') \hat{\delta}(x)$$

$$\hat{B}_{RE}(x) = (0, 1_q') \hat{\delta}(x)$$

خوارزمية أيجاد المقدر الخطي الموضعي اللامعلمي باستعمال توسيع تايلور

(Local Linear Nonparametric Estimator by aTaylor expansion) (LLNE)

ويمكن تلخيص خطوات أجراء خوارزمية أيجاد المقدر الخطي الموضعي اللامعلمي باستعمال توسيع تايلور كالأتي

1 . وصف عرض حزمة ملائم (h_0, h_1) .

2 . أيجاد المقدر اللبي لناداريا واتسون ناداريا واتسون $\hat{\delta}(x)$ بالصيغة (2−15) من خلال أستعمال دالة (Epanechnikov) (or quadratic) بالاعتماد على أيجاد الصيغ (2−12)، (2−13)، (2−14) .

طريقة التقدير ذات المرحلتين **(4)، (6)، (9)، (11)**

(Nonparametric Two – Step Estimation Method in panel data) (Two – Step)

تم أقتراح الطريقة من قبل (Henderson and Ullah) (2005)، وتم تطوير الطريقة بتناول الجانب التطبيقي من قبل ((2014) , Huansha Wang) لإيجاد تقدير أنموذج البيانات المزدوجة اللامعلمية للتأثيرات العشوائية في الصيغة (2−5) :

$$y_{it} = m(x_{it}) + v_i + u_{it} \qquad , i = 1,2,\ldots\ldots\ldots, n \quad , t = 1, 2, \ldots\ldots, T$$

$$y_{it} = m(x_{it}) + w_{it} \qquad , i = 1,2,\ldots\ldots\ldots, n \quad , t = 1, 2, \ldots\ldots, T$$

بإعادة كتابة الصيغة أعلاه بالصورة :

$$y = m + v + u$$
$$y = m + w \qquad \ldots (2-16)$$

إذ أن :

y ،u : متجه $nT \times 1$.

$v = \left(v_1 l_T', v_2 l_T', \ldots\ldots\ldots\ldots, v_n l_T' \right)'$ ، l_T متجه الوحدة بطول T .

$$m = \left(m(x_{11}), \ldots\ldots, m(x_{1T}), \ldots\ldots\ldots, m(x_{n1}), \ldots\ldots\ldots, m(x_{nT}) \right)'$$

$$w = v + u = \left(w_1', w_2', \ldots\ldots\ldots, w_n' \right)$$

$$w_i = \left(w_{i1}, w_{i2}, \ldots\ldots\ldots, w_{iT} \right)' \qquad , \qquad i = 1, 2, 3, \ldots\ldots, n.$$

وان المقدر الخطي الموضعي اللامعلمي (LL) لـ $\delta(x)$ يكون بالصيغة الأتية :

$$\hat{S}_{LL}(x) = \left(Z'(x) K(x) Z(x) \right)^{-1} Z'(x) K(x)\, y \qquad \ldots (2-17)$$

إذ أن :

$Z(x) = \left(1, (x_{it} - x)' \right)'$ مصفوفة من الدرجة $(p+1) \times nT$ ، $t = 1, 2, .., T$ ، $i = 1, 2, .., n$.

$K_x = diag\{ K_h(x_{11} - x), \ldots\ldots, K_h(x_{nT} - x) \}$ الدالة اللبية (Kernel) الموزونة وان تحديد الأوزان تمثل معالجة لمشكلة عدم تجانس فتم أستعمال دالة (Epanechnikov) وهي مصفوفة قطرية من الدرجة $nT \times nT$ ، لغرض استبعاد أثر عدم التجانس في متغيرات الدراسة .

ومن المقدر الخطي الموضعي اللامعلمي (LL) نحصل على

$$\hat{m}_{LL}(x) = e' \hat{\delta}_{LL}(x) \qquad \ldots (2-18)$$

إذ إن $e' = (1, 0, \ldots, 0)$ هو متجه من الدرجة $(p+1) \times 1$ والذي يكون فيه أول حد مساوي إلى الواحد والعناصر المتبقية أصفار وان هذه الخطوات تجرى في المرحلة الأولى، ويمكن تحسين المقدر لان مصفوفة التباين والتباين المشترك للخطأ (w) لم يتم أستغلالها، وان مصفوفة التباين والتباين المشترك للخطأ (w) تكتب بالصيغة الآتية :

$$\mathrm{E} w w' = \Sigma = \sigma_v^2 \left(I_n \otimes J_T \right) + \sigma_u^2 \left(I_n \otimes I_T \right) = \sigma_1^2 P + \sigma_u^2 Q \qquad \ldots (2-19)$$ إذ أن:

I_l : مصفوفة الوحدة بإبعاد l ، $l = nT \,.\, nT$.

J_T : مصفوفة الوحدة بإبعاد T.

$$P = I_n \otimes \bar{J}_T$$
$$Q = I_{nT} - P$$
$$\bar{J}_T = \frac{J_T}{T}$$

$$\sigma_1^2 = T\sigma_v^2 + \sigma_u^2$$

وان $\sum^{-1/2}$ تكون بالصيغة الأتية :

$$\sum^{-1/2} = \frac{1}{\sigma_1} P + \frac{1}{\sigma_u} Q \qquad \ldots (2-20)$$

بإضافة وطرح w $\sum^{-\frac{1}{2}}$ للصيغة (2-20) نحصل على :

$$y = m - \sum^{-\frac{1}{2}} w + w + \sum^{-\frac{1}{2}} w$$
$$y = m + \left(1 - \sum^{-\frac{1}{2}}\right) w + s$$
$$y = m + H\ w + s$$

إذ أن :

$$s = \sum^{-\frac{1}{2}} w$$
$$H = \left(1 - \sum^{-\frac{1}{2}}\right)$$
$$\mathrm{E}(s) = 0 \quad , \quad \mathrm{E}(ss') = 1$$

وبالتالي نحصل على الصيغة الأتية :

$$y^* = y - H\ w \qquad \ldots (2-21)$$
$$y^* = m + s$$

نجري تقدير المقدر الخطي الموضعي (LL) أولا للحصول على انحدار البواقي $\hat{w}$ ، وكذلك نحصل على $\hat{H} = \left(1 - \hat{\sum}^{-\frac{1}{2}}\right)$ بتقدير $\hat{\sum}$ من المقدرات $\hat{\sigma}_u^2$ ، $\hat{\sigma}_v^2$ ، وان :

$$\hat{\sigma}_u^2 = \frac{1}{N(T-1)} \sum_i \sum_t \left(\hat{w}_{it} - \hat{\bar{w}}_i\right)^2$$

إذ أن :

$$\hat{\bar{w}}_i = \frac{\sum_t \hat{w}_{it}}{T} .$$

$\hat{w}_{it} = y_{it} - \hat{m}(x_{it})$: يحتوي على th t , i من عناصر $\hat{w}$.

وبالإضافة إلى ذلك يتم الحصول على ($\sigma_1^2 = T\sigma_v^2 + \sigma_u^2$) عندما($\hat{\sigma}_1^2 = \frac{T\sum_t \hat{\bar{w}}_i^2}{N}$)،

والتي تعطي $\hat{\sigma}_v^2 = \frac{\left(\hat{\sigma}_1^2 - \hat{\sigma}_u^2\right)}{T}$ ويعطي هذا الإجراء $\hat{y}^* = y - \hat{H}\ \hat{w}$ وبالتالي نحصل على مقدر ذات المرحلتين (Two – Step) وكالاتي :

$$\hat{S}_{2S}(x) = \left(Z'(x)K(x)Z(x)\right)^{-1} Z'(x)K(x)\ \hat{y}^* \qquad \ldots (2-22)$$

ومن المقدر الخطي الموضعي اللامعلمي (LL) نحصل على المقدر

$$\hat{m}_{2S}(x) = e'\hat{\delta}_{2S}(x) \qquad \ldots (2-23)$$

وأخيرا أن مقدرات طريقة التقدير ذات المرحلتين (Two–Step) يمكن الحصول عليها بالمرحلة الأولى من خلال أيجاد المقدر الخطي الموضعي (LL) ومن ثم تحسين المقدر باستعمال مصفوفة التباين والتباين المشترك للخطأ (w) في المرحلة الثانية .

خوارزمية أيجاد طريقة التقدير ذات المرحلتين (Two – Step)

(Algorithm Nonparametric Two – Step Estimation Method in panel data)

يمكن تلخيص خطوات أجراء خوارزمية أيجاد طريقة التقدير ذات المرحلتين (Two – Step) كالأتي :

1 . وصف عرض حزمة ملائم (h) .

2 . ايجاد $\hat{S}_{LL}(x)$ بالصيغة (2-17) من خلال أستعمال دالة (Epanechnikov) .

3 . أيجاد $\hat{m}_{LL}(x)$ بالصيغة (2-18) .

4 . ايجاد $\hat{S}_{2S}(x)$ بالصيغة (2-22) بالاعتماد على أيجاد الصيغ (2-19)، (2-20)، (2-21) .

5 . أيجاد مقدر ذات المرحلتين $\hat{m}_{2S}(x)=e'\hat{\delta}_{2S}(x)$ (Two–Step) بالصيغة (2-23) .

اختيار معلمة التمهيد (عرض الحزمة) [1]، [2]، [6]، [7]

(Choosing the Smoothing Parameter or Bandwidth)

أن اختيار عرض الحزمة الملائمة (المعلمة التمهيدية) (h) أو معلمة الانتشار أو معلمة التركيز أو معلمة التباين، مهم لتقريب دالة الانحدار اللامعلمي للدالة الأصلية بحيث يكون الخطأ اقل مايمكن وهو دائما نقطة البداية الجيدة، وان دقة التمهيد تعتمد فقط على عرض الحزمة، وان الموازنة بين التحيز والتباين تسمى معلمة التمهيد أو معلمة عرض الحزمة، فإذا كان عرض الحزمة صغير جدا، نحصل على تعظيم للتباين وتصغير للتحيز، وبالتالي فان المقدر يكون مضطرب للغاية، وعندما يكون عرض الحزمة كبير جد، يحدث العكس، وبالتالي فان المقدر يكون خارج التمهيد، إذ أن عرض الحزمة تتحكم بالبعد والقرب حول x ، وبالتالي تتحكم بكمية التمهيد للمقدر الناتج، فإذا كانت (h) كبيرة نختار قيم كبيرة حول x ، وبالعكس .

معيار تقاطع الشرعية (العبور الشرعي) (Leave – One–Out Cross – Validation)

بعد أن يتم وصف طريقة التقدير لأنموذج التأثيرات العشوائية اللامعلمية للبيانات المزدوجة (2-5)، هناك مشكلة خطيرة هي في اختيار عرض المعلمة التمهيدية (h)، وبالتالي نحن بحاجة لايجاد طريقة سريعة لتحديد دقيق لمعلمة التمهيد، ومعالجة لمشكلة أختيار عرض الحزمة، إذ سيتم تناول معيار تقاطع الشرعية (العبور الشرعي) في حالة (Leave – one–out cross validation) ويسمى ايظا (C V)، وهو من أشهر الطرق التي تقيس جودة الأداء لمقدرات دالة الانحدار اللامعلمية يوصف بالصيغة التالية :

$$LCV(h)=\frac{1}{n}\sum_{i=1}^{n}(\hat{m}_{(-1)}(x_i)-y_i)^2 \qquad 2-24$$

إذ أن : $\hat{m}_{(-1)}(x_i)$: دالة التقدير التمهيدية باستبعاد (2L+1) من مشاهدات المتغير التوضيحي وان (L =0)، وبالتالي فان قيمة (h) هي التي تقابل اقل (C V) .

معايير المقارنة [2]، [4]، [7]

لغرض أيجاد أفضل طريقة لتقدير المعلمات بالنسبة لأنموذج البيانات المزدوجة اللامعلمية للتأثيرات العشوائية في الصيغة (2-5) فأننا سنستعمل المعيارين المذكورة لاحقاً والتي تناولتها أغلب البحوث المنشورة والتي تضمنت موضوع الدراسة، إذ أن تناقص قيم هذه المعايير يشير بالتأكيد إلى جودة التقدير وهي كما يلي :

معيار (RAMSE) (N^{th} Root of the Average of Mean Square Error)

يمثل جذر N لمعدل متوسط مربعات الخطأ

ويأخذ الصيغة الأتية :

$$RAMSE = \sqrt[N]{AMSE} \qquad \dots (2-25)$$

إذ أن:

(N) : يمثل حجم العينة .

(AMSE) : عبارة عن معدل متوسط مربعات الخطأ.

AMAPEمعيار هو عبارة عن معدل متوسط المجموع الكلي للقيم المطلقة لأنحرافات القيم التقديرية عن القيم الحقيقية مقسوماً على القيم الحقيقية وتتمثل قيمته بما يلي :

$$AMAPE = \frac{1}{R}\sum_{i=1}^{R} MAPE(i) \qquad \dots (2-26)$$

أذ أن : (R) : يمثل عدد مرات تكرار التجربة i .

$(MAPE(i))$: تمثل متوسط الخطأ النسبي المطلق في التجربة i وتكون صيغته كما يلي:

$$MAPE(i) = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\theta - \hat{\theta}}{\theta}\right|$$

n . يمثل حجم العينة

$\hat{\theta}$. تشير الى المعلمات المقدرة بواسطة المحاكاة

θ . تشير الى القيم الحقيقية المحتسبة

3 .الجانب التطبيقي: يركز الجانب التطبيقي على حساب المقدرات اللامعلمية لإيجاد كلاً من مقدر الثابت الموضعي اللامعلمي باستعمال توسيع تايلور (Local Constant Nonparametric Estimator by a Taylor expansion) (LCNE)، والمقدر الخطي الموضعي اللامعلمي باستعمال توسيع تايلور (Local Linear Nonparametric Estimator by a Taylor expansion) (LLNE)، ومقدر طريقة التقدير ذات المرحلتين (Two – Step) (Nonparametric Two – Step Estimation Method in panel data)، واستعمال معيار تقاطع الشرعية (العبور الشرعي) لإيجاد عرض الحزمة للطرائق وأخيرا تم تناول أستعمال المعايير (RAMSE)،(AMAPE)، للمقارنة بين طرائق التقدير للأنموذج, فقد تضمن الجانب التطبيقي أخذ بيانات خاصة بالموازنة العامة للدولة مقاسه بملايين الدولارات وتأثيرها في عدد من المتغيرات التفسيرية من التقرير الاقتصادي العربي الموحد(2002–2014) (صندوق النقد الدولي وأخرون 2014) .

وصف البيانات :أن فكرة التطرق للموازنة العامة للدولة جاءت لأنها أداة مالية تستخدمها الدولة لتحقيق مجموعة من الأهداف الاقتصادية والسياسية والتعجيل بها، كما إن لديها أهداف اجتماعية كبيرة فهي وسيلة لإعادة توزيع الدخل القومي عن طريق فرض الضرائب التصاعدية وإعادة أنفاقها لتغطية نفقات الطبقات الفقيرة وهناك أهداف مالية أخرى للموازنة تعبر عن كشف المركز المالي للدولة وتبين فيما أذا كان عجزاً أم فائضاً في الموازنة، ونتيجة لذلك تم دراسة الموازنة العامة للدولة مقاسه بملايين الدولارات الأمريكية كمتغير أستجابة (y) ولعدد من المتغيرات التفسيرية وحسب وجهة نظر أهل الاختصاص وهي المتوسط السنوي لسعر النفط لسلة خامات أوبك مقاسه بالدولار الأمريكي (X_1)، والناتج المحلي الإجمالي بأسعار السوق الجارية مقاسه بالدولار الأمريكي (X_2)، وعدد السكان مقاساً بمليون نسمة (X_3)، الصادرات السلعية (X_4) مقاسه بالدولار الأمريكي، وتمثل البيانات التي تم استعمالها في الجانب التطبيقي بيانات طولية متزنة (Balanced Panel Data) لان المشاهدات المقطعية (الدول العربية) مقاسه لنفس الفترات الزمنية (سنة واحدة) والموجودة في الملحق الثاني، وشملت الدراسة الدول

العربية البالغ عددهم ($N=17$) باستثناء العراق الذي لم تتوفر الإحصائيات الكاملة عنه، والصومال وجيبوتي وفلسطين وجزر القمر لان ليس هناك دقة في تسجيل الإحصائيات، وللمدة الزمنية (2002-2014) ($T=13$) وبحجم عينة ($N\times T=221$) مشاهدة تم الحصول عليها بالاعتماد على التقرير الاقتصادي العربي الموحد للسنوات (2002-2014) .

وصف أنموذج البيانات :

لغرض تحليل البيانات الموصوفة في الفقرة (3-1) سيتم تناول أنموذج البيانات المزدوجة اللامعلمية للتأثيرات العشوائية ليتوافق مع هيكلية البيانات الطولية المتزنة فالأنموذج سيكون كالآتي :

$$y_{it}=m(x_{it})+v_i+u_{it} \qquad ,i=1,2,.....,17 \quad ,t=1,2,.....,13 \qquad \dots (3-1)$$

تحليل البيانات المعتمدة في البحث :فيما يخص بيانات الموازنة العامة للدولة الموضحة بالملحق (2)

1. بالاعتماد على قيم اختبار(1996) (Zheng) في الصيغة (8 -2) يلاحظ وجود مشكلة عدم تجانس التباين (Heteroskedasticity) للأنموذج وكالاتي :-

جدول (1-3)

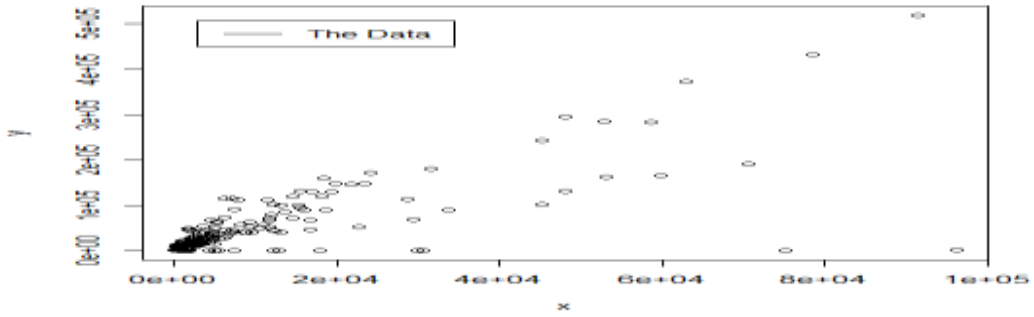
Model	R Square	Adjusted R Square	Std.Error of the Estimate	P-Value (Heteroskedasticity test)	P-Value (A Nonparametric Hausman Test)
NREM	0.727174	0.699891	38278.92	4.53244e-012	0.0901862

من الجدول (1-3) نجد أن قيمة ($p-value=4.53244e-012$) اقل من 0.05 لذا نرفض فرضية العدم ونقبل الفرضية البديلة وهذا يعني انه توجد مشكلة عدم التجانس في تباين بواقي الأنموذج .

2. من الجدول (1-3) بالاعتماد على قيم اختبار (A Nonparametric Hausman Test) في الصيغة (9-2) نجد أن قيمة ($p-value=0.0901862$) أكبر من 0.05 لذا نقبل فرضية العدم وهذا يدل على أن الأنموذج النهائي الملائم لبيانات البحث هو أنموذج التأثيرات العشوائية وبالتالي سيتم تناول المقدرات اللامعلمية لأنموذج التأثيرات العشوائية اللامعلمية للبيانات المزدوجة في حالة وجود مشكلة عدم التجانس .

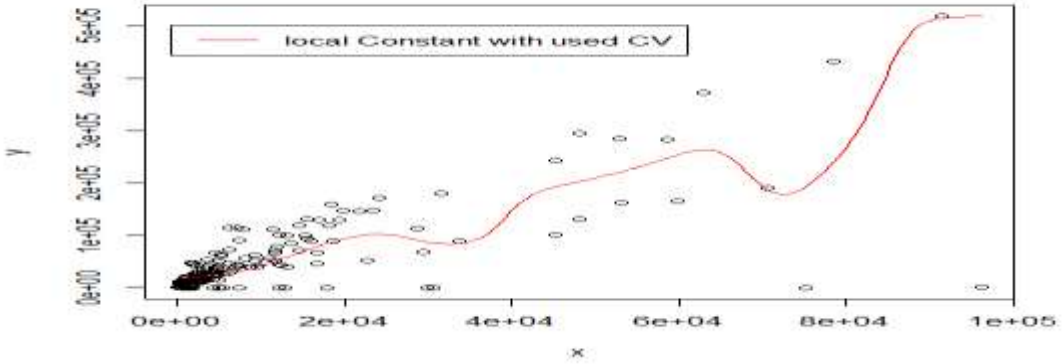
تقدير المعاملات وتحليل النتائج :

فيما يخص بيانات الموازنة العامة للدول العربية الموضحة بالملحق (2)، وبعد تعريف ووصف البيانات وكذلك تعريف الأنموذج سيتم أيجاد المقدرات لأنموذج البيانات المزدوجة اللامعلمية للتأثيرات العشوائية باستعمال طرائق التقدير التي تم تناولها في البحث , ولانجاز ذلك قام الباحث بكتابة برنامج لتقدير لهذا الأنموذج باستعمال برنامج الحزم الجاهزة ((31/10/2014) R.Version.3.1.2) بالملحق (1)، والأشكال التالية تشير إلى رسم نقاط انتشار البيانات الحقيقية للموازنة العامة للدول العربية بالإضافة إلى المقدرات اللامعلمية لأنموذج البيانات المزدوجة اللامعلمية للتأثيرات العشوائية التي تم تناولها في البحث إذ إن (y) تمثل البيانات الحقيقية للموازنة العامة للدول العربية، (x) تمثل كلاً من المتوسط السنوي لسعر النفط، الناتج المحلي الإجمالي، عدد السكان، الصادرات السلعية، للمدة (2002-2014) . وكالأتي :



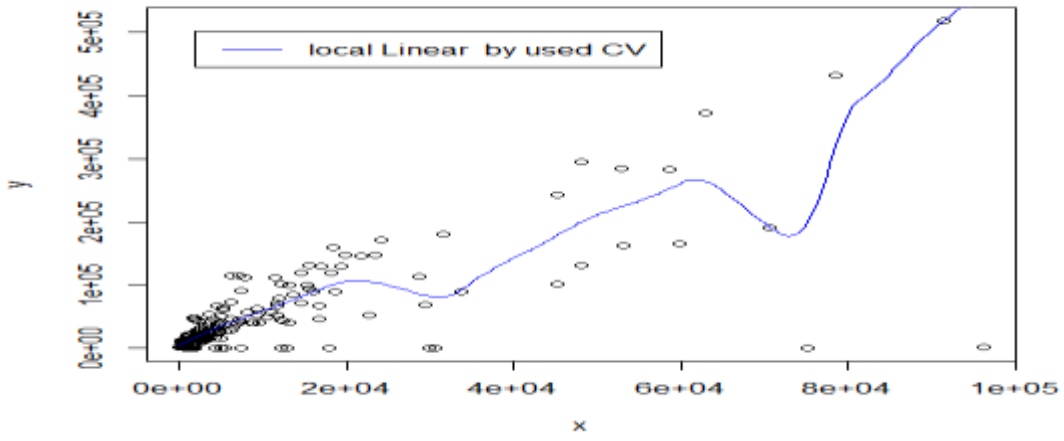
شكل (3-1)

نقاط انتشار البيانات الحقيقية للموازنة العامة للدول العربية (y)، وكلاً من المتوسط السنوي لسعر النفط، الناتج المحلي الإجمالي، عدد السكان، الصادرات السلعية (x) للمدة (2002-2014) .



شكل (3-2)

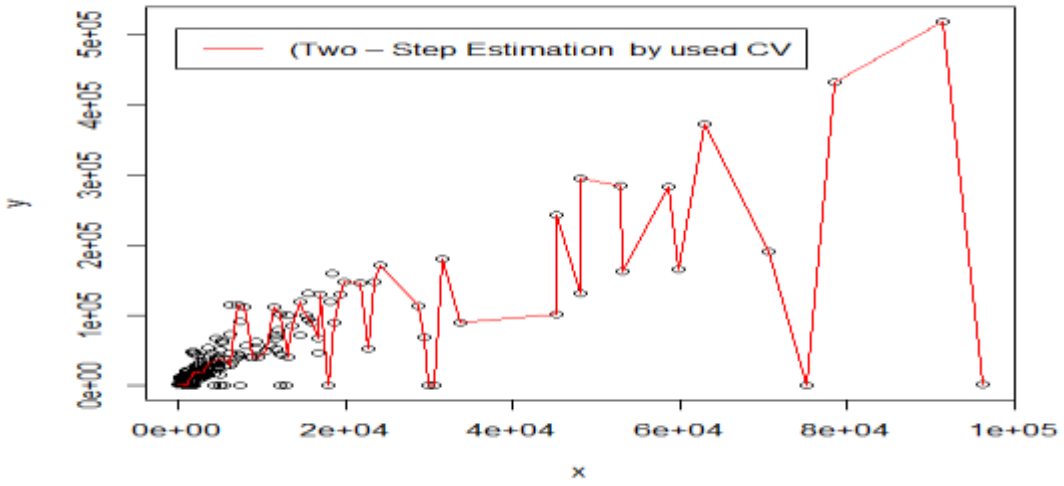
مقدر الثابت الموضعي اللامعلمي باستعمال توسيع تايلور للموازنة العامة للدول العربية للمدة (2002-2014)
إذ إن (y) تشير إلى للموازنة العامة للدول العربية، (x) يشير إلى كلاً من المتوسط السنوي لسعر النفط، الناتج المحلي الإجمالي، عدد السكان ،الصادرات السلعية .



شكل (3 - 4)

مقدر طريقة التقدير ذات المرحلتين للموازنة العامة للدول العربية للمدة (2002-2014) .

إذ إن (y) تشير إلى للموازنة العامة للدول العربية،(x) يشير إلى كلا من المتوسط السنوي لسعر النفط، الناتج المحلي الإجمالي عدد السكان، الصادرات السلعية .



جدول (2-3)

المعاير المعتمدة في طريقة التقدير اللامعلميه الأفضل لإيجاد تقدير أنموذج التأثيرات العشوائية اللامعلمية للبيانات المزدوجة

المعايير المعتمدة			
طريقة التقدير	RAMSE	AMAPE	H
LCNE	461.086	174.086	0.4758496
LLNE	471.0693	1809.112	0.4758496
Two-Step	299.8411	780.4272	0.4758496

الاستنتاجات:

في ظل تحليل نتائج الجانب التطبيقي تم التوصل إلى الاستنتاجات الآتية :

1. من الجدول (1-3) يلاحظ ان الأنموذج يعاني من مشكلة عدم التجانس وان الأنموذج النهائي الملائم لبيانات البحث هو أنموذج التأثيرات العشوائية بالاعتماد على قيم اختبار هاسمان اللامعلمي .
2. من الشكل (1-3) نلاحظ أن النقاط للبيانات منتشرة حول المنحني باتجاه محدد مما يدل على وجود علاقة غير خطية بين كل من المتغيرين (x,y)، و من خلال الشكل (2-3)، (3-3)، (4-3) نلاحظ ديناميكية تغير منحني كل من مقدر الثابت الموضعي اللامعلمي باستعمال توسيع تايلور(LCNE) ، والمقدر الخطي الموضعي اللامعلمي باستعمال توسيع تايلور (LLNE)، و مقدر طريقة التقدير ذات المرحلتين (Two-Step)، ويتم ربط التقديرات بشكل خطي متصاعد تاره ومتنازل تارة أخرى يعكس التذبذب في المتوسط السنوي لسعر النفط، الناتج المحلي الإجمالي، عدد السكان، الصادرات السلعية، للمدة (2002-2014) .
3. من الجدول (2-3) نلاحظ إن أكبر القيم لمتوسط مربعات الخطأ كانت مع كل من مقدر الثابت والمقدر الخطي الموضعي اللامعلمي باستعمال توسيع تايلور لتأثرها بنقاط الحد مما سبب تلك الزيادة وبالتالي عدم كفاءة المقدرين .
4. بالاعتماد على معيار العبور الشرعي المعمم (GCV)، تبين أن أفضل عرض حزمة كان (h=0.4758496 والذي تم اعتماده للمقدرات الثلاث .
5. تبين أن مقدر طريقة التقدير ذات المرحلتين لانموذج التأثيرات العشوائية اللامعلمية أفضل من مقدر الثابت الموضعي اللامعلمي باستعمال توسيع تايلور و المقدر الخطي الموضعي اللامعلمي باستعمال توسيع تايلور بالاعتماد على كل من معيار جذر معدل متوسط مربعات الخطأ(RAMSE) إذ حقق أقل قيمة (299.8411)، ومعيار (AMAPE) إذ حقق أقل قيمة (780.4272)، والسبب في كبر قيم المعايير هو القيم الكبيرة للبيانات الخاصة بموازنات الدول حيث أخذت بملايين الدولارات الأمريكية .

ومن هنا نستنتج أفضلية أستعمال مقدر طريقة التقدير ذات المرحلتين في تقدير الموازنة العامة للدول العربية للمدة (2002-2014) .

التوصيات :

1. نوصي باستعمال طرائق التقدير المذكورة في البحث في تقدير نماذج الانحدار اللامعلمية و الشبه المعلمية .
2. نوصي باستعمال طرق تقدير لامعلمية لمعالجة مشكلتي الارتباط الذاتي والتعدد الخطي ووجود القيم الشاذة أو الملوثة .
3. أجراء دراسات مستقبلية تتناول مقارنة طرائق لامعلمية أخرى مثل طريقة تمهيد الشريحة التكعيبية Cubic (Smoothing Spline) وطريقة الموجة الصغيرة (Wavelet) مع الطرائق التي تم اعتمادها في البحث .

المصادر

1. يوسف، خلود يوسف خمو، (2004)، "مقارنة اساليب بيز مع طرائق اخرى لتقدير منحنى الانحدار اللامعلمي"، اطروحة دكتوراه في الاحصاء، كلية الادارة والاقتصاد، جامعة بغداد..

2**.** Bai, J. (2009). Panel data models with interactive fixed effects. Econometrica 77, 1229-1279.

3**.** Baltagi, B.H. and Q. Li. (1996). On instrumental variable estimation of miparametric dynamic panel data models. Economics Letters 76: 1- 9.

4. Chad Schafer and Larry Wasserman, (2014) . " Tutorial on Nonparametric Inference With R " Carnegie Mellon University working paper .

5. Delaigle , A . and Meister , A. (2007) . " Nonparametric Regression Estimation in the Heteroscedastic Errors-in-Variables Problem " , Journal of the American Statistical Association .

6. Fox , J. (2005) . " Introduction to Nonparametric Regression " , McMaster University Canada , Journal of Economic ans Social Research Council .

7. Guidoum, C, . (2015) . " Kernel Estimator and Bandwidth Selection for Density and its Derivatives " , working paper .

8. Henderson, D.J. and Carroll , Qi. and Li. (2008)." Nonparametric estimation and testing of fixed effects panel data models". J Econom. 2008 ; 144(1): 257–275. doi:10.1016/j.jeconom.2008.01.005 .

9. Huansha Wang , (2014) " . Essays on Semiparametric Ridge-Type Shrinkage Estimation, Model Averaging and Nonparametric Panel Data Model Estimation" A Dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Economics , Uuiversity of California .

10**.** Mundra, K. (2005). Nonparametric slope estimation for fixed effects panel data. Working Paper , Department of Economics, San Diego State University.

11**.** Su, L., A. Ullah, and Y. Wang. (2010). A note on nonparametric regression estimation with general parametric error covariance. Working paper, School of Economics, Singapore Management Univ.

12**.** Tran, K. C. and E. G. Tsionas. (2010). Local GMM estimation of semiparametric panel data with smooth coefficient models.Econometric Reviews 29: 39-61.

13**.**Wasserman , Larry (2006). All of Nonparametric Statistics. Berlin : Springer-Verlag .