مقارنة بعض طرائق تقدير إنموذج شبه معلمي لبيانات طولية

ريم طلال كامل
الباحثة

أ.م.د عماد حازم عبودي
جامعة بغداد/ كلية الإدارة والاقتصاد

P: ISSN : 1813-6729
E : ISSN : 2707-1359

http://doi.org/10.31272/JAE.44.2021.127.17

تأريخ أستلام البحث : 2020/12/23

مقبول للنشر بتأريخ : 2021/1/18

المستخلص

يمكن أن تعرف البيانات الطولية على إنها بيانات تدمج ما بين بيانات السلاسل الزمنية والبيانات المقطعية، بمعنى إنها تدرس البيانات المقطعية وتحركاتها خلال فترة زمنية معينة، لذلك يكون هذا النوع من البيانات ذو محتوى معلوماتي كبير للظاهرة المدروسة وبالتالي يمكننا ذلك من الحصول على تقديرات ذات دقة أعلى. ومن هنا تبرز أهمية البحث في المفاضلة بين أفضلية نماذج البيانات الطولية ، فضلا عن أفضلية طرائق تقدير النماذج الشبه المعلمية والتي يمكن من خلالها وصف طبيعة العلاقة بين متغيرات الانموذج الخطي الجزئي. يهدف البحث الى تقدير إنموذج شبه معلمي يصف وبشكل دقيق طبيعة العلاقة بين قيمة الإنتاج الصناعي لعشرة من المنشآت الصناعية الكبيرة في القطاع العام العراق وللفترة الزمنية (2010-2018) وبين كلا من قيمة مستلزمات واعداد المشتغلين في تلك المنشآت، ولتحقيق ذلك الهدف تم توظيف طريقتين هما طريقة سبيكمان (Speckman) وطريقة روبنسون (Robinson)، وتضمنت الطريقتين توظيف الانحدار الموضعي المتعدد بدرجات مختلفة ودالتي اللب (Gaussian) و (Epanchnikov). فضلا عن إجراء مقارنة بين الانموذجين من خلال استعمال ثلاثة معايير للمقارنة هي متوسط مربعات الخطأ (MSE)، معيار وسيط الانحرافات المطلقة (MDAE)، ومتوسط الانحرافات المطلقة (MAE). وقد أظهرت النتائج التي تم الحصول عليها بأن طريقة التقدير (Speckman) أفضل من طريقة (Robinson) .

الكلمات الرئيسية: البيانات الطولية، أنموذج شبه معلمي، الدالة اللبية، عرض الحزمة، التأثيرات العشوائية.

- البحث مستل من رسالة ماجستير.

مجلة الادارة والاقتصاد
العدد 127 / اذار/ 2021
الصفحات : 249-261

المقدمة: (Introduction)

هناك عدة مصطلحات أو تسميات تصف البيانات الطولية (Longitudinal Data) فقد تسمى أحياناً بالبيانات المدمجة (Panel Data)، أو بيانات السلاسل الزمنية والمقاطع العرضية (Time Series-Cross Sections) (TSCS). أن البيانات الطولية تجمع بين البعد المكاني أو البعد المقطعي والبعد الزماني، فإن هناك صعوبات ناجمة عن الدمج بين هذين البعدين تتمثل في وصف وتوضيح معاملات البعد المقطعي، أي تكون مكونات هذه المعاملات ذات تأثيرات ثابتة أو عشوائية، على هذا الأساس تم تقسيم هذه النماذج الى ثلاثة انواع لتسمح بتحليل هذا النوع من البيانات الى ثلاثة اشكال كالتالي: [5] [4]

1. الإنموذج التجميعي (Pooled Regression Model) (PRM).
2. إنموذج التأثيرات الثابتة (Fixed Effected Model) (FEM) .
3. إنموذج التأثيرات العشوائية (Random Effected Model) (REM) .

في هذه البحث سيتم التعامل مع إنموذج البيانات الطولية ذو التأثيرات العشوائية. من مميزات وإيجابيات تحليل البيانات الطولية مقارنةً بتحليل البيانات المقطعية بمفردها او البيانات الزمنية بمفردها ،التحكم في التباين الذي يؤدي الى نتائج متحيزة عند ظهورها في حالة البيانات المقطعية او في حالة بيانات السلسلة الزمنية كما و يعتبر هذا النوع من البيانات مناسب لدراسة فترات حالات اقتصادية كثيرة مثل البطالة والفقر ولها امكانية افضل لدراسة ديناميكية التعديل والتي يكون التغير فيها بمقاطع السلسلة الزمنية التي يمكن ان تخفيها البيانات المقطعية. كما وان هذا النوع من البيانات تكون ذات محتوى معلوماتي كبير يخص الظاهرة المدروسة وبالتالي يمكننا الحصول على تقديرات ذات دقة أعلى .كما تتميز البيانات الطولية عن غيرها بعدد درجات حرية أكبر وكذلك بكفاءة أفضل وتشارك في الحد من ظهور مشكلة المتغيرات المهملة التي نتجت من خصائص للمفردات غير المشاهدة والتي بدورها تؤدي الى مقدرات متحيزة في الانحدارات المنفردة.[7]

2) أساليب إختيار الإنموذج الملائم للبيانات الطولية:

كما ذكرنا سابقاً بوجود ثلاثة أنواع رئيسة من نماذج البيانات الطولية وهي الإنموذج التجميعي ، إنموذج التاثيرات الثابتة، وإنموذج التاثيرات العشوائية ،لذا يجب علينا إختيار الإنموذج الأكثر ملائمة للبيانات المدروسة من بين هذه النماذج بالاعتماد على الإختبارات الخاصة التي تم وضعها لهذا الغرض، إذ تمت عملية اختيار الإنموذج الأفضل بمرحلتين الأولى بأستعمال إختبار مضاعف لاكرانج (Lagrange multiplier Test) اذ تكون المقارنة بين إنموذج الإنحدار التجميعي وإنموذجي التاثيرات الثابتة والعشوائية ، والمرحلة الثانية عن طريق إختبار هوسمان (Hausman Test) الذي يتضمن المفاضلة بين إنموذجي التاثيرات العشوائية والتأثيرات الثابتة إذا لم يكن الإنموذج التجميعي ملائماً للبيانات الدراسة . [1]

(1-2) إختبار مضاعف لاكرانج: (Lagrange multiplier Test)

يعتمد هذا الاختبار على إحصاءة تتبع توزيع χ^2 بدرجة حرية واحدة وذلك للمفاضلة بين الإنموذج التجميعي و الإنموذجين العشوائي والثابت، عرف هذا الإختبار بمضاعف لاكرانج (Lagrange multiplier Test)(LM)، و لإختبار الفرضية التالية: [6]

$H_0: \sigma_u^2 = 0$ إنموذج الانحدار التجميعي هو الانموذج الملائم

$H_1: \sigma_u^2 \neq 0$ إنموذج الانحدار التجميعي غير ملائم

وإن إحصاءة الإختبار لهذه الفرضيات هي:

$$LM = \frac{nt}{2(t-1)}\left[\frac{\sum_{i=1}^{n}\left(\sum_{j=1}^{t} e_{ij}\right)^2}{\left(\sum_{i=1}^{n}\sum_{j=1}^{t} e_{ij}\right)^2} - 1\right] \quad \dots(1)$$

or

$$LM = \frac{nt}{2(t-1)}\left[\frac{t^2\,\bar{e}\,\bar{e}'}{ee'} - 1\right]^2 \sim \chi^2_{(1)} \quad \dots(2)$$

إذ إن:

n: تمثل عدد المشاهدات في الوحدات المقطعية.

t : يمثل الفترة الزمنية في الوحدات المقطعية .

e: متجه بواقي الإنموذج التجميعي للبيانات الطولية.

$\bar{e}$: متجه متوسطات المجاميع لبواقي الإنموذج التجميعي للبيانات الطولية.

إذا كانت قيمة LM المحسوبة أكبر من قيمة $\chi^2_{(1)}$ الجدولية أو إذا كانت قيمة (P-value) أقل من مستوى المعنوية، عندها ترفض فرضية العدم وهذا يعني إن الإنموذج التجميعي غير ملائم للبيانات المدروسة وإن إنموذجي التأثيرات الثابتة والعشوائية هما الأكثر ملائمة. كما إقترح (Honda 1985) إختبار آخر لمضاعف لاكرانج وذلك لتجاوز الشرط الذي يجب توفره وفقاً لإسلوب (Breusch-Pagan) في الفرضية البديلة وهو إن التباين لا يساوي صفر أي شرط عدم إنعدام التباين، فقد قدم (Honda) إختبارا أحادي الجانب وسهل الاستخدام يتضمن مقارنة الجذر التربيعي لإحصاءة (Breusch-Pagan) بقيمة التوزيع الطبيعي القياسي، وبالتالي فإن إحصاءة إختبار مضاعف لاكرانج بإسلوب (Honda) تكون بالشكل التالي:[1] [6]

$$LMH = \sqrt{\frac{nt}{2(t-1)}\left[\frac{\sum_{i=1}^{n}\left(\sum_{j=1}^{t} e_{it}\right)^2}{\left(\sum_{i=1}^{n}\sum_{j=1}^{t} e_{it}\right)^2} - 1\right]^2} \sim N(0,1) \qquad \dots(3)$$

وذلك لإختبار الفرضية التالية:

$H_0: \sigma_u^2 = 0$ إنموذج الانحدار التجميعي هو الانموذج الملائم

$H_1: \sigma_u^2 > 0$ إنموذج الانحدار التجميعي غير ملائم

إن رفض فرضية العدم هنا يعني أن التأثيرات العشوائية بالنسبة للمجاميع معنوية أي إن إنموذج الجمع بين البيانات المقطعية والسلاسل الزمنية معنوي، بمعنى آخر فإن الإنموذج التجميعي غير ملائم للبيانات المدروسة وإن إنموذجي التأثيرات الثابتة والعشوائية هما الإنموذجين الأكثر ملائمة للبيانات المدروسة.

(2-2) إختبار هوسمان : (Hausman Test)

تتم المفاضلة بين إنموذجي التأثيرات الثابتة والعشوائية لتحديد الإنموذج النهائي الملائم لبيانات الدراسة من خلال إستعمال إختبار هوسمان (Hausman test)، ويستعمل لقياس التأثيرات الثابتة والعشوائية إذ تفترض فرضية العدم بأن مقدرات التأثيرات العشوائية تكون هي الأفضل، وذلك من خلال إستعمال إحصاءة الإختبار (H) التي لها توزيع (χ^2) بدرجة حرية (k) والتي تعرف وفق الصيغة التالية: [1]

$$H = (\hat{\beta}_{FEM} - \hat{\beta}_{REM})'\left[Var(\hat{\beta}_{FEM}) - Var(\hat{\beta}_{REM})\right]^{-1}(\hat{\beta}_{FEM} - \hat{\beta}_{REM}) \qquad \dots(4)$$

إذ تمثل:

$Var(\hat{\beta}_{FEM})$: التباين والتباين المشترك للمعلمات المقدرة لإنموذج التأثيرات الثابتة.

$Var(\hat{\beta}_{REM})$: التباين والتباين المشترك للمعلمات المقدرة لإنموذج التأثيرات العشوائية.

وإن فرضيات هذا الاختبار هي:

H_0 : إنموذج التأثيرات العشوائية هو الانموذج الملائم

H_1 : إنموذج التأثيرات العشوائية إنموذج غير ملائم

إذا كانت قيمة إحصاءة هوسمان (H) المحسوبة أكبر من قيمة $\chi^2_{(K)}$ الجدولية يكون إنموذج التأثيرات الثابتة هو الإنموذج الملائم ، وعليه يمكن رفض فرضة العدم القائلة بأن التأثيرات العشوائية متسقة، أي إن إنموذج التأثيرات العشوائية هو إنموذج غير ملائم للبيانات المدروسة، أي ان إنموذج التأثيرات الثابتة هو الإنموذج الأكثر ملائمة لتلك البيانات.

3) إنموذج البيانات الطولية شبه المعلمي ذو التأثيرات العشوائية:

(Semi-Parametric Longitudinal Data Model)

يعد الإنموذج الخطي الجزئي (Partial Linear Model) (PLM) من أشهر النماذج شبه المعلمية ويطلق عليه احياناً بالإنموذج شبه المعلمي البسيط ، ويتألف الإنموذج الخطي الجزئي من جزء خطي يمثله الإنحدار المعلمي وجزء لا خطي يمثله الإنحدار اللامعلمي، و يمكن وصف إنموذج البيانات الطولية الخطي الجزئي ذو التأثيرات العشوائية رياضياً وفق الصيغة التالية: [12][9]

$$y_{ij} = x_{ij}'\beta + m(z_{ij}) + \varphi_{ij} \quad , i = 1, \dots\dots, n \ , \ j = 1, \dots\dots, t \qquad \dots(5)$$
$$\varphi_{ij} = u_i + v_{ij}$$

إذ أن:

y_{ij} : متجه بالأبعاد $(nt \times 1)$ ويمثل المتغير المعتمد في الوحدة المقطعية i عند الفترة الزمنية j.

(x_{ij}) و (z_{ij}) هي متجهات المتغيرات التوضيحية بالأبعاد $(q \times 1)$ و $(d \times 1)$ على الترتيب.

$m(Z)$: تمثل دالة الجزء اللامعلمي التي يمكن تقديرها بطرائق التقدير اللامعلمية.

u_i : الخطأ في الوحدة المقطعية (i).
v_{ij} : متجه الخطأ العشوائية بالأبعاد (t×1).
ويخضع الإنموذج شبه المعلمي المعرف في الصيغة (5) الى الشروط والافتراضات التالية:

$$E(v_{ij}|x_{ij},z_{ij})=0 \quad ، \quad E(v_{ij}v_{ij}'|x_{ij},z_{ij},u_i)=\sigma_{ij}^2\, I_t$$
$$E(u_i|x_{ij},z_{ij},u_i)=0 \quad ، \quad E({u_i}^2|x_{ij},z_{ij},u_i)=\sigma_u^2$$
$$E(\varphi_{ij}|z_{ij})=0$$

وبذلك فإن الإنموذج وفق الإفتراضات أعلاه سيكون وفق الصيغة التالية:

$$E(y_{ij}|x_{ij},z_{ij})=x_{ij}'\beta+m(z_{ij}) \qquad \dots(6)$$

4) عرض الحزمة: (Bandwidth)

وتسمى بمعلمة التمهيد ويرمز لها عادة بالرمز (h)، تملك معلمة التمهيد تأثيراً قوياً على نتائج الدوال اللبية وبالتالي فأن أي مقدر من مقدرات التمهيد اللامعلمي اللبي سيكون حساساً جداً تجاه التغيرات في قيمة (h) والذي سيؤدي بدوره الى تغيير بشكل منحنى التمهيد. وأساليب تقدير أو إختيار عرض الحزمة متعددة لكننا أخترنا في هذا البحث طريقة العبور الشرعي (Cross Validation Method) وإن من أكثر أساليب التمهيد أستعمالاً هي الدوال اللبية التي تمتاز بمرونتها في عملية التقدير كما وتتصف بكونها دوال حقيقية مستمرة ومتماثلة حول الصفر وسيتم استعمال دالـتي اللـب (Gaussian وEpanchnikov) وسيتم توظيف طريقة المربعات الصغرى للعبور الشرعي (LSCV) (Least Square Cross Validation) التي تعمل على جعل متوسط مربعات الخطأ التكاملية Mean Integrated Squared Error(MISE) اقل ما يمكن . [3] [5]

5) طرائق التقدير (Estimation Methods)

Speckman Method (1-5)

تتلخص طريقة Speckman بتقدير الجزء اللامعلمي وذلك بأخذ التوقع الشرطي للإنموذج المعرف في الصيغة (5) بالنسبة الى (z_{ij}) وكما يلي: [8] [9] [11]

$$E(y_{ij}|z_{ij})=E(x_{ij}|z_{ij})'\beta+E[m(z_{ij}|z_{ij})]+E[m(\varphi_{ij}|z_{ij})] \qquad \dots(7)$$

وتحت شروط الإنموذج الخطي الجزئي فإن $E[m(\varphi_{ij}|z_{ij})]=0$، وإن $E[m(z_{ij}|z_{ij})]=m(z_{ij})$ ، لذلك بطرح الصيغة (7) من الصيغة (5) نحصل على:

$$y_{ij}-E(y_{ij}|z_{ij})=[x_{ij}-E(x_{ij}|z_{ij})]'\,\beta+\varphi_{ij} \qquad \dots(8)$$

وهنا قد تم إستبعاد الجزء اللامعلمي من الإنموذج الخطي الجزئي، ويكتب هذا الإنموذج المختزل بصيغة المصفوفات والمتجهات بالشكل التالي:

$$\check{Y}=\check{X}'\beta+\varphi_{ij} \qquad \dots(9)$$

وبتطبيق طريقة المربعات الصغرى (OLS) على الصيغة (9) يمكن أن نحصل على تقديرات المعلمات (β) وكما يلي:

$$\check{\beta}=(\check{X}'\check{X})^{-1}(\check{X}'\check{Y}) \qquad \dots(10)$$

اذ ان:
$\check{X}$: مصفوفة بالأبعاد $(nt\times q)$ صفوفها تمثل $\check{X}_{ij}$.
$\check{Y}$: متجه صفي بالأبعاد $(nt\times 1)$ تشمل صفوفه قيم $\check{Y}_{ij}$.
ويمكن تقدير الحدين غير المعلومين $E(y_{ij}|z_{ij})$ و $E(x_{ij}|z_{ij})$ في الصيغة (8) بإستعمال إحدى الطرائق اللامعلمية ولتكن طريقة نادرايا واتسون التي تمثل حالة خاصة من المقدر الخطي الموضعي متعدد الحدود وذلك عندما (p=0) بصيغته التالية:

$$\hat{m}(z_{ij})_{NW}=\frac{\sum_{i=1}^{n}\sum_{j=1}^{t}y_{ij}\,K\left(\frac{z_{ij}-z}{h}\right)}{\sum_{i=1}^{n}\sum_{j=1}^{t}\;K\left(\frac{z_{ij}-z}{h}\right)} \qquad \dots(11)$$

فإذا رمزنا لهذين التقديرين $\hat{X}, \hat{Y}$ على الترتيب، عندها فإن:

$$\hat{X} = \hat{E}(X_{ij}|Z_{ij}) = \frac{1}{nth^p} \frac{\sum_{r=1}^{n}\sum_{s=1}^{t} X_{rs}\, K_h(Z_{ij} - Z_{rs})}{\sum_{r=1}^{n}\sum_{s=1}^{t} K_h(Z_{ij} - Z_{rs})} \quad \dots(12)$$

$$\hat{Y} = \hat{E}(Y_{ij}|Z_{ij}) = \frac{1}{nth^p} \frac{\sum_{r=1}^{n}\sum_{s=1}^{t} Y_{rs}\, K_h(Z_{ij} - Z_{rs})}{\sum_{r=1}^{n}\sum_{s=1}^{t} K_h(Z_{ij} - Z_{rs})} \quad \dots(13)$$

إذ تمثل :

p : درجة متعدد الحدود(polynomial).

وبتعويض الصيغتين (12) و (13) في الصيغة (10) نحصل على تقدير الجزء المعلمي من الإنموذج الخطي الجزئي للبيانات الطولية:

$$\hat{\beta}_S = (\hat{X}'\hat{X})^{-1}\ (\hat{X}'\hat{Y}) \quad \dots(14)$$

وان تقدير الجزء اللامعلمي للإنموذج الخطي الجزئي يتم من خلال تقليل الاخطاء العشوائية للتقدير الموضعي الخطي المتعدد الحدود من الدرجة الاولى لنحصل على الصيغة التالية:

$$\hat{m}_s(Z) = I_{nt}\ (R'\ K_z\ R)^{-1}\ R'\ K_z\ (Y - X\hat{\beta}) \quad \dots(15)$$

R : مصفوفة بالأبعاد $(nt) \times (1+p)$،إذ أن:

$$R = \begin{bmatrix} 1 & (Z_{11} - z)' \\ \vdots & \\ 1 & (Z_{nt} - z)' \end{bmatrix}$$

Profile Least Square Method(Robinson) (2-5)

وتعتمد هذه الطريقة على تحويل الإنموذج شبه المعلمي الى إنموذج لامعلمي وذلك بطرح الجزء المعلمي من الإنموذج الشبه المعلمي، ويمكن تطبيق هذه طريقة على إنموذج البيانات الطولية المعرف في الصيغة (5) كما هو موضح في الخطوات التالية: [9] [2]

$$y_{ij} - x_{ij}'\beta = m(z_{ij}) + \varphi_{ij} \qquad ,i = 1, \dots\dots, n \ \ , \ \ j = 1, \dots\dots, t$$

$$y_{ij}^* = y_{ij} - x_{ij}'\beta$$

$$y_{ij}^* = m(z_{ij}) + \varphi_{ij} \quad \dots(16)$$

إذ يمكن تقدير الدالة $\boldsymbol{m}(.)$ أولاً باستعمال التقدير الخطي الموضعي متعدد الحدود الذي تم التطرق الية في المصادر السابقة، وذلك بإفتراض أن $\boldsymbol{\theta_1}, \boldsymbol{\theta_0}$ تمثل الحلول الناتجة من تصغيرمجموع مربعات الخطأ للإنموذج العشوائي، عندها فإن:

$$\sum_{i=1}^{n}\sum_{j=1}^{t}[(y_{ij} - x_{ij}'\boldsymbol{\beta}) - \boldsymbol{\theta_0} - \boldsymbol{\theta_1}(z_{ij} - z_i)]^2\, K_h(z_{ij} - z_i) \quad \dots(17)$$

عندها فإن $\widehat{\boldsymbol{\theta}}' = (\widehat{\boldsymbol{\theta}}_1\ \ \widehat{\boldsymbol{\theta}}_0)$ تمثل تقديرات لكل من $\boldsymbol{m_0}(z_{ij})$ و $\boldsymbol{m_1}(z_{ij})$ على الترتيب، أي ان:[9]

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{I_{nt}}\ (\boldsymbol{H}'\ \boldsymbol{K_z}\ \boldsymbol{H})^{-1}\ \boldsymbol{H}'\ \boldsymbol{K_z}\ (\boldsymbol{Y} - \boldsymbol{Z\beta}) = \boldsymbol{S}(\boldsymbol{Y} - \boldsymbol{Z\beta}) \quad \dots(18)$$

إذ أن:

$\boldsymbol{I_{nt}}$: $\mathbf{q}$ من المتجهات التي عناصرها 0 وان اول $\mathbf{p}$ من عناصرها تساوي1.

$\boldsymbol{S}$: مصفوفة التمهيد $\boldsymbol{S} = \boldsymbol{I_{nt}}\ (\boldsymbol{H}'\ \boldsymbol{K_z}\ \boldsymbol{H})^{-1}\ \boldsymbol{H}'\ \boldsymbol{K_z}$.

$\boldsymbol{K}$: مصفوفة الأوزان لدالة اللب القطرية بالأبعاد $(\boldsymbol{nt} \times \boldsymbol{nt})$،

$$\boldsymbol{diag(K)} = \left[\boldsymbol{K_h}(\boldsymbol{Z_{11}} - \boldsymbol{Z})', \dots, \boldsymbol{K_h}(\boldsymbol{Z_{nm_i}} - \boldsymbol{Z})'\right]$$

$\boldsymbol{Y}$: متجه المتغير المعتمد بالأبعاد $(\boldsymbol{nt} \times \mathbf{1})$، $\boldsymbol{Y} = [\boldsymbol{Y_{11}}, \dots, \boldsymbol{Y_{nt}}]$.

$\boldsymbol{X}$: مصفوفة المتغيرات التوضيحية بالأبعاد $(\boldsymbol{nt} \times \boldsymbol{q})$، $\boldsymbol{X} = [\boldsymbol{X_{11}}, \dots, \boldsymbol{X_{nt}}]$.

$\boldsymbol{H}$: مصفوفة بالأبعاد $(\boldsymbol{nt}) \times (\mathbf{1} + \boldsymbol{p})$،إذ أن:

$$H = \begin{bmatrix} 1 & (Z_{11} - z)' \\ \vdots & \\ 1 & (Z_{nt} - z)' \end{bmatrix}$$

وبالتعويض عن $m(z_{ij})$ في الصيغة (16) عندها فإن إنموذج الإنحدار للبيانات الطولية بصيغة المصفوفات والمتجهات يصبح على الشكل التالي:[10]

$$\hat{Y} = \hat{X}'\beta + \varphi^* \quad , i = 1, \dots\dots, n \ , \ j = 1, \dots\dots, t \qquad \dots(19)$$

إذ أن:

$$\hat{Y} = (I_{nt} - S)\, Y \quad , \hat{Y} = (\hat{Y}_{11}, \dots, \hat{Y}_{nt})$$
$$\hat{X} = (I_{nt} - S)\, X \quad , \hat{X} = (\hat{X}_{11}, \dots, \hat{X}_{nt})$$
$$\varphi^* = (I_{nt} - S)\, \varphi_{ij} + (I_{nt} - S)\, m(Z) \quad ، \quad \varphi^* = ({\varphi_{11}}^*, \dots, {\varphi_{nt}}^*)$$
$$m(Z) = [m(Z_{11}), \dots, m(Z_{nt})]$$

وبتطبيق طريقة المربعات الصغرى يمكن ان نحصل على التقدير شبه المعلمي لمتجه المعلمات (β) ودالة التمهيد $m(Z)$ على الترتيب كما يلي: [9] [12]

$$\hat{\beta}_{PLLS} = (\hat{X}'\hat{X})(\hat{X}'\hat{Y}) \qquad \dots(20)$$

$$\hat{m}_{PLLS}(Z) = I_{nt}\,(H'K\,H)^{-1}\,H'\,K\,(Y - X\hat{\beta}_{PLLLS}) \qquad \dots(21)$$

6) وصف البيانات: (Data description)

تمثل البيانات المدروسة المبينة في الشكل (أ) والجدول رقم (2) قيمة الانتاج كمتغير معتمد والمتغيرات التوضيحية الرئيسة المؤثرة عليه وهي قيمة مستلزمات الانتاج واجور المشتغلين، وهي بيانات طولية تتضمن عشرة من المنشآت الصناعية الكبيرة في العراق (n=10) والتي تمثل المشاهدات المقطعية مقاسة لمدة تسع سنوات (t=9) للفترة (2010-2018) والتي تمثل بدورها السلسلة الزمنية، وقد تم الحصول على البيانات من الجهاز المركزي للإحصاء. ويتم تشخيص نوع إنموذج البيانات الطولية من خلال إختبار مضاعف لاكرانج وفق اسلوب هوندا لتحديد مطابقة البيانات للإنموذج التجميعي، ونتائج الإختبار مبينة في الجدول رقم (1) أدناه:

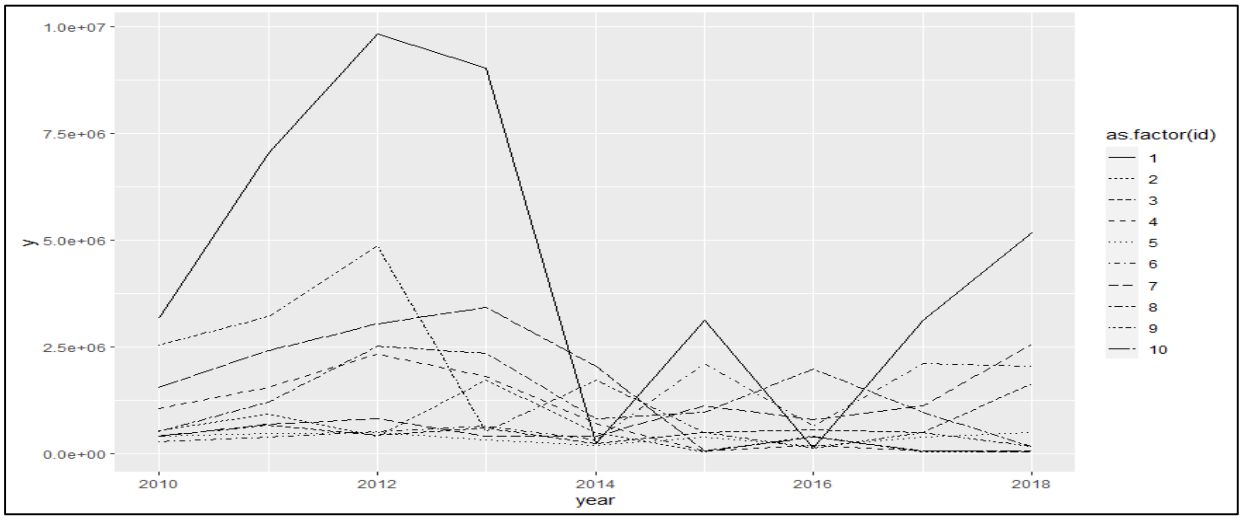
جدول رقم (1)

إختبار مضاعف لاكرانج وإختبار هوسمان

الإختبار	قيمة إحصاءة إختبار مربع كاي	مستوى الدلالة
إختبار مضاعاف لاكرانج	6.7369	$8.1\ e^{-12}$
إختبار هوسمان	0.28401	0.8676

المصدر :من إعداد الباحثين بالإعتماد على نتائج التقدير ولغة R.

يتضح من الجدول (1) أعلاه إن مستوى الدلالة لإختبار مضاعف لاكرانج بلغت ($8.1\ e^{-12}$)وهو اقل من مستوى المعنوية (5%) ، وهذا يعني رفض فرضية العدم أي إن الإنموذج التجميعي غير ملائم لبيانات الدراسة، بينما كانت قيمة مستوى الدلالة لإحصاءة اختبار هاوسمان (0.8676) وهي أكبر من مستوى المعنوية وهذا يعني قبول فرضية العدم أي إن إنموذج التأثيرات العشوائية أكثر ملائمة من إنموذج التأثيرات الثابتة.



الشكل (أ)

بيانات قيمة الانتاج للمنشآت الصناعية الكبيرة للقطاع العام في العراق للفترة (2010-2018)

جدول رقم (2)

بيانات قيمة الانتاج الصناعي للمنشآت الصناعية الكبيرة للقطاع العام في العراق للفترة (2010-2018)

اعداد المشتغلين	قيمة مستلزمات الانتاج (مليون دينار)	قيمة الانتاج (مليون دينار)	السنة	رمز المنشاة	اعداد المشتغلين	قيمة مستلزمات الانتاج (مليون دينار)	قيمة الانتاج (مليون دينار)	السنة	رمز المنشاة
30000	241750	280800	2010	6	23000	1809583	3158855	2010	1
30000	238800	373750	2011	6	22000	3009550	7052375	2011	1
31000	300992	510450	2012	6	23000	4696864	9847855	2012	1
31000	332102	640200	2013	6	23000	3851234	9046750	2013	1
30000	581208	305000	2014	6	26000	9119489	230860	2014	1
26000	1016660	2102200	2015	6	19000	1551800	3111477	2015	1
22000	278866	635580	2016	6	17000	2163615	117760	2016	1
26000	1016660	2102200	2017	6	19000	1551800	3111477	2017	1
25000	1007350	2043900	2018	6	20000	2757164	5176621	2018	1
42000	215724	406000	2010	7	18000	269147	533750	2010	2
42000	250465	689780	2011	7	19000	252578	916250	2011	2
44000	403547	811965	2012	7	25000	67766	402794	2012	2
34000	199345	399345	2013	7	24000	551765	1709750	2013	2
40000	5689484	408375	2014	7	19000	549200	490000	2014	2
18000	308452	1120810	2015	7	13000	4500	42000	2015	2
18000	416097	783360	2016	7	18000	137046	402000	2016	2
18000	308452	1120810	2017	7	13000	4500	42000	2017	2
22000	542650	2569125	2018	7	13000	3696	39240	2018	2
15000	240682	499200	2010	8	71000	165488	400985	2010	3
21000	644302	1196800	2011	8	71000	372371	657300	2011	3
42000	1280890	2512000	2012	8	18000	241398	421540	2012	3
17000	1062256	2342400	2013	8	18000	345312	602700	2013	3
19000	563003	815520	2014	8	71000	2662044	228930	2014	3
41000	514925	967675	2015	8	27000	226930	483912	2015	3
43000	1333060	1975506	2016	8	29000	259477	545175	2016	3
41000	514925	967675	2017	8	27000	226930	483912	2017	3
34000	56552	174640	2018	8	23000	767028	1639750	2018	3
54000	1027738	2533100	2010	9	34000	416015	1053885	2010	4
54000	234000	3204800	2011	9	30000	713883	1551060	2011	4
54000	1642535	4878235	2012	9	28000	1078726	2326450	2012	4
71000	379530	506004	2013	9	29000	867662	1808797	2013	4
54000	2056880	1713680	2014	9	42000	2790070	701150	2014	4

100000	184429	478795	2015	9	14000	27876	50604	2015	4
100000	79577	130219	2016	9	18000	104274	191035	2016	4
100000	184429	478795	2017	9	14000	27876	50604	2017	4
87000	216007	164768	2018	9	18000	24721	40379	2018	4
108000	1295912	1546747	2010	10	25000	8027578	406485	2010	5
157000	1749453	2399058	2011	10	25000	69648	459285	2011	5
168000	1621108	3027240	2012	10	22000	67271	488025	2012	5
191000	1381292	3424286	2013	10	39000	70563	324224	2013	5
89000	600598	2049732	2014	10	25000	314984	189883	2014	5
44000	42667	46448	2015	10	32000	58673	388600	2015	5
82000	358296	373349	2016	10	32000	36618	174388	2016	5
44000	42667	46448	2017	10	32000	58673	388600	2017	5
46000	25230	50647	2018	10	32000	77995	495450	2018	5

ونظرا لعدم توافق وحدات القياس فقد تم تحويل البيانات الى الصيغة القياسية (standardization)

$Z = \frac{X-\mu}{\sigma}$.

ولتجنب الخلط بين تفاصيل طرائق التقدير للانموذج عند تفسير النتائج فقد تم اعطاؤها التسميات الموضحة في الجدول رقم (3)

جدول رقم (3)

وصف النماذج المقدرة حسب طرائق التقدير

التسمية	تفاصيل طريقة التقدير	التسمية	تفاصيل طريقة التقدير
Semi Parametric Model I	Speckman ، NW، Gaussian Kernel	Semi Parametric Model V	Robinson ، NW، Gaussian Kernel
Semi Parametric Model II	Speckman ، NW Epanechnikov Kernel	Semi Parametric Model VI	Robinson ، NW، Epanechnikov Kernel
Semi Parametric Model III	Speckman ، Local Gaussian Kernel	Semi Parametric Model VII	Robinson ، Local Gaussian Kernel
Semi Parametric Model IV	Speckman ، Local ، Epanechnikov Kernel	Semi Parametric Model VIII	Robinson ، Local Epanechnikov Kernel

تم تقدير ثلاثة معايير للمقارنة بين طرائق التقدير وهي (MSE)، (MAE)، (MDAE) ووضع النتائج المستخرجة في الجدول (4) ادناه.

جدول رقم (4)

قيم معايير المقارنة (MSE, MAE,MDAE) بين النماذج المقدرة

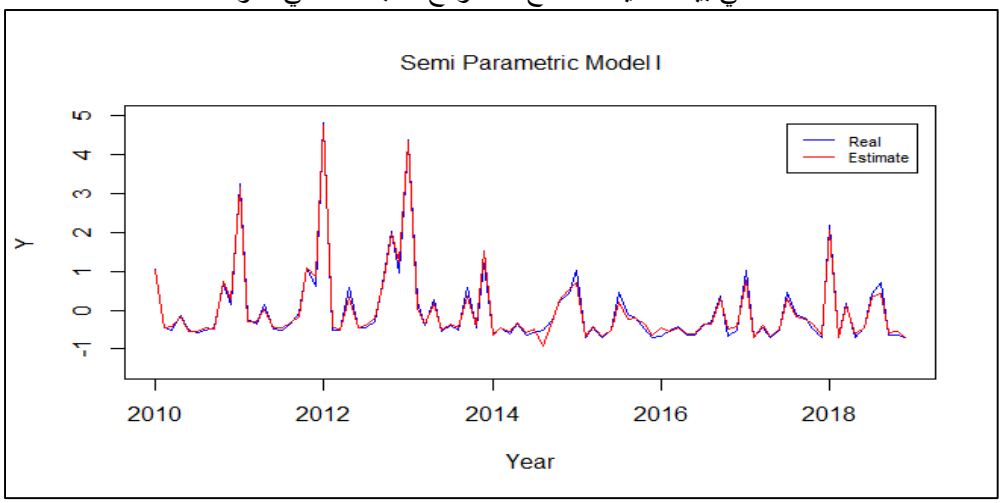
Model	MSE	MAE	MDAE	Best
Semi Parametric Model III	**0.00881**	**0.06339**	**0.03914**	**1**
Semi Parametric Model II	0.01289	0.07367	0.04006	2
Semi Parametric Model I	0.01403	0.08485	0.06100	3
Semi Parametric Model IV	0.02177	0.09583	0.04916	4
Semi Parametric Model VI	0.06177	0.19245	0.14299	5
Semi Parametric Model V	0.07723	0.21969	0.17933	6
Semi Parametric Model VII	0.17815	0.27157	0.17922	7
Semi Parametric Model VIII	0.25713	0.34099	0.21134	8

ويتضح من خلاله فيما يخص النماذج الشبه المعلمية المقدرة بطريقة سبيكمان فإن الأنموذج الاكثر كفاءة كان إنموذج البيانات الطولية الشبه المعلمي (Semi Parametric Model III) أي الانموذج المقدر بالاعتماد على التقدير الموضعي الخطي بالدالة اللبية (Gaussian)، كونه اعطى اقل قيم لمعايير المقارنة الثلاثة، وحل بالترتيب الثاني إنموذج البيانات الطولية الشبه المعلمي (Semi Parametric Model II) وهو الانموذج المقدر بطريقة نادارايا - واتسون بدالة اللب (Epanechnikov) ، وحل ثالثاً (Semi Parametric Model I) اي الانموذج المقدر بطريقة نادارايا - واتسون ودالة اللب (Gaussian)، في حين جاء بالترتيب الرابع الانموذج (Semi Parametric Model IV) اي الانموذج المقدر بطريقة

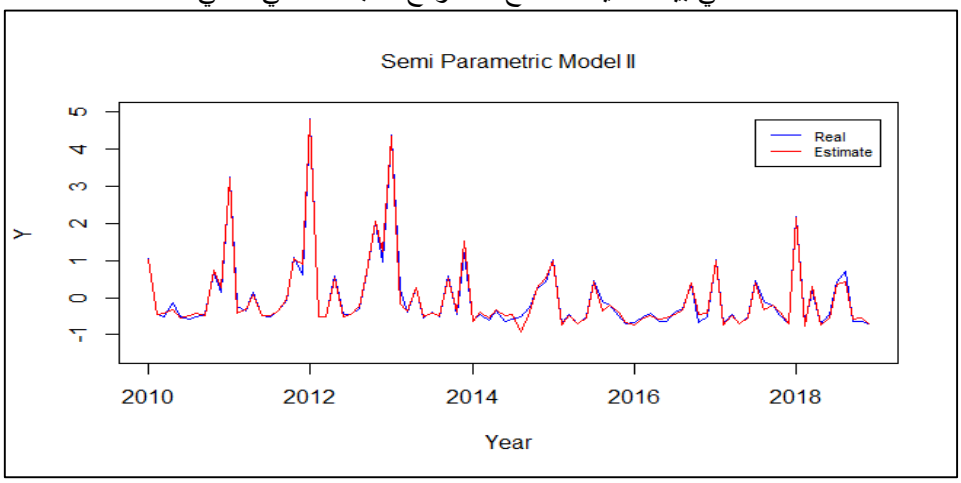
الانحدار الموضعي بالدالة اللبية (Epanechnikov). اما النماذج الشبه المعلمية المقدرة بطريقة روبنسون (Profile least square) أظهرت تفوق الانموذج المقدر بطريقة نادرايا – واتسون بالدالة اللبية (Epanechnikov) اي الانموذج (Semi Parametric Model VI)) كونه امتلك اقل قيم لـ (MSE) ، (MAE) و (MDAE) ، يليه الانموذج (Semi Parametric Model V) والمقدر بطريقة نداريا واتسن بالدالة اللبية (Gaussian) ، تبعه الانموذج الانموذج Semi Parametric (Model VII) والمقدر بطريقة الانحدار الموضعي بالدالة اللبية (Gaussian)، وحل بالترتيب الرابع الانموذج المقدر (Semi Parametric Model VIII) والمقدر بطريقة الانحدار الموضعي بالدالة اللبية (Epanechnikov).

ولتوضيح عملية المقارنة بصورة اكثر دقة تم رسم المنحنيات الحقيقية والمقدرة والخاصة بالانموذج شبه المعلمي للبيانات الطولية العشوائية لمتغير قيمة الانتاج وكما هو موضح بالاشكال من (1) الى (8):

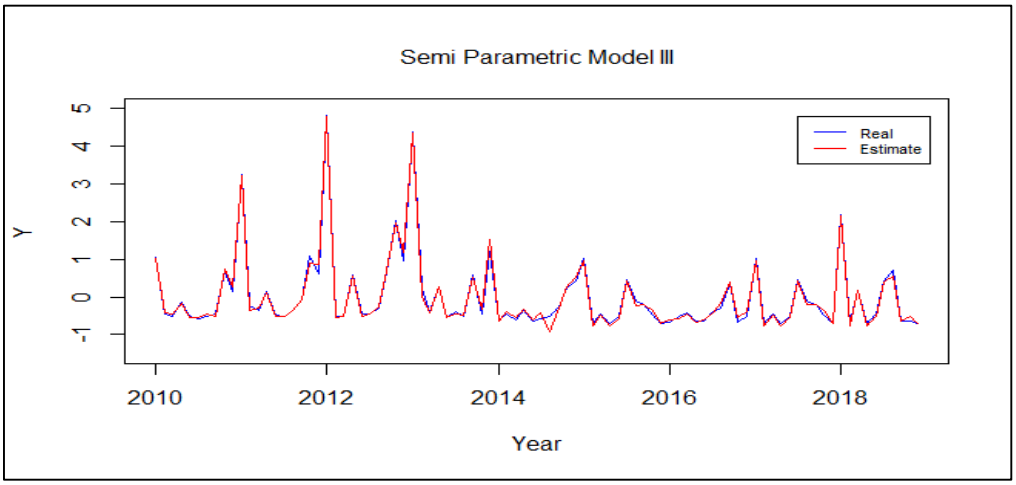
شكل رقم (1)
منحني بيانات قيمة الانتاج للانموذج الشبه المعلمي الأول



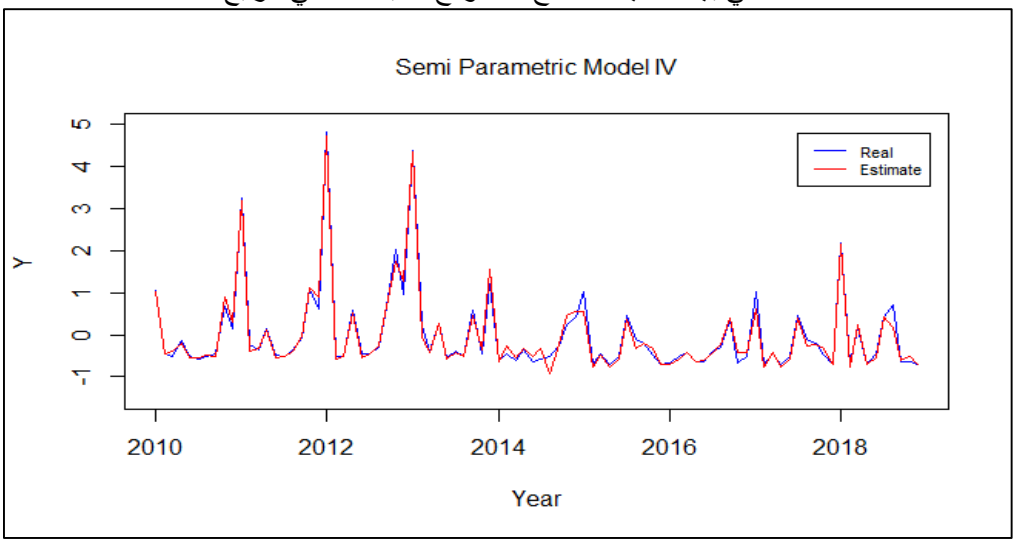
شكل رقم (2)
منحني بيانات قيمة الانتاج للانموذج الشبه المعلمي الثاني



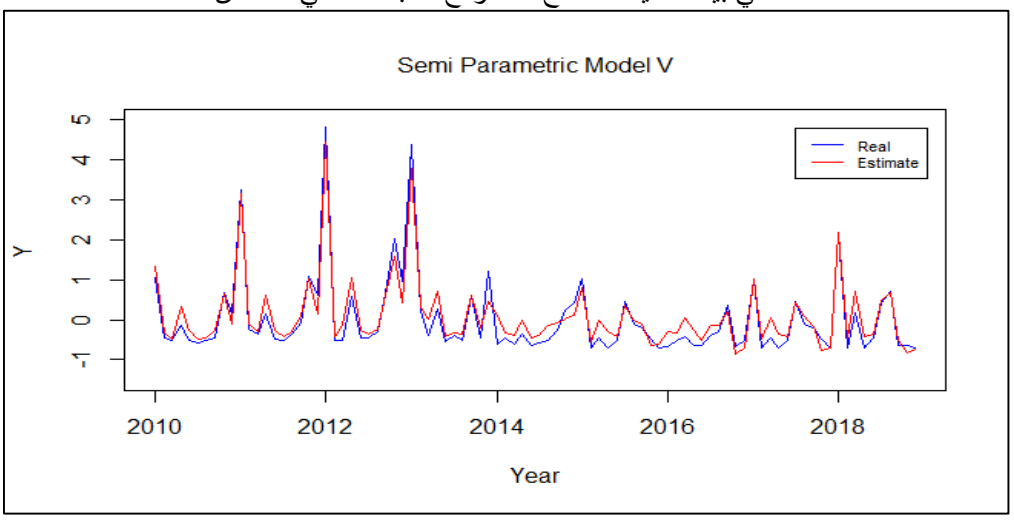
شكل رقم (3)
منحني بيانات قيمة الانتاج للانموذج الشبه المعلمي الثالث



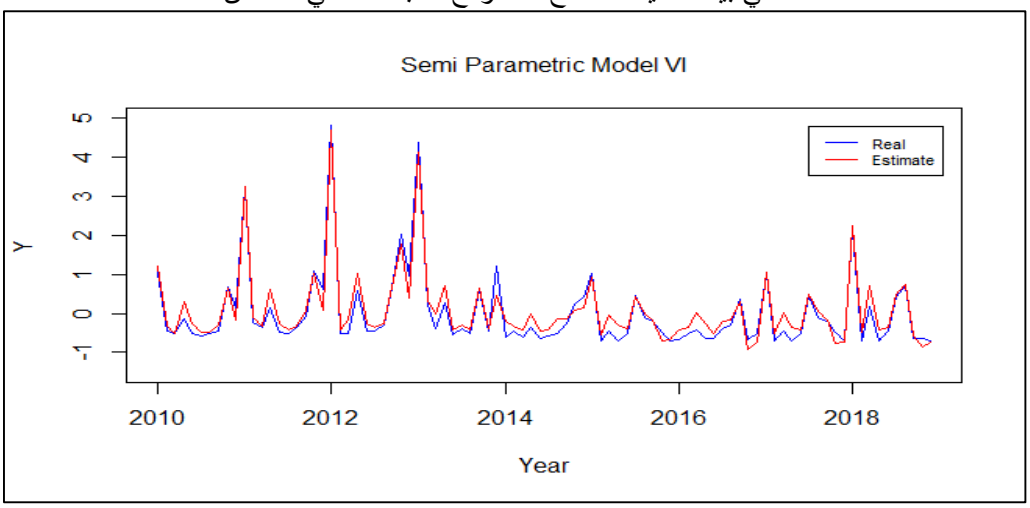
شكل رقم (4)
منحني بيانات قيمة الانتاج للانموذج الشبه المعلمي الرابع



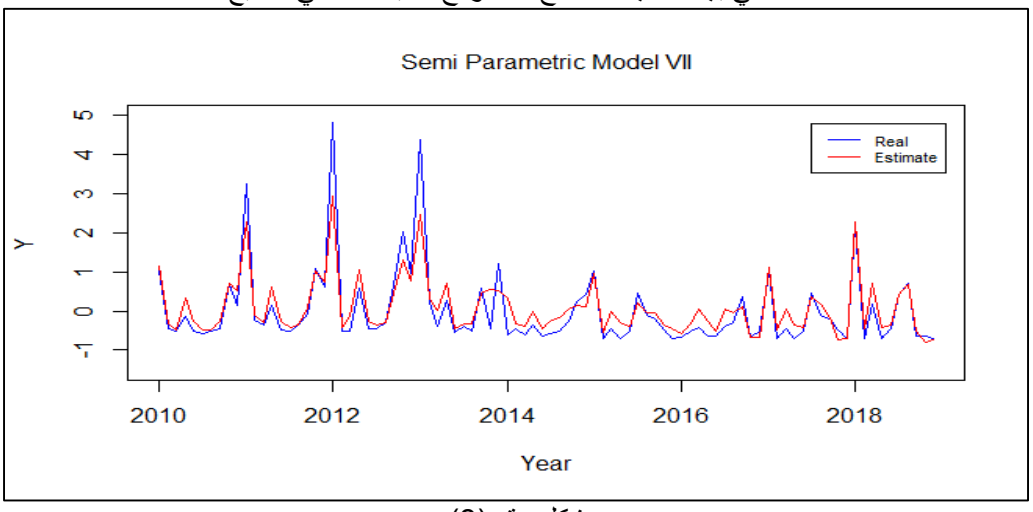
شكل رقم (5)
منحني بيانات قيمة الانتاج للانموذج الشبه المعلمي الخامس



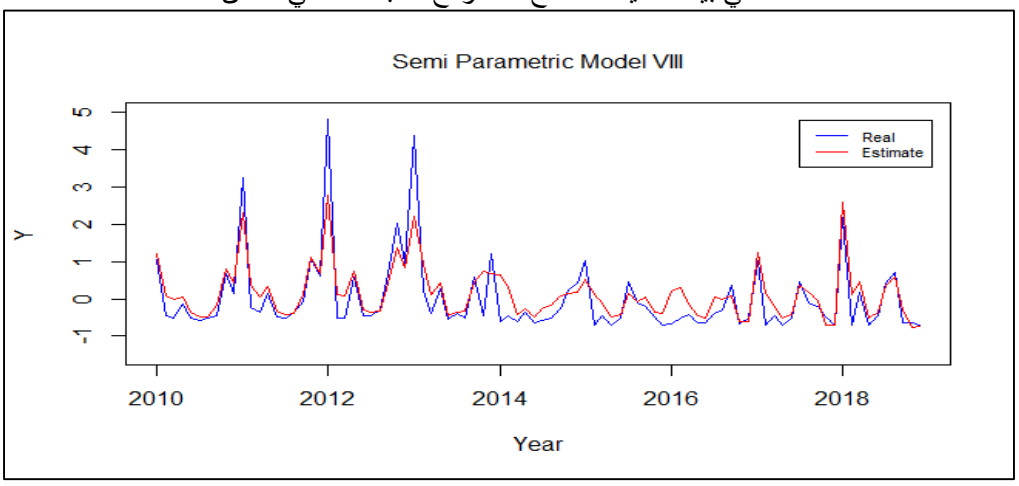
شكل رقم (6)
منحني بيانات قيمة الانتاج للانموذج الشبه المعلمي السادس



شكل رقم (7)
منحني بيانات قيمة الانتاج للانموذج الشبه المعلمي السابع



شكل رقم (8)
منحني بيانات قيمة الانتاج للانموذج الشبه المعلمي الثامن



من الواضح من اشكال النماذج المقدرة بأن التمهيد الشبه المعلمي للانموذج (Semi Parametric Model I) الموضح في الشكل (1) كان أفضل من النماذج (الرابع والخامس و السادس والسابع والثامن)، إذ نلاحظ وجود تقارب بين القيم الحقيقية والقيم التقديرية لمشاهدات المتغير المعتمد مع وجود بعض الفروقات عند السنوات 2014 ،2016 و2018، لا يختلف كثيرا التمهيد للانموذج (Semi Parametric Model II) في الشكل (2) عن سابقه المذكور في الشكل (1) سوى باقتراب التمهيد عند السنة 2014 ،التمهيد بالانموذج الشبه المعلمي (Semi Parametric Model III) في الشكل (3) يبدو مطابقاً بشكل كبير لمشاهدات المتغير المعتمد الحقيقية، ويمكن القول ان التمهيد وفق هذا الانموذج هو الافضل من بين كافة النماذج، مع ملاحظة فرق بسيط في التمهيد عند السنوات 2014 و 2018، الشكل (4) يخص التمهيد بالانموذج الشبه المعلمي (Semi Parametric Model IV)، من الواضح وجود فرق في التمهيد عند السنوات 2014، 2016 و 2018،و الاشكال من (5) الى (8) تمثل وعلى الترتيب التمهيد للنماذج الشبه المعلمية (VIII, VII,VI,V)، ويظهر من خلالها فرق أكبر بين المشاهدات المقدرة عن المشاهدات الحقيقية وخاصة عند المستويات الطبيعية لقيم المتغير المعتمد عن ما كانت عليه في النماذج الشبه المعلمية اعلاه، مع ملاحظة أفضلية الانموذج (Semi Parametric Model VI) هو الافضل من بين النماذج المذكورة في هذه الفقرة.

7) الاستنتاجات (Conclusion)

من خلال النتائج التي توصل اليها الباحث يمكن ان نستنتج ما يلي:

1. إن الانموذج الاكثر ملائمة للبيانات الطولية لوصف طبيعة العلاقة بين قيمة الإنتاج وكلا من قيمة مستلزمات الإنتاج وعدد المشتغلين لبيانات الشركات الصناعية العشرة الكبيرة المدروسة في هذا البحث وللفترة الزمنية من 2010-2018 هو الانموذج العشوائي.
2. إن أفضل انموذج شبه معلمي يمكن توظيفه في وصف بيانات قيمة الإنتاج للشركات الصناعية الكبيرة هو الانموذج المقدر بطريقة سبيكمان.
3. التقدير بإستعمال الإنحدار الموضعي المتعدد من الدرجة الاولى أكثر دقة من التقدير وفق الانحدار الموضعي بثابت عند دراسة إنموذج البيانات الطولية العشوائي لقيمة الإنتاج الصناعي للشركات العشرة.
4. دالة اللب (Gaussian) تعطي تقديرات اكثر ملائمة من دالة اللب (Epanechnikov) للبيانات المدروسة .
5. هناك تطابق تام في النتائج التي تم التوصل اليها من خلال معايير المقارنة الثلاثة (MSE)، (MAE)، MDAE) من جهة، وتطابق بين قيم تلك المعايير مع دقة اشكال منحنيات التمهيد من جهة أخرى.

المصـــادر (References)

1. Baltagi , B,H, "Econometrics Analysis of Panel Data" , 3rd Edition, John Wiely And Sons, Ltd, West Sussex,2005.
2. Fan, J. and Huang, T. (2005). "Profile likelihood inferences on semiparametric varying-coefficient partially linear models". Bernoulli, 11, 1031–1057.
3. Flashier,"NonparametricEconometrics"www.vcharite.univmrs.fr/nonparametric-chap3.pdf
4. Green W.,(2005). " Econometric Analysis", 5ed, Prentice Hall, New Jersey.
5. Hardle, W., Muller,M., Sperlich,S., H., Werwatz,A., (2004). "Nonparametric and Semi parametric Models an Introduction". Springer Edition
6. Honda, Y.,(1985)." Testing the error components model with nonnormal disturbances" , review of economic studies. 52(4),681-690.
7. Hsiao, C., (2006) . " Analysis of Panel Data ," Second Edition, Institute for the University of Southern California.
8. Lin, X. and Carroll, R. J., (2006), "Semiparametric estimation in general repeated measures problems", Journal of the Royal statistical society, 68(1), 69-88.
9. Poo, M., Soberonm A., (2016). " Nonparametric and Semiparametric Panel data Model: Recent Development". Journal of Economic Surveys, 00(0), 1–37.
10. Robinson, P.M. (1988). "Root-n-consistent semiparametric regression". Econometrica 56(4), 931–954
11. Speckman, P. (1988). "Kernel smoothing in partial linear models". Journal of the Royal Statistical Society: Series B 50: 413–436.
12. Yun, W., (2012)." Essays on Nonparametric and Semiparametric Models and Continuous Time Models".Doctor thesis in economics, University of California, USA.

Comparison Of Some Methods For Estimating a Semi-parametric Model For Longitudinal Data

Reem Tallal Kamil AL- adilee / University of Baghdad /College Of Physical Education& Sports Sciences For Girls
reem.t.kamel.88@gmail.com
Ass.Prof.Dr. Emad Hazim Aboudi / University of Baghdad/ College of Administration &
Economics / statistics Department
dremad622@yahoo.com

Abstract:

Longitudinal data can be defined as data that combines between time-series data and cross-sectional data, meaning that it studies cross-sectional data and its movements over a certain period of time, so this type of data has a large informational content for the phenomenon studied and thus enables us to obtain estimates of higher accuracy . Hence, the importance of research appears in the comparison between the preference of longitudinal data models, as well as the preference of methods for estimating semi-parametric models through which the nature of the relationship between the variables of the partial linear model can be described. The research aims to estimate a semi-parametric model that accurately describes the nature of the relationship between the value of the industrial production of ten large industrial establishments in the public sector in Iraq for the period of time (2010-2018) and both the value of the requirements and the number of workers in those establishments, and to achieve that aims, two methods were employed: Speckman method and Robinson method, and the two methods included employing multiple local regression with different degrees and the Kernel functions (Gaussian) and (Epanchnikov). As well as making a comparison between the two models through the use of three comparison criteria, which are the mean squares of error (MSE), the median absolute deviation error (MDAE), and the mean of absolute deviations (MAE). The results obtained showed that the estimation method (Speckman) is better than the method. (Robinson).

Keywords: Longitudinal Data ، Semi-Parametric Model ، Kernel Function ،Bandwidth ، Random Effected .

…….
……
……