

مقارنة بعض طرائق تقدير أنموذج شبه معلمي لبيانات طولية

ريم طلال كامل
الباحثة

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

P: ISSN : 1813-6729

<http://doi.org/10.31272/JAE.i127.109>

E : ISSN : 2707-1359

مقبول للنشر بتاريخ : 2021/1/18

تاريخ استلام البحث : 2020/12/23

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 أفضل نماذج البيانات الطولية ، فضلا عن أفضل طرائق تقدير النماذج الشبه المعلمية والتي يمكن من خلالها وصف طبيعة العلاقة بين متغيرات الانموذج الخطي الجزئي. يهدف البحث الى تقدير أنموذج 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 الموضوعي 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
مجلد 46 / العدد 127 / اذار / 2021
الصفحات : 261-249

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 الزمني، 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 إلى ثلاثة أنواع لتسمح بتحليل هذا النوع من البيانات إلى ثلاثة أشكال كالتالي: [5] [4]

1. النموذج التجميعي (Pooled Regression Model) (PRM).
2. نموذج التأثيرات الثابتة (Fixed Effect Model) (FEM).
3. نموذج التأثيرات العشوائية (Random Effect Model) (REM).

في هذه البحث سيتم التعامل مع نموذج البيانات الطولية ذو التأثيرات العشوائية. من مميزات وإيجابيات تحليل البيانات الطولية مقارنةً بتحليل البيانات المقطعية بمفردها أو البيانات الزمنية بمفردها، التحكم في التباين الذي يؤدي إلى نتائج متحيزة عند ظهورها في حالة البيانات المقطعية أو في حالة بيانات السلسلة الزمنية كما و يعتبر هذا النوع من البيانات مناسب لدراسة فترات حالات اقتصادية كثيرة مثل البطالة والفقر ولها إمكانية أفضل 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 المهمة التي نتجت من خصائص للمفردات غير المشاهدة والتي بدورها تؤدي إلى مقدرات متحيزة في الانحدارات المنفردة. [7]

(2) أساليب اختيار النموذج الملائم للبيانات الطولية:

كما ذكرنا سابقاً بوجود ثلاثة أنواع رئيسية من نماذج البيانات الطولية وهي النموذج التجميعي ، نموذج التأثيرات الثابتة، ونموذج التأثيرات العشوائية، لذا يجب علينا اختيار النموذج الأكثر ملائمة للبيانات المدروسة من بين هذه النماذج بالاعتماد على الاختبارات الخاصة التي تم وضعها لهذا الغرض، إذ تمت عملية اختيار النموذج الأفضل بمرحلتين الأولى بأستعمال اختبار مضاعف لاكرانج (Lagrange multiplier Test) إذ تكون المقارنة بين نموذج الانحدار التجميعي ونموذجي التأثيرات الثابتة والعشوائية ، والمرحلة الثانية عن طريق اختبار هوسمان (Hausman Test) الذي يتضمن المقاضلة بين نموذجي التأثيرات العشوائية والتأثيرات الثابتة إذا لم يكن النموذج التجميعي ملائماً للبيانات المدروسة. [1]

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

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

$$\begin{aligned} H_0: \sigma_u^2 &= 0 & \text{نموذج الانحدار التجميعي هو النموذج الملائم} \\ H_1: \sigma_u^2 &\neq 0 & \text{نموذج الانحدار التجميعي غير ملائم} \end{aligned}$$

وإن إحصاءة الاختبار لهذه الفرضيات هي:

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

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

إذ إن:

n : تمثل عدد المشاهدات في الوحدات المقطعية.

t : يمثل الفترة الزمنية في الوحدات المقطعية .

e : متجه بواقي النموذج التجميعي للبيانات الطولية.

$\bar{e}$: متجه متوسطات المجاميع لبواقي النموذج 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 أن النموذج التجميعي غير ملائم للبيانات المدروسة وإن نموذجي التأثيرات الثابتة والعشوائية هما الأكثر ملائمة. كما اقترح (Honda 1985) اختبار 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) اختباراً أحادي الجانب وسهل الاستخدام يتضمن مقارنة الجذر التربيعي لإحصاءة (Breusch-Pagan) بقيمة التوزيع الطبيعي القياسي، وبالتالي فإن إحصاءة اختيار مضاعف لاكرانج بإسلوب (Honda) تكون بالشكل التالي: [1] [6]

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

وذلك لإختبار الفرضية التالية:

$$H_0: \sigma_u^2 = 0 \quad \text{إنموذج الانحدار التجميعي هو الانموذج الملائم}$$

$$H_1: \sigma_u^2 > 0 \quad \text{إنموذج الانحدار التجميعي غير ملائم}$$

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]

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

إذ تمثل:

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

$Var(\hat{\beta}_{REM})$: التباين والتباين المشترك للمعاملات 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 فرضية العدم القائلة بأن التأثيرات العشوائية متنسقة، أي إن إنموذج التأثيرات العشوائية هو إنموذج غير ملائم للبيانات المدروسة، أي إن 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 (PLM) (Partial Linear Model) من أشهر النماذج شبه المعلمية ويطلق عليه أحياناً بالإنموذج شبه المعلمي البسيط ، ويتألف الإنموذج الخطي الجزئي من جزء خطي يمثل الإنحدار المعلمي وجزء لا خطي يمثل الإنحدار اللامعلمي، ويمكن وصف إنموذج البيانات الطولية الخطي الجزئي ذو التأثيرات العشوائية رياضياً وفق الصيغة التالية: [9] [12]

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

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

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 $(t \times 1)$.

ويخضع النموذج شبه المعلمي المعروف في الصيغة (5) الى الشروط والافتراضات التالية:

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

وبذلك فإن النموذج وفق الافتراضات أعلاه سيكون وفق الصيغة التالية:

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

(4) عرض الحزمة: (Bandwidth)

وتسمى بمعلمة التمهيد ويرمز لها عادة بالرمز (h) ، تملك معلمة التمهيد تأثيراً قوياً على نتائج الدوال اللبية وبالتالي فإن 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 اخترنا 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 دالتي اللب (Epanchnikov و Gaussian) وسيتم توظيف طريقة المربعات الصغرى للعبور الشرعي (Least Square Cross Validation) (LSCV) التي تعمل على جعل متوسط مربعات الخطأ التكاملية Mean Integrated Squared Error(MISE) أقل ما يمكن . [3] [5]

(5) طرائق التقدير (Estimation Methods)

Speckman Method (1-5)

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

$$\begin{aligned} E(y_{ij}|z_{ij}) &= E(x_{ij}|z_{ij})'\beta + E[m(z_{ij}|z_{ij})] + E[m(\varphi_{ij}|z_{ij})] \quad \dots (7) \\ \text{وتحت شروط النموذج الخطي الجزئي فإن } E[m(\varphi_{ij}|z_{ij})] &= 0, \text{ وإن } E[m(z_{ij}|z_{ij})] = m(z_{ij}), \text{ لذلك بطرح الصيغة (7) من الصيغة (5) نحصل على:} \end{aligned}$$

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

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

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

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

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

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:

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

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 الموضوعي 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 المعروف في الصيغة (5) كما هو موضح في الخطوات التالية: [9] [2]

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

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

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

إذ يمكن تقدير الدالة $m(\cdot)$ أولاً باستعمال التقدير الخطي الموضوعي متعدد الحدود الذي تم التطرق اليه في المصادر السابقة، وذلك بإفترض أن θ_1, θ_0 تمثل الحلول الناتجة من تصغير مجموع مربعات الخطأ للإنموذج العشوائي، عندها فإن:

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

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

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

إذ أن:

I_{nt} : q من المتجهات التي عناصرها 0 وان اول p من عناصرها تساوي 1.

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

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

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

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

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

H : مصفوفة بالأبعاد $(nt) \times (1 + p)$ ، 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]

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

إذ أن:

$$\hat{Y} = (I_{nt} - S) Y \quad , \hat{Y} = (\hat{Y}_{11}, \dots, \hat{Y}_{nt})$$

$$\hat{X} = (I_{nt} - S) X \quad , \hat{X} = (\hat{X}_{11}, \dots, \hat{X}_{nt})$$

$$\varphi^* = (I_{nt} - S) \varphi_{ij} + (I_{nt} - S) m(Z) \quad , \varphi^* = (\varphi_{11}^*, \dots, \varphi_{nt}^*)$$

$$m(Z) = [m(Z_{11}), \dots, m(Z_{nt})]$$

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

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

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

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 (2018-2010) 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)

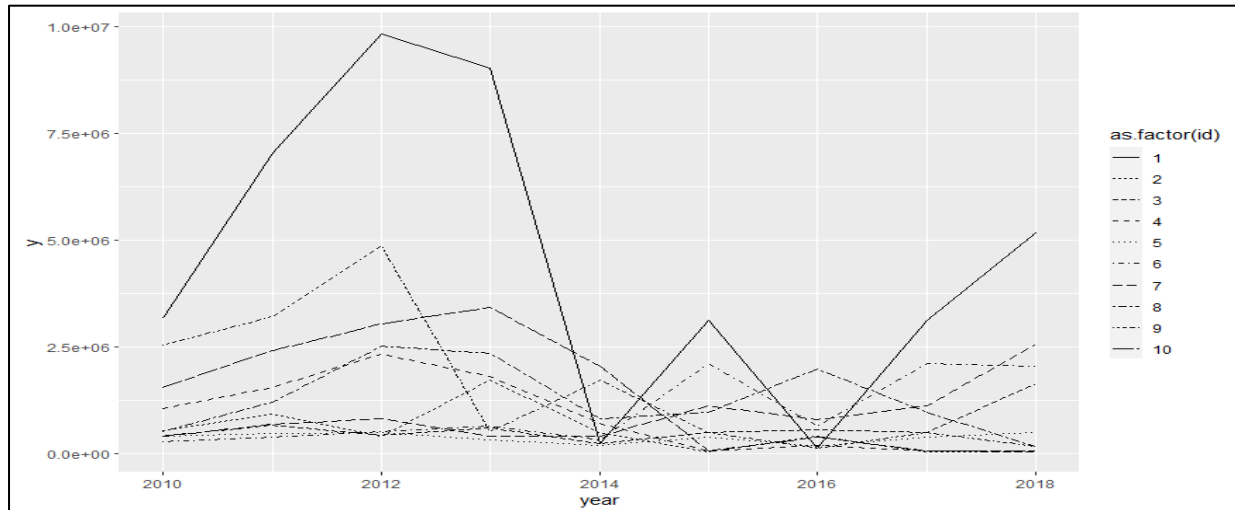
إختبار مضاعف لاكرانج وإختبار هوسمان

الإختبار	قيمة إحصاء إختبار مربع كاي	مستوى الدلالة
إختبار مضاعف 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 أقل من مستوى المعنوية (5%) ، وهذا يعني رفض فرضية العدم أي إن الإنموذج التجميعي غير ملائم لبيانات الدراسة، بينما كانت قيمة مستوى الدلالة لإحصاء إختبار هوسمان (0.8676) وهي أكبر من مستوى المعنوية وهذا يعني قبول فرضية العدم أي إن إنموذج التأثيرات العشوائية أكثر ملائمة من إنموذج التأثيرات الثابتة.

مقارنة بعض طرائق تقدير النموذج شبه معلمي لبيانات طولية



الشكل (أ)

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

جدول رقم (2)

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

رمز المنشأة	السنة	قيمة الانتاج (مليون دينار)	قيمة مستلزمات الانتاج (مليون دينار)	اعداد المشتغلين	رمز المنشأة	السنة	قيمة الانتاج (مليون دينار)	قيمة مستلزمات الانتاج (مليون دينار)	اعداد المشتغلين
1	2010	3158855	1809583	23000	6	2010	280800	241750	30000
1	2011	7052375	3009550	22000	6	2011	373750	238800	30000
1	2012	9847855	4696864	23000	6	2012	510450	300992	31000
1	2013	9046750	3851234	23000	6	2013	640200	332102	31000
1	2014	230860	9119489	26000	6	2014	305000	581208	30000
1	2015	3111477	1551800	19000	6	2015	2102200	1016660	26000
1	2016	117760	2163615	17000	6	2016	635580	278866	22000
1	2017	3111477	1551800	19000	6	2017	2102200	1016660	26000
1	2018	5176621	2757164	20000	6	2018	2043900	1007350	25000
2	2010	533750	269147	18000	7	2010	406000	215724	42000
2	2011	916250	252578	19000	7	2011	689780	250465	42000
2	2012	402794	67766	25000	7	2012	811965	403547	44000
2	2013	1709750	551765	24000	7	2013	399345	199345	34000
2	2014	490000	549200	19000	7	2014	408375	5689484	40000
2	2015	42000	4500	13000	7	2015	1120810	308452	18000
2	2016	402000	137046	18000	7	2016	783360	416097	18000
2	2017	42000	4500	13000	7	2017	1120810	308452	18000
2	2018	39240	3696	13000	7	2018	2569125	542650	22000
3	2010	400985	165488	71000	8	2010	499200	240682	15000
3	2011	657300	372371	71000	8	2011	1196800	644302	21000
3	2012	421540	241398	18000	8	2012	2512000	1280890	42000
3	2013	602700	345312	18000	8	2013	2342400	1062256	17000
3	2014	228930	2662044	71000	8	2014	815520	563003	19000
3	2015	483912	226930	27000	8	2015	967675	514925	41000
3	2016	545175	259477	29000	8	2016	1975506	1333060	43000
3	2017	483912	226930	27000	8	2017	967675	514925	41000
3	2018	1639750	767028	23000	8	2018	174640	56552	34000
4	2010	1053885	416015	34000	9	2010	2533100	1027738	54000
4	2011	1551060	713883	30000	9	2011	3204800	234000	54000
4	2012	2326450	1078726	28000	9	2012	4878235	1642535	54000
4	2013	1808797	867662	29000	9	2013	506004	379530	71000
4	2014	701150	2790070	42000	9	2014	1713680	2056880	54000
4	2015	50604	27876	14000	9	2015	478795	184429	100000

مقارنة بعض طرائق تقدير إنموذج شبه معلمي لبيانات طولية

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

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

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

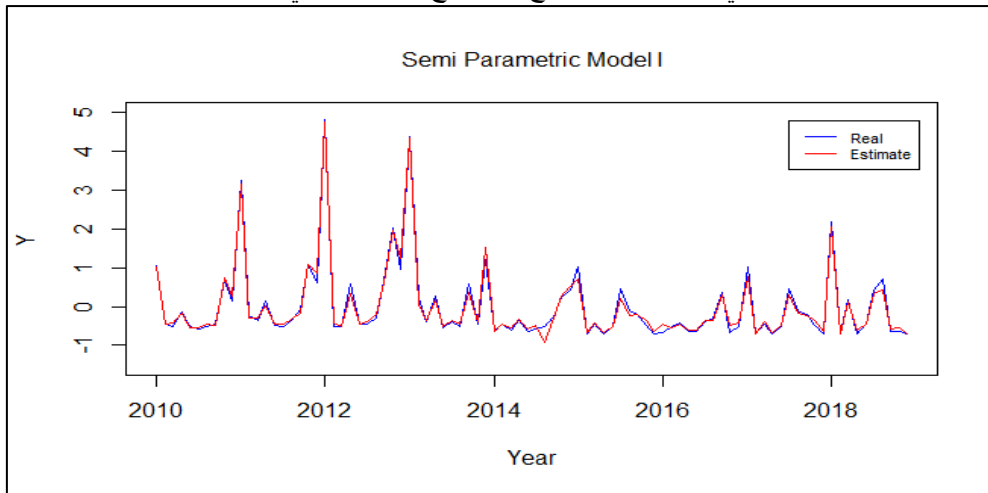
Model	MSE	MAE	MDAE	Best
Semi Parametric Model III	0.00881	0.06339	0.03914	1
Semi Parametric Model II	0.01289	0.07367	0.04006	2
Semi Parametric Model I	0.01403	0.08485	0.06100	3
Semi Parametric Model IV	0.02177	0.09583	0.04916	4
Semi Parametric Model VI	0.06177	0.19245	0.14299	5
Semi Parametric Model 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 الأكثر كفاءة كان إنموذج البيانات الطولية الشبه المعلمي (Semi Parametric Model III) أي الأنموذج 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 الأنموذج المقدر بطريقة ناداراي - واتسون بدالة اللب (Epanechnikov)، وحل ثالثاً (Semi Parametric Model I) اي الأنموذج المقدر بطريقة ناداراي - واتسون ودالة اللب (Gaussian)، في حين جاء بالترتيب الرابع الأنموذج (Semi Parametric Model IV) اي الأنموذج 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 نادرايا 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):

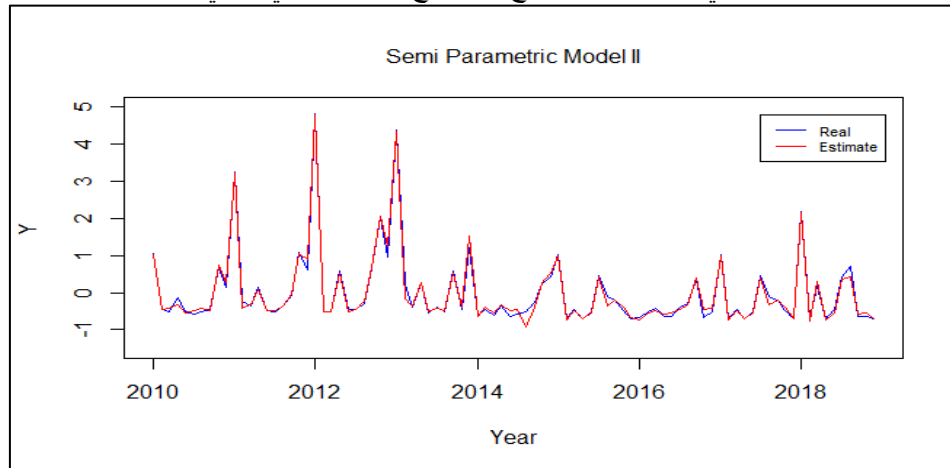
شكل رقم (1)

منحني بيانات قيمة الانتاج للانموذج الشبه المعلمي الأول



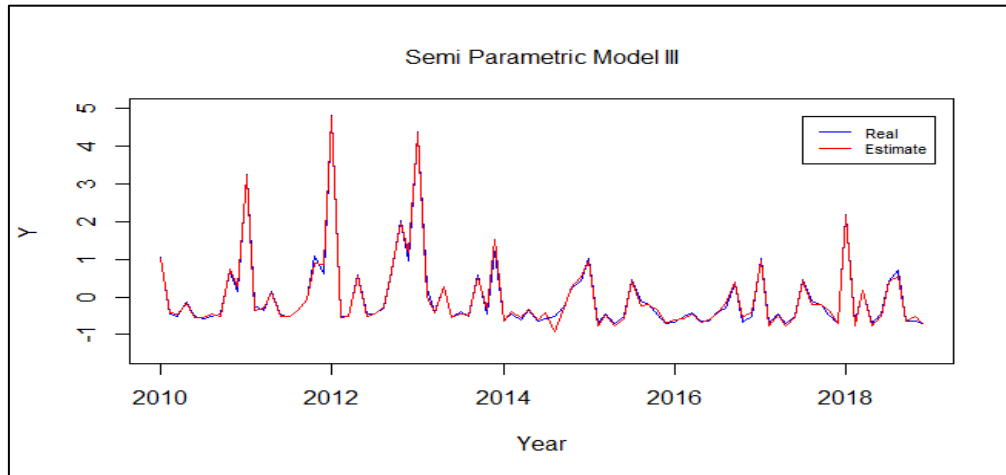
شكل رقم (2)

منحني بيانات قيمة الانتاج للانموذج الشبه المعلمي الثاني

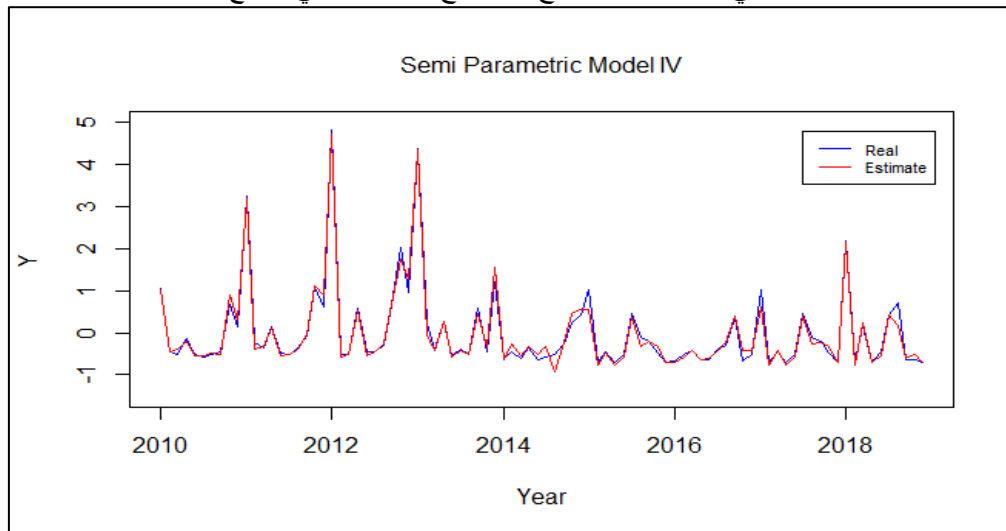


شكل رقم (3)

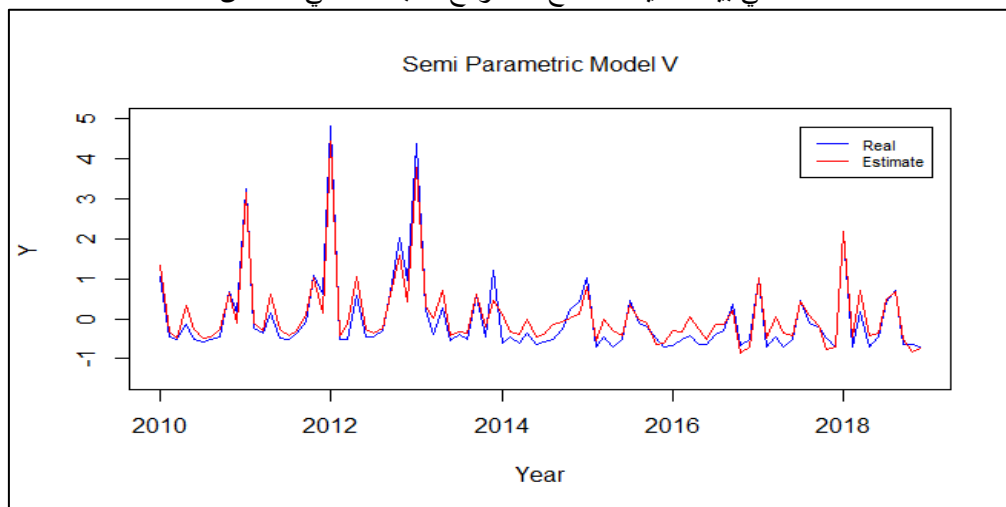
منحني بيانات قيمة الانتاج للانموذج الشبه المعلمي الثالث



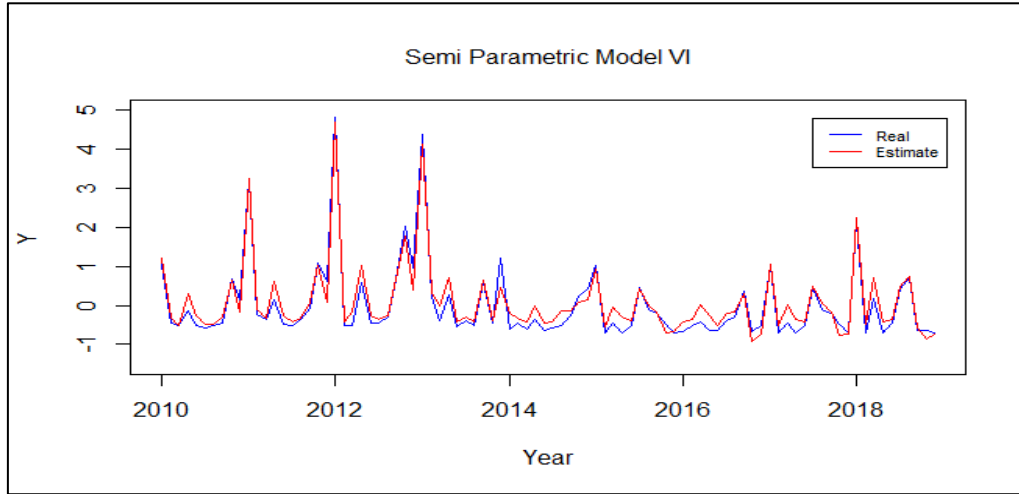
شكل رقم (4)
منحني بيانات قيمة الانتاج للنموذج الشبه المعلمي الرابع



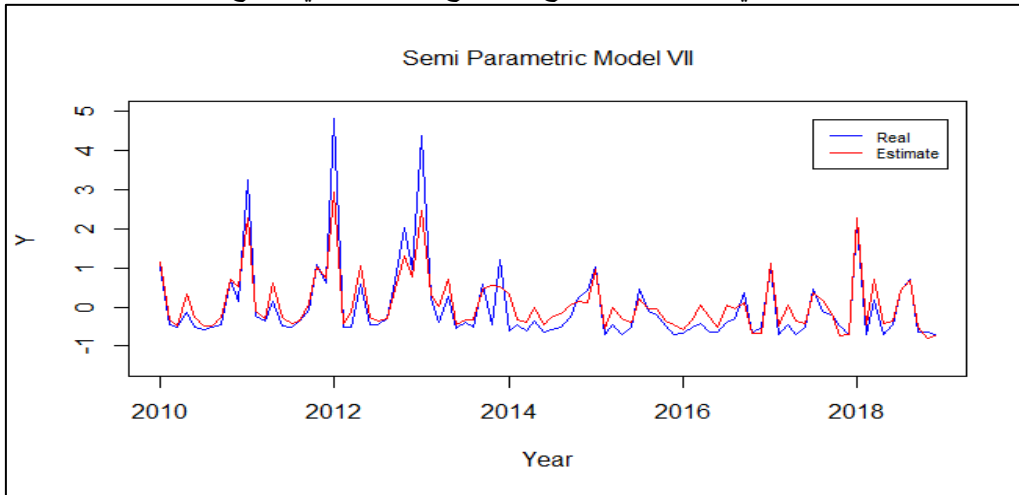
شكل رقم (5)
منحني بيانات قيمة الانتاج للنموذج الشبه المعلمي الخامس



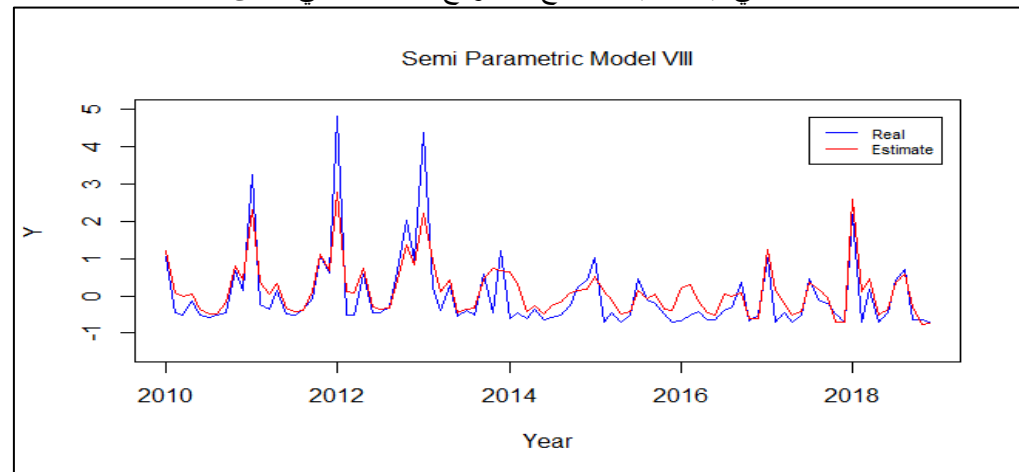
شكل رقم (6)
منحني بيانات قيمة الانتاج للنموذج الشبه المعلمي السادس



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



شكل رقم (8)
منحني بيانات قيمة الانتاج للنموذج 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 لملاحظات المتغير المعتمد مع وجود بعض الفروقات عند السنوات 2014، 2016 و 2018، 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 لملاحظات 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 الملاحظات المقدرة عن الملاحظات 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)

1. من خلال النتائج التي توصل اليها الباحث يمكن ان نستنتج ما يلي:
إن الانموذج الأكثر 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 الانحدار الموضوعي المتعدد من الدرجة الاولى أكثر دقة من التقدير وفق الانحدار الموضوعي بثابت عند دراسة انموذج البيانات الطولية العشوائي لقيمة الإنتاج الصناعي للشركات العشرة.
4. دالة اللب (Gaussian) تعطي تقديرات أكثر 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.

(References) المصادر

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, "Nonparametric Econometrics" 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., Soberon 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 Effectd .

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