

مقارنة مقدرات طريقة المربعات الصغرى السامة للانموذج المعلمي مع مقدرات طريقة ناداريا واتسون الموزونة وغير الموزونة للانموذج المعلمي للبيانات الطولية

عادل ماجد حسن / الباحث

أ.م.د. عدي طه رحيم / المشرف

كلية الإدارة والاقتصاد / الجامعة المستنصرية

المستخلص

يتلوث الماء نتيجة وجود مواد غريبة فيه سواء كانت عضوية او غير عضوية، تعمل هذه المواد على افساد نوعية المياه وهو ما يحصل لنهر دجلة. وتتسبب هذه الفضلات بتلوث مياه نهر دجلة تضر بصحة الانسان مسببة له الامراض، ومن هذه الامراض مرض الاميبا التي اخذت بياناته الممثلة بأعداد اصابات الاشخاص بالمرض لجانب الكرخ والرصافة في بغداد لعام 2018 () كمتغير معتمد (Y) بسلسلة كاملة (12) شهرا"، واخذت تراكيز ملوثات الماء على عدد الاصابات بالمرض تم الاعتماد على سبعة تراكيز ملوثة وهي تمثل المتغيرات التوضيحية (X_i) وهي تمثل بيانات مقطعية لعشرة محطات واقعة على ضفتي نهر دجلة لجانب الكرخ والرصافة. وبسبب التعامل مع المقاطع العرضية والسلسلة الزمنية فان بيانات بانل هي افضل تسمية لها، استعملنا نماذج بيانات بانل المعلمية واللامعلمية بعدة مقدرات لالمعلمية وهي مقدر ناداريا واتسون الموزون و مقدر ناداريا واتسون غير الموزون لتحديد اي من المقدرات هو الاكفأ الذي يعطي افضل انموذج، وكانت عدد الاصابات المسجلة في المستشفيات والمراكز الصحية في جانبي الكرخ والرصافة والتي تعاني من مشكلة عدم تجانس التباين وان المتغيرات لتراكيز الملوثات تعاني من مشكلة الارتباط الذاتي وان الانموذج المعلمي المناسب هو انموذج التأثير العشوائي لبيانات بانل حسب اختبار هاوسمان).

الكلمات المفتاحية: البيانات المقطعية- بيانات بانل - مقدر ناداريا واتسون الموزون- مقدر ناداريا واتسون غير الموزون - عدم تجانس التباين- الارتباط الذاتي- الايمودج المعلمي - اختبار هاوسمان

Comparison Of The Estimators Of The General Least Squares Method Of The Parametric Model With The Estimators Of The Nadaria-Watson Weighted And Unweighted Method Of The Non-Parametric Model Of Longitudinal Data.

Adel Majid Hassan adelmajid093@gmail.com

Auday Taha Rahem audaytaha73@gmail.com

Abatract

The water is polluted as a result of the presence of foreign substances, whether organic or inorganic. These substances spoil the water quality, which is what happens to the Tigris River. These waste products pollute the waters of the Tigris River, harmful to human health, causing diseases, and among these diseases is an amoeba, whose data represented by the numbers of people infected with the disease were taken on both sides of Karkh and Rusafa in Baghdad for the year 2018 as a dependent variable (Y) in a complete series (12) month. The concentrations of water pollutants were taken on the number of cases of the disease. Seven pollutant concentrations were relied upon, which represent the explanatory variables (Xi), which represent cross-sectional data for ten stations located on the banks of the Tigris River on both sides of Karkh and Rusafa. Because of dealing with cross-sections and time series, Panel data is the best name for it. We used parametric and non-parametric Panel data models with several non-parametric estimators, which are the Nadaria Watson weighted estimator and the Nadaria Watson unweighted estimator to determine which of the estimators is the most efficient that gives the best model, and the number of recorded injuries In hospitals and health centers on both sides of Karkh and Rusafa, which suffer from the problem of heterogeneity of variance, and that the variables of pollutant concentrations suffer from the problem of autocorrelation, and that the appropriate parametric model is the random effect model for the panel data according to the Houseman test.

Keywords: Longitudinal Data. General Least Squares. Nadaria-Watson Weighted And Unweighted.

المقدمة

ان النماذج المقدره لبيانات بانل (*panel data*) تكون تقديراتها اكثر كفاءة ومعنوية من التقديرات التي تحصل عليها باستعمال بيانات السلاسل الزمنية فقط او بيانات المقاطع العرضية فقط و تمتاز بإعطائها صورة اوضح للظاهرة قيد الدراسة من تلك التي تعتمد على نوع واحد من البيانات فأنموذج التأثيرات العشوائية المعلمي *Random Effect Model* () يستعمل عندما لا تتحقق فرضيات الانموذج الثابت المعلمي، وايضا توجد مشكلة الارتباط الذاتي في هذه النماذج اللامعلمية وكذلك مشكلة عدم تجانس التباين في هذه النماذج . ومن خلال الدراسات السابقة في عام (2014) قامت الباحثة محمد [7] باستعمال بيانات بانل لتقدير دالة النمو الاقتصادي وذلك بتقديم وتوضيح كيفية الاختيار والتوفيق بين انموذج الاختيار التجميعي وانموذج التأثيرات الثابتة وانموذج التأثيرات العشوائية وتم الاعتماد في ذلك على معيار معامل التحديد واختبار فيشر-ودارين واتسون لاختيار افضل نموذجين جزئيين للأنموذج العام من خلال دراسة ستة متغيرات تفسيرية تؤثر في النمو الاقتصادي حيث شملت الدراسة عشرة دول عربية للفترة من (1980-2010) وفي نفس العام تناول الباحثان (^{10]} Lee & Mukherjee) التقديرات اللامعلمية لأنموذج التأثيرات الثابتة للبيانات المزدوجة باستعمال طريقة المربعات الصغرى الخطية الموضعية ((*LLLS*) لأنموذج التأثيرات الثابتة باستعمال التحويلات لمقدر (*LLLS*)) وبالمقارنة بين تقديرات ((*LLLS*) قبل وبعد اجراء التحويلات توصلنا الى ان مقدرات ((*LLLS*) قبل اجراء التحويلات تمتلك كفاءة تقاربية اكثر من مقدرات ((*LLLS*) بعد اجراء التحويلات، وتتجلى ذلك من خلال تطبيقها في دراسة العلاقة غير الخطية بين الاوكسجين و النتروجين باستعمال منحني Kuznets).

هدف البحث: يهدف البحث الى المقارنة ما بين مقدرات طريقة المربعات الصغرى العامة للأنموذج المعلمي مع مقدرات طريقة ناداريا واتسون الموزونة وغير الموزونة للأنموذج اللامعلمي للبيانات الطولية لتحديد ايهما الافضل والاكفا الذي يعطي اقل قيمة لمعيار المقارنة وذلك عن طريق المقارنة بعض المقدرات المعلمية واللامعلمية. بالإضافة الى تطبيق هذه النماذج على بيانات واقعية تمثل ملوثات ماء نهر دجلة كمتغيرات توضيحية متمثلة بالمتغيرات (*PO*)-- (*SO4*)-- ((*Ca-Mg*)) - ((*T.H.*) - (*BOD5*) - (*4PH*))، لدراسة تأثير هذه الملوثات على عدد الاصابات لمرض الاميبا في محافظة بغداد للقطاعات الصحية لجاني الكرخ والرصافة لمعرفة اي الجانبين قد ازدادت فيه اصابات الاميبا نتيجة التأثير بهذه الملوثات المائية في نهر دجلة.

مشكلة البحث: تتكون البيانات الطولية (*panel data*) من دمج بيانات السلسلة الزمنية مع بيانات المقاطع العرضية ، بالتالي فان البيانات الطولية تعاني من مشكلة عدم تجانس التباينات التي تظهر على الاقل لمقطع واحد من المقاطع العرضية ، وايضا تعاني من مشكلة الارتباط الذاتي التي تظهر بسبب التعامل مع السلسلة الزمنية ، ولذلك فعند

التعامل مع النماذج المعلمية و اللامعلمية لبيانات بانل فيجب ان نختبر الانموذج لمعرفة هل تعاني بياناته من اي مشكلة من هاتين المشكلتين.

4- الجانب النظري

1-4] **بيانات بانل**: تعرف هذه البيانات بانها قياسات متكررة (*Repeated Measurement*) تأخذ لكل قطاع من قطاعات المشاهدات لسلسلة خلال فترة زمنية (اسبوع، شهر، سنة) وتعرف ايضا بانها بيانات نحصل عليها من المشاهدات المتكررة لظاهرة ما والى (n) من المقاطع العرضية (*cross –sections*) وخلال سلسلة زمنية (T) معينة⁽³⁾⁽⁴⁾، تقرا هذه البيانات لظاهرة معينة على اسلويين الاول قراءة فترة من فترات السلسلة الزمنية لكل المقاطع العرضية، اما السلوب الثاني فتقرا بيانات مقطع من المقاطع العرضية لكل فترات هذه السلسلة الزمنية⁽¹⁾، كذلك فانه عند تسجيل هذه البيانات لظاهرة معينة فان هذه الظاهرة ستتغير على مستويين الاول التغير على المستوى الافقي (العرض) فتكون البيانات بيانات مقطعية (*cross –sections data*) اما المستوى الثاني فهو يمثل التغير على المستوى العمودي (الطول) وعندها تكون البيانات عبارة عن بيانات السلسلة الزمنية، ان دمج البيانات المقطعية مع الفترات الزمنية تجعل نماذج البيانات تتمتع بالعديد من المزايا التي وضعها الباحث⁽⁸⁾ (*Blatagi*) منها: 1- التحكم في عدم تجانس التباين الخاص الذي يظهر في حالة البيانات المقطعية او في حالة البيانات الزمنية. 2- تعطي البيانات الطولية الكفاءة الافضل وزيادة في درجات الحرية واقل تعددية خطية بين المتغيرات ومعلومات اكثر ويمكن التعبير عن هذه البيانات كما في الجدول (4)- 1) اذ ان العدد الكلي لجميع المشاهدات المعتمدة هو ($N=n*T$) وكما في الجدول ادناه وهو تصميم لمجموعة من البيانات الطولية وكما يأتي:⁽¹⁰⁾

الجدول (1-4)

Subject	Repeated measurement	y_{ij}	X_{ijr}
1	1	y_{11}	$\chi_{111}, \dots, \chi_{11d}$
1	2	y_{12}	$\chi_{121}, \dots, \chi_{12d}$
1	t_1	y_{1t_1}	$\chi_{1t_11}, \dots, \chi_{1t_1d}$
n	1	y_{n1}	$\chi_{n11}, \dots, \chi_{n1d}$
.	2	y_{n2}	$\chi_{n21}, \dots, \chi_{n2d}$
n	t_n	y_{nt_n}	$\chi_{nt_n1}, \dots, \chi_{nt_nd}$

ولذلك فان البيانات المزدوجة تمتلك تأثيرات المقطع العرضي او تأثيرات الزمن او قد تمتلك كلاهما، و تكون من النوع الثابت او من النوع العشوائي⁽³⁾⁽⁵⁾، ولذلك فان البيانات الطولية

لها ثلاثة اشكال رئيسة هي انموذج الانحدار التجميعي (Pooled Regression Model) وانموذج التأثيرات الثابتة (Fixed Effect Model) وانموذج التأثيرات العشوائية (Random Effect Model) عندها تكون لدينا N من المشاهدات المقطعية مقاسة في T من الفترات الزمنية. تتكون البيانات المزدوجة من بيانات سلسلة زمنية وبيانات مقطعية ولذلك فان البيانات المزدوجة تمتلك تأثيرات المقطع العرضي او تأثيرات الزمن او قد تمتلك كلاهما عندها تكون لدينا N من المشاهدات المقطعية مقاسة في T من الفترات الزمنية، عليه فان انموذج بيانات بانل يعرف بالصيغة التالية $y_{it} = \beta_{o(i)} + \sum_{j=1}^k \beta_j x_j(it) + u_{it}$ ⁽¹⁾⁽²⁾

$$y_{it} = \beta_{o(i)} + \sum_{j=1}^k \beta_j x_j(it) + u_{it} \quad \dots (1)$$

اذ ان: y_{it} و $t = 1, 2, \dots, N$, $i = 1, 2, \dots, T$ اذ ان قيمة متغير الاستجابة للمشاهدة (i) للفترة الزمنية t

$\beta_{o(i)}$: قيمة نقطة التقاطع في المشاهدة (i) و β_j : قيمة ميل خط الانحدار (j) و $x_j(it)$: قيمة المتغير التفسيري (j) في المشاهدة (i) عند الفترة الزمنية (t) و u_{it} : قيمة الخطأ في المشاهدة i عند الفترة (t)

1-2-4) انموذج الانحدار التجميعي (Pooled Regression Model):

يعد هذا الانموذج من ابسط نماذج بيانات بانل حيث تكون فيه جميع المعاملات $\beta_{o(i)}$ و β_j ثابتة لجميع الفترات الزمنية (يهمل اي تأثير للزمن) بإعادة كتابة الانموذج في المعادلة (1) نحصل على انموذج الانحدار التجميعي بالصيغة الآتية: ⁽²⁾⁽⁵⁾⁽³⁾

$$y_{it} = \beta_o + \sum_{j=1}^k \beta_j x_j(it) + u_{it} \quad \dots (2) \quad E(u_{it}) = 0 \quad v(u_{it}) = \sigma_e^2$$

اذ ان: عدد المعلمات $1+k$ و y_{it} : قيمة متغير الاستجابة للمشاهدة (i) عند الفترة الزمنية t و $\beta_{o(i)}$: قيمة نقطة التقاطع في المشاهدة (i) و β_j : قيمة ميل خط الانحدار و $x_j(it)$: قيمة المتغير التفسيري (j) في المشاهدة (i) عند الفترة الزمنية (t) و u_{it} : قيمة الخطأ في المشاهدة i عند الفترة الزمنية (t) .

2-2-4) انموذج التأثيرات الثابتة (Fixed effect model) (FEM):

ان انموذج التأثيرات الثابتة يفترض اختلاف الحدود الثابتة خلال المقاطع العرضية (Cross-sections) خلال وحدات الزمن ، ولذلك تظهر المتغيرات الوهمية (Dummy variables) مع الحدود الثابتة لأنموذج التأثير الثابت ، ويمكن الوصول للصيغة الرياضية للانموذج من خلال فرض انموذج الانحدار التالي: ⁽¹⁶⁾⁽²⁾

$$Y_{it} = (\alpha + \mu_i) + \beta' X_{it} + u_{it} \quad \dots (3)$$

$$Y_{it} = \alpha_i + \beta' X_{it} + u_{it} \quad \dots (4)$$

الذي يمكن إعادة كتابته كالآتي

اذ ان Y_{it} : يمثل المشاهدة (t) من مشاهدات المتغير المعتمد الخاص بالمقطع العرضي (i)، يمثل معلمة الحد الثابت لأنموذج الانحدار الخاص بالمقطع العرضي (i) و B : موجه صفي ذو مرتبة ($k1$) * يضم معلمات الانحدار ($k, \dots, 12$) الخاصة بأنموذج انحدار المقطع العرضي (i) و X_{it} : مصفوفة ($k1$) تضم المشاهدات (t) من مشاهدات المتغيرات التوضيحية الخاصة بأنموذج انحدار المقطع العرضي (i) و u_{it} : يمثل حد الخطأ العشوائي (error term) (t) الاخطاء العشوائية لانحدار المقطع العرضي (i).

4-2-3 | النموذج التأثير العشوائي | Random Effect Model

يفترض انموذج التأثيرات العشوائية اختلاف تباينات الخطأ العشوائي المتكون من حاصل جمع حدين من الخطأ او اكثر ناتج من استعمال انموذج انحدار بوجود الحد الثابت العشوائي (Random constant term) أي وجود (معلمة عشوائية) ($Random Parameter$) وعليه يمكن الوصول الى الصيغة الرياضية لأنموذج التأثير العشوائي ⁽¹³⁾ ($Random effect model$)، ونكتب المعادلة (1) وكالآتي: ⁽¹⁴⁾

$$Y_{it} = (\alpha + \mu_i) + \beta' X_{it} + u_{it}, \quad t = 1, 2, \dots, T$$

$$\dots (4) Y_{it} = \alpha + \beta' X_{it} + (\mu_i + u_{it}), \quad t = 1, 2, \dots, T$$

اذ ان μ_i : يمثل تأثير المقطع (i) وهي مركبة ثابتة خلال الزمن، u_{it} : تمثل تأثير كل من المقطع العرضي (i) والزمن (t) وهي مركبة الاخطاء المشتركة بين المقاطع العرضية والزمن وبافتراض: $\mu_i \sim N(0, \sigma_\mu^2)$ و $u_{it} \sim N(0, \sigma_u^2)$ و $E(\mu_i u_{it}) = 0 \quad \forall i = it$

4-3 | تقدير معلمات الانموذج المعلمي

4-3-1 | تقدير معلمات الانموذج المعلمي الثابت

4-3-2 | تقدير معلمات النموذج التأثير العشوائي

يمكن كتابة دالة الامكان (Likelihood function) لانموذج التأثير العشوائي كالآتي:

$$\dots (5) L(Y | h, \Omega) = \pi^{-0.5N} (0.5)^{0.5N} |\Omega|^{-0.5} \exp[-0.5(Y - Zh)' \Omega^{-1} (Y - Zh)]$$

وعند اخذ اللوغارتم الطبيعي للدالة (5) نحصل على الآتي:

$$\dots (6) LnL = 0.5N [Ln0.5 - Ln\pi] - 0.5Ln|\Omega| - 0.5(Y - Zh)' \Omega^{-1} (Y - Zh)$$

وبأجراء التفاضل الجزئي للدالة اعلاه بالنسبة لموجه المعلمات (h) وهو معلمة التمهيد المؤثرة في نتائج الدوال اللبية وجعل المشتقات الجزئية مساوية للصفر ومن ثم حل المعادلات نحصل على التالي: ⁽¹¹⁾

$$\dots (7) \hat{h} = (Z' \Omega^{-1} Z)^{-1} Z' \Omega^{-1} Y$$

عند استعمال التكرار (Iteration) لاجاد المقدرات $\hat{\beta}_{ML}$ و $\hat{\phi}^2$ نحصل على مقدرات (ML) للمعلمات (ϕ^2, β) اي (الحصول على قيمتين متتاليتين متقاربتين جدا) ويمكن استعمال المقدر $(\hat{\beta} = (X'X)^{-1}XY)$ كقيمة اولية لعملية التقدير.

4-4] النماذج الالعملية للبيانات الطولية

4-4-1] النموذج التأثيرات الثابتة الالعملية للبيانات الطولية

4-4-2] النموذج التأثيرات العشوائية الالعملية للبيانات الطولية

(Nonparametric Panel Data Models With Random Effects) (NREM)

تكون معلمات النموذج التأثيرات العشوائية غير متحيزة وعادة مايفرض بان تباين الخطا غير ثابت لجميع المشاهدات المقطعية وهناك ارتباط ذاتي خلال الزمن بين كل مجموعة من

مجاميع المشاهدات المقطعية في فترة زمنية محددة ، وان حد الخطا $u_{it} \sim (0, \sigma_{\varepsilon}^2)$ والانموذج يكون بالصيغة الآتية: ⁽¹²⁾
... (8)

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

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

اذ ان y_{it} تمثل قيمة متغير الاستجابة في المشاهدة i عند الفترة الزمنية t ، من الدرجة $(1nT*)$ و x_{it} تمثل قيمة المتغير التوضيحي للمشاهدة i عند الفترة الزمنية t ، من

الدرجة $(nT*1)$. $x_{it} \in R^q \quad x_{it} \in R$

$m(.)$: دالة التمهيد غير المعرفة ويمكن تقديرها $E[y_{it}/x_{it}=x]$

$$E[Y_{it} | X_{it} = x] = m(x)$$

v_i : التأثيرات العشوائية المفردة ، $v_i = (v_{1T}, v_{2T}, \dots, v_{nT})'$ ، $i=1,2,\dots,n$ ، $T=1,2,\dots,T$ بدرجة $(T*1)$

5-4] طرق تقدير النموذج التأثيرات العشوائية الالعملية للبيانات الطولية

4-5-1] طريقة المقدر الليي لناداريا واتسون (Nadaraya – Watson Kernel Estimator)

Nadaraya تم استعمال هذا المقدر لأول مرة من قبل كل من الباحثين ((Nadaraya (1961) لما يسمى (Tukey (1964) ان الطريقة تعود الى تقديم سابق لـ (Watson (1961) الذي استعمل فكرة الرسم البياني نفسها في التقدير وهو من Regressogram المقدرات الشائعة الاستعمال في التطبيق لنماذج الانحدار الالعلمي ويمتاز بانه دالة

محددة ومستمرة وذات قيم موجبة وغير سالبة وتكملها مساويا" للواحد ، وبفرض

دالة التوزيع الاحادية . $f(x) = \int f(y, x) dy$ ، $f(Y/X) = \frac{f(y, x)}{f(x)}$ ، فان دالة الانحدار (Y_i/X_i) تكون بالصيغة الآتية: (17) (18)

$$E(Y_i/X_i = x) = \int y f(y/x) dy = \int y \frac{f(x, y)}{f(x)} dy = m(x) \quad \dots(9)$$

ومن ثم فان الهدف الرئيس للانحدار اللامعلمي هو تقدير دالة $\hat{m}(x)$ لإيجاد تقدير انموذج التأثيرات العشوائية اللامعلمية للبيانات المزدوجة بالصيغة (9) ، باستعمال المعدل الموزون للبيانات ، فان الصيغة لهذا الممهد تكون :

$$\hat{m}(x) = \frac{\sum_{i=1}^n \sum_{t=1}^T y_{it} K\left(\frac{x_{it} - x}{h}\right)}{\sum_{i=1}^n \sum_{t=1}^T K\left(\frac{x_{it} - x}{h}\right)} \quad \dots(10)$$

اذ ان (h) : معلمة التمهيد (عرض الحزمة) وقيمته اكبر من الصفر، ويمكن تلخيص خطوات اجراء المقدّر اللبي لناداريا واتسون باستبعاد اثر الارتباط الذاتي من كل مشاهدة من مشاهدات العينة المدروسة بايجاد $\hat{r}_q$ والذي يمثل قيمة معامل الارتباط الذاتي والذي يمثل قيمة (L) عند ايجاد (LCV) لتنقية بيانات العينة من اثر وجوده ثم بعد ذلك يتم ايجاد $\hat{m}_{(x)}$ لإيجاد تقدير انموذج التأثيرات العشوائية اللامعلمية للبيانات الطولية بالصيغة (10) (17) (21)

4-5-2 طريقة مقدر المربعات الصغرى الخطية الموضعية الموزونة

(Aweighted local linear least squares estimator) (AWLLLS)

وهي من طرائق تقدير انموذج التأثيرات العشوائية اللامعلمية للبيانات الطولية عندما ((في عام Li Racine تكون هناك مشكلة عدم التجانس ، وقد تم تناول المقدّر من قبل (2007) ، ولتوضيح عمل هذا المقدّر وصيغة لإيجاد تقدير انموذج التأثيرات العشوائية :وعند استعمال توسيع (21) (18) (19) اللامعلمية للبيانات الطولية نستعمل الصيغة التالية

$$\begin{aligned} \text{تاييلور فان}^{(6)}: y_{it} &= m(x_{it}) + v_i + u_{it} \\ y_{it} &= m(x) + (x_{it} - x) B(x) + v_i + u_{it} \\ y_{it} &= (1, (x_{it} - x)) (m(x), B'(x))' + v_i + u_{it} \\ y_{it} &= Z_{it} \delta(x) + \varepsilon_{it} \quad y_{it} = z_{it} \delta(x) + u_{it} \quad \dots \end{aligned} \quad (11)$$

ويمكن كتابة (11) بصيغة المصفوفات وكالآتي :

$$Y = Z(x)\delta(x) + \varepsilon \quad Y = Z(x)\delta(x) + u \dots$$

(12)

$$Y = (y'_{11}, y'_{21}, \dots, y'_{nT})' : \text{متجه بدرجة } (1)$$

$$\text{و } (nT \times 1) \text{ وهو متجه بدرجة } (1) : Z(x) = (Z'_{11}, Z'_{21}, \dots, Z'_{nT})'$$

تشير الى دالة $(\delta(x))$ و $(nT \times 1)$ وهو متجه الاخطاء بدرجة : $\varepsilon = (\varepsilon'_{11}, \varepsilon'_{21}, \dots, \varepsilon'_{nT})'$ تمهيدية بقيمة واحدة .

ان مصفوفة التباين والتباين المشترك (Ω) تمثل معالجة لمشكلة عدم تجانس وتم استعمالها بغرض استبعاد اثر عدم تجانس التباين في متغيرات الدراسة وتكون بالصيغة الآتية :

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

ينتج: ... (13)

ان $\hat{\delta}(x)$ تمثل مقدرات المربعات الصغرى الخطية الموضعية الموزونة لكل من $B(x)$ و $m(x)$ لأنموذج التأثيرات العشوائية اللامعلمية للبيانات الطولية.

4-6 | **تشخيص الانموذج الملائم للبيانات الطولية**

4-6-1 | **اختبار الانموذج المعلمي للبيانات الطولية**

4-6-1-1 | **اختبار فرضية التجانس للانموذج الثابت**

نستعمل اختبار F وكالآتي
(14) ... (16) (13)

$$F_o = \frac{(u'u_p - u'u_f)/(n-1)}{(u'u_f)/(N-n-k)}$$

اذ $u'u_p$: مجموع مربعات البواقي لأنموذج الانحدار المقيد (المدمج) $(u'u_f)$: مجموع مربعات البواقي لأنموذج التأثير الثابت (غير المقيد)، وبمقارنة قيمة F الحسابية في الصيغة (14) مع قيمة F الجدولية بدرجة حرة $(1-n)$ و $(N-n-k)$ على التوالي ، فاذا كانت $(F_o \leq F_{tab})$ تقبل الفرضية H_0 وان جميع الحدود الثابتة لكل المقاطع العرضية متساوية ومن ثم دمج بيانات السلسلة الزمنية والمقاطع العرضية مع بعض ثم التعامل معها كعينة واحدة تضم (N) من المشاهدات ولها حد ثابت واحد، وبالعكس ذلك نقبل

الفرضية (1H) التي تعني تغير الحد الثابت خلال المقاطع العرضية وبالتالي اعتماد انموذج التأثير الثابت.

2-1-2-4 اختبار تأثير مركبة المقاطع العرضية للانموذج العشوائي

باقترح الباحثان ⁽¹³⁾ (Breusch and Pagan) في عام (1980) اختبارا $H_1: \sigma_\mu^2 \neq 0$ لمدى دقة مركبة المقاطع العرضية الذي يسمى اختبار التأثيرات العشوائية (Testing Random effects) وضعا الفرضية التالية: $H_0: \sigma_\mu^2 = 0$ ضد الفرضية البديلة $H_1: \sigma_\mu^2 \neq 0$ وباستعمال اختبار مضاعف لاكرانج ((Lagrange multiple test) يتم اختبار الفرضية اعلاه وكالاتي: ⁽⁹⁾

$$LM = N(2(T-1))^{-1} \left[\hat{W}' \hat{W} \right]^{-1} \left[\hat{W}' \left(I_n \otimes J_T J_T' \right) \hat{W} \right] - 1 \quad \dots (15)$$

اذ ان:

$\hat{W}' \hat{W}$: مجموع مربعات الخطأ للانموذج المدمج وان الصيغة (15) اعلاه تتوزع توزيع مربع كاي χ^2 وبدرجة حرية واحدة وتقارن قيمة LM المحسوبة مع قيمة مربع كاي χ^2 الجدولية فاذا كان $(LM \leq \chi^2)$ نقبل فرضية العدم وبعبكسه نقبل الفرضية البديلة.

2-2-4-4 الاختبار اللامعلمي لكشف مشكلة عدم تجانس التباين للخطأ :

(Nonparametric test for the detection of the problem of heterogeneity)

توجد العديد من الاختبارات المعلمية المستعملة للكشف عن مشكلة عدم التجانس ، لكن المشكلة تكمن في ايجاد اختبار لامعلمي للكشف عن مشكلة عدم التجانس ، واختبار فرضية العدم $(H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2)$ ضد الفرضية البديلة

$(H_1: \sigma_1^2 \neq \sigma_2^2 \neq \dots \neq \sigma_n^2)$ على الاقل واحد من فرض ثبات تجانس التباين ⁽²⁰⁾ (12)

$$\dots \frac{1}{(n-1) \sum_{i=1}^n R_{i,n}^2} \quad (16)$$

، $R_{i,n}^2 = \frac{1}{2} (Y_{i,n} - Y_{i-1,n})^2$ تحت فرضية تحقق $E[R_{i,n}^2] \approx \sigma^2(x_{i,n})$ ، ولاختبار

الفرضية $E(uu') = \sigma_u^2 I_n$ وقد قام العديد من الباحثين بتطوير الاختبار فتم التوصل للصيغة النهائية ⁽¹⁹⁾ وكالاتي:

17))...

$$T_n = \frac{1}{(n-1)(n-2)h} \sum_{|i-j| \geq 2} K \left(\frac{x_{i,n} - x_{j,n}}{h} \right) (R_{i,n}^2 - \bar{R}_n^2) (R_{j,n}^2 - \bar{R}_n^2)$$

$i = 1, \dots, n, j = 0, \dots, n$

ويعتمد القرار النهائي على ايجاد قيمة احصاء الاختبار فاذا كانت قيمة الاحصاء اكبر من قيمة χ^2 المجدولة لدرجة حرية k (عدد المتغيرات التوضيحية) وفيها نقبل H_1 وهذا يعني ان الانموذج يعاني من مشكلة عدم تجانس التباين وبعبكسه نقبل الفرضية H_0 ⁽¹⁴⁾

4-7] اختيار الانموذج الملائم اختبار مضاعف لاكرانج

[Lagrange multiplier Test

اقترح كل من (1980 Breusch-Pagan) يعتمد على إحصاءه تتبع توزيع χ^2 بدرجة حرية واحدة وذلك للمفاضلة بين الانموذج التجميعي والا نموذجين الثابت والعشوائي ، وعرف هذا الاختبار بمضاعف لاكرانج (LM) (Lagrange multiplier Test)، ويعتمد على الاخطاء المقدرة بطريقة المربعات الصغرى الاعتيادية (OLS) لاختبار الفرضية الآتية: ⁽¹⁶⁾

$$H_1: \sigma_u^2 \neq 0$$

$$H_0: \sigma_u^2 = 0 \quad \text{فان احصاء الاختبار للفرضيات}$$

هي: $LM =$

$$\frac{nt}{2(t-1)} \left[\frac{\sum_{i=1}^n (\sum_{j=1}^t u_{ij})^2}{(\sum_{i=1}^n \sum_{j=1}^t u_{ij})^2} - 1 \right]^2 \quad \dots (18)$$

$$LM = \frac{nt}{2(t-1)} \left[\frac{t^2 \underline{u} \underline{u}'}{\underline{u} \underline{u}'} - 1 \right]^2 \sim \chi^2_{(1)} \quad \dots (19)$$

$\underline{u}$: متجه بواقي الانموذج التجميعي للبيانات الطولية و $\underline{u}$: متجه متوسطات المجاميع لبواقي الانموذج التجميعي. اذا كانت قيمة LM المحسوبة اكبر من قيمة $\chi^2_{(1)}$ الجدولية او اذا كانت قيمة (P-value) اقل من مستوى المعنوية ، عندها ترفض فرضية العدم وهذا يعني ان الانموذج التجميعي غير ملائم للبيانات المدروسة وان أنموذج التأثيرات الثابتة والعشوائية هما الاكثر ملائمة.

كما اقترح (1985 Honda) اختبار اخر لمضاعف لاكرانج لتجاوز الشرط الذي يجب توفره وفقا لأسلوب (Breusch-Pagan) في الفرضية البديلة وهي ان التباين لا يساوي صفراي بشرط عدم انعدام التباين، فقد قدم (Honda) اختبارا احادي الجانب وسهل

الاستعمال يتضمن مقارنة الجذر التربيعي لإحصاء (Breusch-Pagan) بقيمة التوزيع الطبيعي القياسي، فان احصاء اختبار مضاعف لاكرانج بأسلوب (Honda) تكون:

$$LMH = \sqrt{\frac{nt}{2(t-1)}} \left[\frac{\sum_{i=1}^n \left(\sum_{j=1}^t u_{it} \right)^2}{\left(\sum_{i=1}^n \sum_{j=1}^t u_{it} \right)^2} - 1 \right] \sim N(0,1) \quad \dots (20)$$

وذلك لاختبار الفرضية التالية:

فرضية العدم) انموذج الانحدار التجميعي هو الانموذج $H_0: \sigma_u^2 = 0$ (الملائم)

(الفرضية البديلة) انموذج الانحدار التجميعي غير ملائم $H_1: \sigma_u^2 > 0$

4-8) معايير اختيار افضل نموذج للبيانات الطولية.

في كثير من التطبيقات الاحصائية يرغب صانع القرار في الوصول الى الانموذج الافضل بحكم البيانات قيد الدراسة عن طريق ايجاد افضل طريقة تقدير بالنسبة للانموذج المعلمي واللامعلمي للبيانات الطولية لبناء القرار الامثل عليه، وسوف نتطرق الى استعمال المعايير التالية :

اولا: معيار متوسط مربعات الاخطاء

(Average mean square error)

$$AMSE = \frac{1}{R} \sum_{i=1}^R Mse(i) \quad \dots (21)$$

(Mse(i) تمثل متوسط مربعات الخطأ

ومعيار الجذر (N) لمعدل متوسط مربعات الاخطاء

(N th root of the avarge of mean square error) وهو يمثل الجذر N لمعدل

متوسط مربعات الاخطاء ورمزه (RAMSE)

$$RAMSE = \sqrt[N]{AMSE} \quad \dots (22)$$

5- الجانب التطبيقي

5-1) المقدمة.

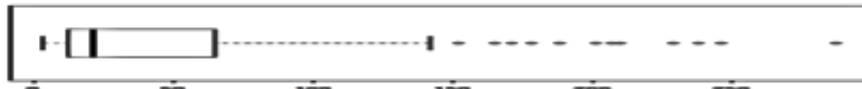
اعتمدت هذه الدراسة على بيانات تراكيز بعض ملوثات الماء في نهر دجلة (وهو اساس الموارد المائية في بغداد) وبحسب موقع المحطات على ضفتي النهر في جانب الكرخ او في جانب الرصافة و بيانات اعداد الاصابات بمرض الاميبا كمتغير معتمد ولجانب الكرخ (والرصافة لعام 2018) وقد استعمل الباحث البرنامج الجاهز (R) في التطبيق

[2-5] وصف البيانات ان المتغيرات التوضيحية: تتمثل بالملوثات التي لها تأثير على نسبة تلوث المياه ، وحسب المحددات البيئية لنظام صيانة الانهار رقم ((25 لسنة 1967) في العراق تم اختيار الملوثات السبعة التالية والتي اخذت ونسبها هي: PH ويدل على حموضة المياه ويجب ان تقع قيمته بين 6.5-8.5 ملغم/لتر. $5BOD$: يدل على الحاجة البيولوجية للأوكسجين اي وجود الاوكسجين في الماء والمحدد البيئي لهذا المتغير الذي يبلغ ($5 >$) ملغم/لتر. $T.H$: ويمثل عسرة الماء تنتج من ارتفاع وجود Ca و Mg والمحدد البيئي لها ان تكون اقل ((180 ملغم/لتر. Ca : الكالسيوم وان قيم لعينات الماء تختلف في الفصول وتكون في الصيف (10-1042) ملغم/لتر وفي الشتاء تتراوح ما بين (16-954) ملغم/لتر. Mg : المغنيسيوم وتتراوح قيمته في الصيف ما بين (12-256) ملغم /لتر و الشتاء يتراوح ما بين (29-386) ملغم /لتر.

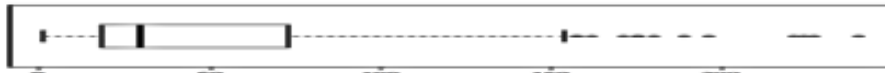
$4:SO$ الكبريتات المحدد البيئي لهذا المتغير الذي يبلغ اقل من (200) ملغم /لتر. $4PO$: الفوسفات وقيمته يجب ان لا تزيد عن ((0-4 ملغم / لتر.

[3-5] عينة البحث والمتغيرات المستعملة: لقد اخذت عينة بحجم 240 مشاهدة بواقع عينة واحدة من كل محطة لكل شهر وذلك من قبل الاختصاصيين العاملين في قسم ملوثات المياه في مديرية بيئة بغداد التابعة لدائرة حماية وتحسين البيئة في منطقه الوسط في وزارة الصحة / البيئة / بواقع سلسلة كاملة 12 شهرا وبيانات مقطعية بعدد (10) محطات ببيانات كاملة اذ تمثل البيانات المستعملة في الجانب التطبيقي نتائج الفحوصات المختبرية لمياه نهر دجلة خلال عام (2018) للمحطات الواقعة على ضفتي النهر لجانب الكرخ والرصافة وهذه المحطات هي ($17T$) جسر-المثنى، ($18T$) جسر-الائمة، ($19T$) جسر-الشهداء، ($21T$) جسر-الجادرية، ($22T$) مشروع ماء الرشيد، ($23T$) محطة ماء الزعفرانية، ($24T$) مشروع ماء الدورة، ($16D$) نهر دياي / جسر-ديالى الجديد، ($15T$) نهر المصب العام في ابي غريب، ($25T$) نهر المصب العام في المحمودية وكل محطة مرتبطة بالقطاعات الصحية المكونة من مستشفيات ومراكز صحية تابعة لوزارة الصحة أجري فيها الفحص المختبري لأشخاص لأصابتهم بمرض الاميبا .

[5-4] تحليل البيانات: تم تطبيق طريقة ($Boxplot$) للكشف القيم الشاذة في المتغير المعتمد (عدد الإصابات بمرض الاميبا في مدينة بغداد في جانبي الكرخ والرصافة) كما هو مبين في الشكلين (1-5 و 2-5)، وقد تبين وجود قيم شاذة في لكلا الجانبين.



الشكل (1-5) (الرسم الصندوقي لأعداد المصابين بمرض الاميبا في بغداد / الكرخ لعام 2018)



الشكل (2-5) (الرسم الصندوقي لأعداد المصابين بمرض الاميبا في بغداد / الرصافة لعام 2018) ونظرا لعدم توافق وحدات القياس للبيانات المدروسة تحول الى الصيغة القياسية $Z = \frac{X - \mu}{\sigma}$ (standardization). ولتشخيص نوع إنموذج بيانات البائل تم استعمال اختبار مضاعف لاكرانج وفق اسلوب (هوندا) لتحديد مطابقة البيانات للأنموذج التجميبي، وذلك لاختبار فرضية عدم القائلة بأن الأنموذج التجميبي يلائم البيانات المدروسة، وفي حالة عدم مطابقة البيانات للأنموذج التجميبي عندها فإنه يجب المفاضلة بين إنموذجي التأثيرات العشوائية والتأثيرات الثابتة من خلال اختبار (هوسمان) وذلك للبت في فرضية عدم القائلة بأن إنموذج التأثيرات العشوائية أكثر ملائمة من إنموذج التأثيرات الثابتة، فضلا عن اجراء اختبارات الكشف عن وجود مشكلة عدم تجانس تباينات المقاطع وذلك من خلال استعمال اختبار (Breusch-Pagan test) وذلك لاختبار فرضية عدم التي تنص على عدم وجود مشكلة عدم تجانس تباينات المقاطع العرضية لبيانات البائل، ونتائج الاختبارات المذكورة في الجدولين (1-5) و (2-5) أدناه:

جدول (5-1) اختبار مضاعف لاكرانج واختبار هوسمان واختبار عدم تجانس التباين لبيانات الكرخ

الاختبار	قيمة إحصاء الاختبار	(p-value)
اختبار مضاعف لاكرانج لبيانات الكرخ	12.237	2.2 e-16
اختبار هوسمان لبيانات الكرخ	0.7176	0.9982
اختبار عدم تجانس تباينات المقاطع لبيانات الكرخ	21.408	0.0447

جدول (5-2) اختبار مضاعف لاكرانج واختبار هوسمان واختبار عدم تجانس التباين لبيانات الرصافة

الاختبار	قيمة إحصاء الاختبار	(p-value)
اختبار مضاعف لاكرانج لبيانات الرصافة	11.887	2.2 e-16
اختبار هوسمان لبيانات الرصافة	5.2236	0.6327
اختبار عدم تجانس تباينات المقاطع لبيانات الرصافة	32.417	0.00119

يتضح من الجدول (1-5) والذي يخص اختبارات بيانات الكرخ ان قيمة (p-value) لاختبار مضاعف لاكرانج قد بلغت (2.2 e-16) وهي اقل من مستوى المعنوية (5%)، وهذا يعني رفض فرضية عدم أي إن الأنموذج التجميبي غير ملائم لبيانات الكرخ. بينما بلغت قيمة (p-value) لإحصاء اختبار هوسمان لبيانات الكرخ (0.9982) وهي أكبر من مستوى المعنوية، وهذا يدفعنا الى قبول فرضية عدم، أي إن الانموذج العشوائي أكثر ملائمة من الإنموذج الثابت في تمثيل بيانات الإصابات بمرض الاميبا لبيانات جانب الكرخ. في حين بلغت قيمة (p-value) الخاصة باختبار (Breusch-Pagan test) لبيانات الكرخ (0.0447)، وهذا يؤدي الى رفض فرضية عدم H_0 التي تنص على تجانس في تباينات

المقاطع العرضية، وبذلك فان بيانات الإصابة بمرض الاميبا لبيانات جانب الكرخ غير متجانسة في تباينات المقاطع العرضية.

كما ويتضح من خلال الجدول (2-5) الذي يمثل اختبارات بيانات الرصافة ان قيمة p -value) اختبار مضاعف لاكرانج قد بلغت (2.2-e-16) وهي اقل من مستوى المعنوية (5%)، وهذا يعني رفض فرضية العدم أي إن الأنموذج التجميعي غير ملائم لبيانات الرصافة. بينما بلغت قيمة p -value) لإحصاء اختبار هاوسمان لبيانات الرصافة (0.6327) وهي أكبر من مستوى المعنوية، وهذا يدفعنا الى قبول فرضية العدم، أي إن الانموذج العشوائي أكثر ملائمة من الأنموذج الثابت في تمثيل بيانات الإصابات بمرض الاميبا لبيانات جانب الرصافة. في حين بلغت قيمة p -value) الخاصة باختبار (Breusch-Pagan test) لبيانات الكرخ (0.00119) وهي اقل من مستوى المعنوية، وهذا يؤدي الى رفض فرضية العدم التي تنص على تجانس في تباينات المقاطع العرضية، وبذلك فان بيانات الإصابة بمرض الاميبا لبيانات جانب الرصافة غير متجانسة في تباينات المقاطع العرضية.

3-5 تقدير الإنموذج: بعد التعرف على طبيعة البيانات المدروسة ووصف إنموذج بيانات البائل الملائم لتلك البيانات وهو الانموذج العشوائي، سيتم تطبيق طرائق التقدير التي تم تناولها في الجانب النظري على بيانات مرض الاميبا لجانب الكرخ والرصافة، تم تقدير معلمات الانموذج المعلمي العشوائي لجانب الكرخ والرصافة ووضع النتائج في الجدول (3-5). الجدول (3-5) (القيم التقديرية لمعلمات الانموذج العشوائي لبيانات الكرخ والرصافة)

Parameter	β_1	β_2	β_3	β_4	β_5	β_6	β_7
الكرخ	0.0158	0.0064	0.0183	-0.0151	0.0331-	-0.0272	0.0179
الرصافة	0.0189	-0.0267	0.0335	-0.0269	0.0255	-0.1104	0.1254

وللمقارنة بين تلك طرائق التقدير المختلفة وبيان الانموذج الأفضل تم استعمال معايير المقارنة معدل متوسط مربعات الخطأ (RAMSE)، معدل وسيط الانحرافات المطلقة (AMADE)، معدل متوسط الانحرافات المطلقة (AMAE) ومعدل متوسط الانحرافات المطلقة النسبية (AMAPE)، وذلك للحصول على افضل انموذج عشوائي يمثل تلك البيانات. ولتجنب الخلط بين طرائق التقدير المختلفة فقد تم وضع أسماء وتفاصيل المقارنة بين تلك الطرائق في الجداول (4-5) و (5-5).

جدول (4-5) النماذج المستعملة في التقدير

	Model Name	Method
1	Parametric I	GLS
2	Non Parametric II	NW, Gaussian
3	Non Parametric III	NW, Epanchincov
4	Non Parametric IV	WNW, Gaussian
5	Non Parametric V	WNW, Epanchincov

جدول (5-5) قيم معايير المقارنة بين النماذج المقدرة لبيانات الكرخ

Model	AMSE	AMDAE	AMAE	AMAPE	Best
1	0.9692556	0.61667	0.746665	1.024608	5
2	0.7579374	0.4865029	0.6422889	2.289197	4
3	0.753898	0.4789538	0.636211	2.392592	3
4	0.00168707	0.00517717	0.01962688	0.0546324	1
5	0.01526787	0.00770137	0.04914785	0.1070358	2

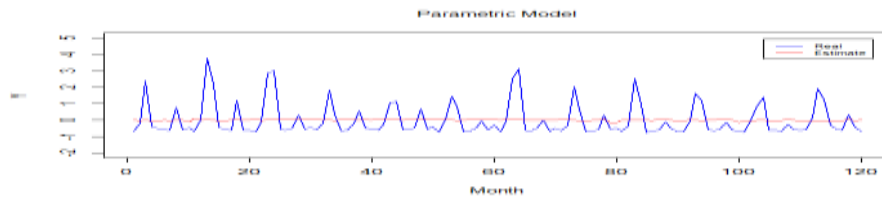
يتضح من خلال الجدول (5-5) الذي يخص معايير المقارنة المقدرة للنماذج المختلفة لبيانات الكرخ ان النموذج اللامعلمي (Non Parametric IV) المقدر وفق طريقة ندارايا واتسون الموزونة بالدالة اللبية (Gaussian) هو الأنموذج الأكثر كفاءة من بين النماذج كافة كونه اعطى اقل قيم لمعايير المقارنة الأربعة، وحل بالترتيب الثاني الانموذج اللامعلمي (Epanechnikov (Non Parametric Model V)) حسب قيمة معيار المقارنة (AMSE) والمقدر بطريقة ندارايا واتسون الموزونة بالدالة اللبية (Epanechnikov) وبالترتيب الثالث حسب معيار المقارنة (AMSE) جاء الانموذج اللامعلمي والمقدر وفق الانحدار الموضعي الخطي الموزون بالدالة اللبية (Gaussian Non Parametric II) فقد حل بالترتيب الرابع لكافة معايير المقارنة، في حين حل الانموذج اللامعلمي الانموذج العشوائي المعلمي بالترتيب الخامس بأعلى قيمة لمعايير المقارنة ((AMSE،AMDAE) و (AMAE).

جدول (6-5) قيم معايير المقارنة بين النماذج المقدرة لبيانات الرصافة

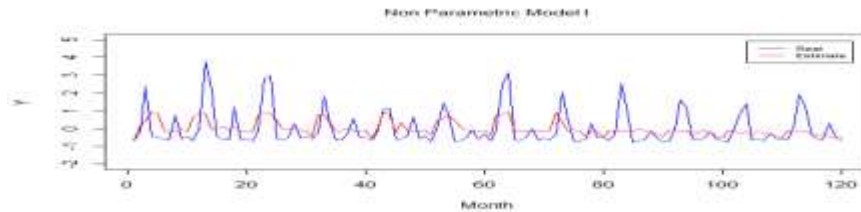
Model	RAMSE	AMDAE	AMAE	AMAPE	Best
L	0.9692556	0.616167	0.7749073	0.9863455	5
II	0.8961828	0.649214	0.7617201	1.11913	4
III	0.8958098	0.6448439	0.7606514	1.115425	3
IV	0.0399002	0.0400698	0.1050109	0.2428438	1
V	0.0766568	0.06150961	0.1516696	0.3061616	2

كما ويتضح من خلال الجدول (6-5) الذي يخص معايير المقارنة المقدرة للنماذج المختلفة لبيانات الرصافة ان النموذج اللامعلمي (Non Parametric V) المقدر وفق طريقة الانحدار الموضعي الخطي الموزون بالدالة اللبية (Gaussian) هو الأنموذج الأكثر كفاءة من بين النماذج كافة كونه اعطى اقل قيم لمعايير المقارنة (AMSE) و (AMAE) وحل بالترتيب الثاني الانموذج اللامعلمي (Non Parametric VI) والمقدر وفق طريقة الانحدار الموضعي الخطي الموزون بالدالة اللبية ((Epanechnikov حسب معيار المقارنة (AMSE) و (AMAE) وجاء الانموذج اللامعلمي (Non Parametric Model II) والمقدر بطريقة ندارايا واتسون بالدالة اللبية (Epanechnikov) الموزونة بالترتيب الثالث حسب معايير المقارنة الأربعة ، وحل رابعا الانموذج اللامعلمي والمقدر بطريقة

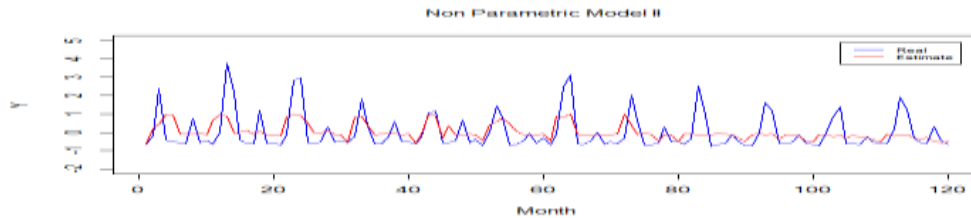
نداريا واتسون بالدالة اللبية (Gaussian) ولكافة معايير المقارنة، في حين جاء خامسا" الانموذج المعلمي بأعلى قيمة لمعايير المقارنة ((AMSE،AMDAE) و (AMAE). ولتوضيح عملية المقارنة بصورة اكثر دقة تم رسم المنحنيات الحقيقية والمقدرة والخاصة بنماذج بيانات البائل العشوائية لمتغير اعداد الاصابة بمرض الاميبا وكما هو موضح بالأشكال من (3-5) الى (12-5) كما في ادناه.



شكل (3-5) منحنى بيانات اعداد الاصابة بمرض الاميبا للأنموذج Parametric I لبيانات الكرخ



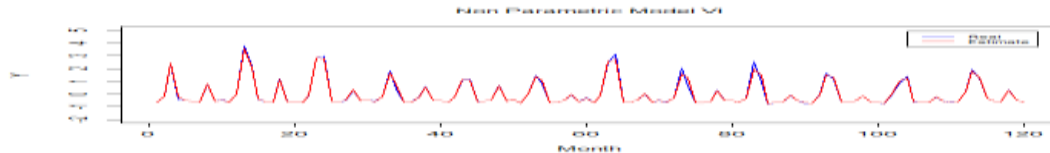
شكل (4-5) منحنى بيانات اعداد الاصابة بمرض الاميبا للأنموذج Non Parametric II لبيانات الكرخ



شكل (5-5) منحنى بيانات اعداد الاصابة بمرض الاميبا للأنموذج Non Parametric III لبيانات الكرخ



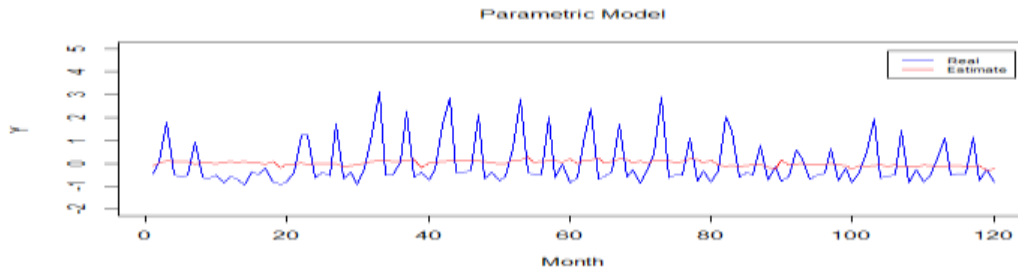
شكل (6-5) منحنى بيانات اعداد الاصابة بمرض الاميبا للأنموذج Non Parametric IV لبيانات الكرخ



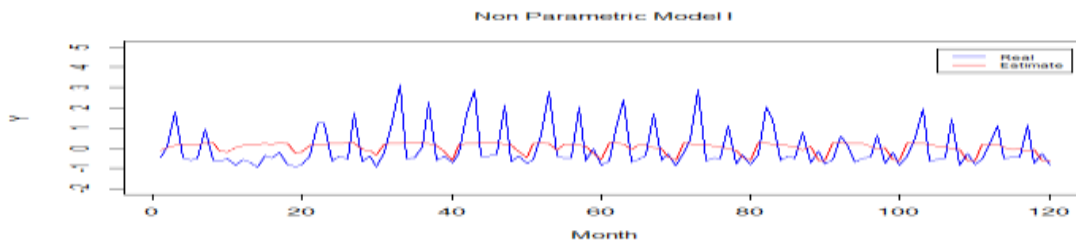
شكل (7-5) منحنى بيانات اعداد الاصابة بمرض الاميبا للأنموذج Non Parametric V لبيانات الكرخ

وقد تبين من خلال الاشكال من (4-5) الى (7-5) والتي تخص بيانات جانب الكرخ ما يأتي:

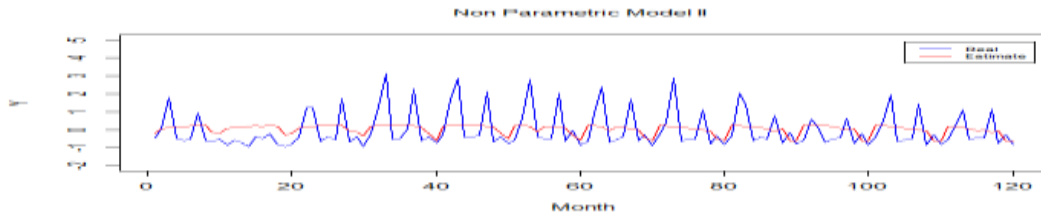
1. التمهيد وفق الانموذج اللامعلمي (Non Parametric Model IV) والموضح في الاشكال (5-6) يبدو الاقرب لمشاهدات المتغير المعتمد الحقيقية، ويمكن القول ان التمهيد وفق هذا الانموذج هو الافضل من بين كافة النماذج، كونه يكاد يكون مطابقا لمنحني المتغير المعتمد الحقيقي.
2. الشكل (5-7) يمثل التمهيد للانموذج اللامعلمي (Non Parametric Model V) يحتل المرتبة الثانية من حيث الأفضلية، ويظهر من خلاله فرق بسيط في التمهيد بين المشاهدات المقدرة والمشاهدات الحقيقية .
3. التقدير بالاعتماد على الانموذج اللامعلمي (Non Parametric Model IV) والموضح في الشكل (5-5) جاء في الترتيب الثالث، وذلك من خلال تطابق القيم المقدرة من القيم الحقيقية.
4. الشكل (5-4) يخص منحنى الانموذج اللامعلمي (Non Parametric Model III) ، اذ حل هذا الانموذج بالترتيب الرابع اذ نلاحظ ابتعاد المشاهدات للمنحنى الممهد عن المشاهدات الحقيقية .
5. الشكل (5-3) يمثل منحنى التمهيد للانموذج المعلمي ويتضح من خلاله ان هذا الانموذج قد حل بالترتيب الأخير من بين كافة النماذج، إذ يمكن ملاحظة ان المشاهدات المقدرة بعيدة عن المشاهدات الحقيقية للمتغير المعتمد.



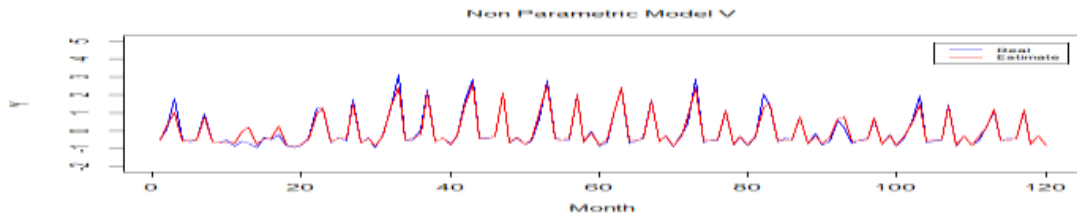
شكل (5-8) منحنى بيانات اعداد الاصابة بمرض الاميبا للانموذج I Parametric لبيانات الرصافة



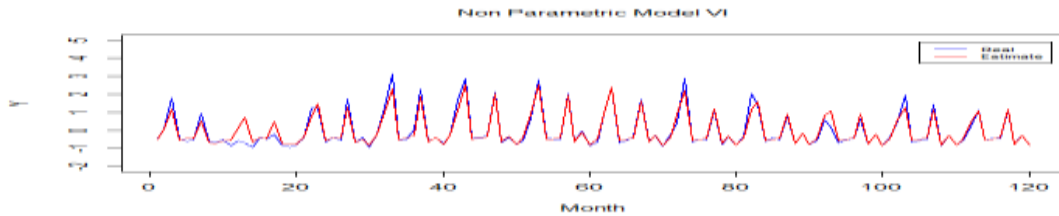
شكل (5-9) منحنى بيانات اعداد الاصابة بمرض الاميبا للانموذج II Non Parametric لبيانات الرصافة



شكل (10-5) منحنى بيانات اعداد الاصابة بمرض الاميبا للانموذج Non Parametric III لبيانات الرصافة



شكل (11-5) منحنى بيانات اعداد الاصابة بمرض الاميبا للانموذج Non Parametric IV لبيانات الرصافة



شكل (12-5) منحنى بيانات اعداد الاصابة بمرض الاميبا للانموذج Non Parametric V لبيانات الرصافة

ان الاشكال من (8-5) الى (12-5) فتخص منحنيات تمهيد المقدرة لبيانات جانب الرصافة كما يأتي:

1. التمهيد وفق الانموذج اللامعلمي (Non Parametric Model IV) والموضح في الشكل (11-5) يبدو الاقرب لملاحظات المتغير المعتمد الحقيقية، ويمكن القول ان التمهيد وفق هذا الانموذج هو الافضل من بين كافة النماذج، كونه يكاد يكون مطابقا لمنحنى المتغير المعتمد الحقيقي.

2. الشكل (12-5) يمثل التمهيد للانموذج اللامعلمي (Non Parametric Model V) يحتل المرتبة الثانية من حيث الأفضلية، ويظهر من خلاله فرق بسيط في التمهيد بين المشاهدات المقدرة والمشاهدات الحقيقية .

3. التقدير بالاعتماد على الانموذج اللامعلمي (Non Parametric Model III) والموضح بالشكل (10-5) جاء في الترتيب الثالث، وذلك من خلال تطابق القيم المقدرة من القيم الحقيقية.

4. الشكل (5-9) يخص منحنى الانموذج اللامعلمي (Non Parametric Model II) ، اذ حل هذا الانموذج بالترتيب الرابع اذ نلاحظ ابتعاد المشاهدات للمنحنى الممهد عن المشاهدات الحقيقية .

5. الشكل (5-8) يمثل منحنى التمهيد للانموذج المعلمي ويتضح من خلاله ان هذا الانموذج قد حل بالترتيب الأخير من بين كافة النماذج، إذ يمكن ملاحظة ان المشاهدات المقدرة بعيدة عن المشاهدات الحقيقية للمتغير المعتمد.

الاستنتاجات والتوصيات

الاستنتاجات

من خلال النتائج ومنحنيات التمهيد التي تم الحصول في الجانب التطبيقي للرسالة فقد توصل الباحث الى الاستنتاجات الآتية:

1. بيانات اعداد الاصابة بمرض الاميبا التي تخص جانبي الكرخ والرصافة هي بيانات ملوثة، تعاني من مشكلة عدم تجانس التباين حسب اختبار (Breusch-Pagan)، فضلا عن كونها تعاني من ارتباط بيانات المتغيرات التوضيحية السبعة التي تمثل خصائص المياه الصالحة للشرب مع الخطأ العشوائي.

2. عدم التجانس لتباينات المقاطع يظهر في بيانات جانب الكرخ بشكل اكبر من بيانات جانب الرصافة.

3. ان الانموذج الملائم لتحليل ودراسة بيانات اعداد الاصابة حسب اختبار مضاعف لاكرانج بأسلوب (هوندا) هو انموذج بيانات البانل ذو التأثيرات العشوائية حسب اختبار (هاوسمان).

4. من خلال انموذج بيانات البانل ذو التأثيرات العشوائية المعلمي لجانب الكرخ تبين ان علاقة المتغيرات التوضيحية السبعة التي تمثل خصائص المياه الصالحة للشرب بمتغير اعداد الاصابة كما يأتي: الأول (PH) والثاني (5BOD) والثالث (T.H) والسابع (4PO) هي علاقة طردية، بينما ترتبط المتغيرات التوضيحية الرابع (Ca) والخامس (Mg) والسادس (4SO) بعلاقة عكسية مع المتغير المعتمد الذي يمثل اعداد الاصابة بالمرض.

5. فيما يخص بيانات جانب الرصافة فان انموذج بيانات البانل ذو التأثيرات العشوائية المعلمي يشير الى ان علاقة المتغيرات التوضيحية السبعة التي تمثل خصائص المياه الصالحة للشرب بمتغير اعداد الاصابة كما يأتي: الأول (PH) والثالث (T.H) والخامس (Mg) والسابع (4PO) هي علاقة طردية، بينما ترتبط المتغيرات التوضيحية الثاني (5BOD) والرابع (Ca) والسادس (4SO) بعلاقة عكسية مع المتغير المعتمد الذي يمثل اعداد الاصابة.

6. من خلال قيم معايير المقارنة ومنحنيات التمهيد تبين ان الانموذج الأكثر كفاءة لوصف بيانات جانب الكرخ لأعداد الاصابة هو الانموذج اللامعلمي المقدر وفق طريقة ناداريا واتسون الموزونة بالدالة اللبية (Gaussian)، بينما كان الانموذج الاكفاً لبيانات جانب

الرصافة المقدر هو الانموذج المقدر وفق طريقة الانحدار الموضوعي الخطي الموزون بالدالة اللبية (Gaussian)، وذلك بسبب عدم تجانس التباينات بين المقاطع.

7. حسب قيم معاملات التحديد اللامعلمية لبيانات جانب الكرخ فان المتغيرات التوضيحية الممهدة وفق الانموذج اللامعلمي المقدر بطريقة ندارايا واتسون الموزونة بالدالة اللبية (Gaussian) تفسر- (99.86%) من التغيرات الحاصلة في المتغير المعتمد الممهد لأعداد الاصابة ، في حين فسرت المتغيرات التوضيحية الممهدة بواسطة الانحدار الموضوعي الخطي المتعدد الموزون بالدالة اللبية (Epanchincov) ما نسبته (98.71%) من التغيرات الطارئة على متغير اعداد الاصابة ، اما المتغيرات التوضيحية الممهدة وفق الانموذج اللامعلمي المقدر بطريقة ندارايا واتسون الموزونة بالدالة اللبية (Epanchincov) فقد فسرت التغيرات الحاصلة في اعداد الاصابة بنسبة (98.51%)، بينما فسرت المتغيرات التوضيحية الممهدة وفق الانحدار الموضوعي الخطي المتعدد الموزون بدالة اللب (Gaussian) التغيرات الحاصلة في متغير اعداد الاصابة بنسبة (88.82%).

8. اما فيما يخص بيانات جانب الرصافة فقد فسرت المتغيرات التوضيحية الممهدة وفق الانحدار الموضوعي الخطي المتعدد بدالة اللب (Gaussian) ما نسبته (99.44%) من التغيرات الحاصلة في متغير اعداد الاصابة.

9. مما تقدم يتضح ان اكفاً انموذج يفسر- طبيعة العلاقة بين اعداد الاصابة بمرض الاميبا وخصائص المياه هو الانموذج المقدر بطريقة ندارايا واتسون الموزونة بالدالة اللبية (Gaussian) لجانب الكرخ.

10. تتفاوت في بعض الأحيان نتائج المقارنة بين المقدرات باختلاف معايير المقارنة، اذ كان معيار المقارنة (AMSE) اكثر ثباتاً ودقة من معايير المقارنة ((MAPE، AMAE)، (AMADE))، فمنحنيات التمهيد لبيانات اعداد الاصابة لجانب الكرخ والرصافة تطابق في نتائجها النتائج التي تم الحصول عليها من خلال معدل متوسط مربعات الخطأ.

التوصيات

1-دراسة انموذج التأثير الثابت اللامعلمي لبيانات بانل ومقارنته مع انموذج التأثير الثابت المعلمي لبيانات بانل .

2-استعمال طرق تقدير اخرى لامعلمية ومقارنتها مع هذه الطرق لغرض الوصول للأنموذج الأكثر جودة الذي يمثل بيانات بانل فضلاً عن استعمال طرق تمهيد ماعدا الطرق اللبية للنماذج اللامعلمية.

3-ضرورة العمل على توفير كادر متقدم مختص بأجراء الفحوصات المختبرية لعينات المياه وحساب تراكيز الملوثات عن طريق ارسالهم لبعثات ودورات خارج العراق ليتمكنوا من ادارة الاجهزة المتوفرة لديهم الخاصة بالتحليلات المختبرية فضلاً عن ذلك تجهيز مختبرات الصحة والبيئة بأجهزة متطورة تخص هذا المجال وكذلك توفير الحماية لهم

عند الخروج للعمل الميداني لغرض سحب العينات من مياه الانهر لما قد يتعرضون له عند الذهاب بشكل منفرد.

4- ضرورة قيام الدولة المتمثلة بالحكومة المركزية والحكومات المحلية بتعميم قرارات تمنع فيها مؤسساتها الحكومية اولا من رمي المخلفات السائلة والصلبة في نهر دجلة واعتباره جريمة يحاسب عليها القانون لما يتسبب من ضرر كبير في المجال الصحي والبيئي وتسببه بأمراض خطيرة واثنا المؤسسات والشركات والمصانع الاهلية بمحاسبته قانونيا ويعتبر هذا موازي للضريبة السنوية لهذه الجهات و حسب المحددات البيئية لنظام صيانة الانهار رقم (25) لسنة (1967) في العراق .

5- ضرورة اجراء بحوث احصائية اخرى تهتم بأمراض اخرى خطيرة على حياة الانسان تسببها هذه الملوثات مثل الامراض السرطانية وغيرها مع توعية الناس بخطورتها ومعرفة ماهية التراكيز الملوثة المسببة لهذه الامراض.

المصادر العربية

1- التميمي، سهاد علي شهيد ، (2016)، " دراسة مقارنة لبعض طرائق تقدير النموذج الديناميكي المكاني الخاصة بـ (Panel data) مع تطبيق عملي "، اطروحة دكتوراه مقدمة الى كلية الادارة والاقتصاد – الجامعة المستنصرية.

2- الجمال، زكريا يحيى ، (2012)، " اختبار النموذج في النماذج البيانات الطولية الثابتة و العشوائية "، المجلة العراقية للعلوم الاحصائية ، العدد (21) ، ص ص {266 – 285}.

3- الدليمي ،ناظم عبد الله،(1994)، " اساليب دمج السلاسل الزمنية والبيانات المقطعية في تحليل بعض الظواهر الاقتصادية " اطروحة دكتوراه في الاحصاء ، كلية الادارة والاقتصاد، جامعة بغداد.

4- الزيدي، ازهار سلمان زامل،(1986)، " استخدام اسلوب الدمج بين السلاسل الزمنية والبيانات المقطعية لتحليل سلوك المستهلك العراقي في ضوء بحث ميزانية الاسرة لعام 1984 " رسالة ماجستير في الاحصاء، كلية الادارة والاقتصاد، جامعة بغداد.

5- القيسي، باسم شليبه ،(2009)، " التحليل البيزي لنماذج الانحدار الخاصة بالبيانات المزدوجة (paneldata) "، اطروحة في الاحصاء ، كلية الادارة والاقتصاد ، جامعة بغداد.

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

7- محمد ، رتيعة ، (2013) ، "استخدام نماذج بيانات البانل في تقدير دالة النمو الاقتصادي للدول العربية "، المجلة الجزائرية للاقتصاد والمالية ، العدد (2) ، ص {152 – 170} .

المصادر الاجنبية

- 8- Baltagi ,badi, (2006) "Forecasting with panel data"Econometric studies, series1,NO.25.
- 9- Cai, Z. & Li, Q., (2008), "Nonparametric Estimation of Varying Coefficient Dynamic Panel Data Models", Econometric theory, Vol. 24, pp, (1321-1342) , printed in the united states of America .
- 10-Chen , J., Li,D. & Gao,J. ,(2013) , " Non- and Semi-Parametric Panel Data Models: A Selective Review" , Oxford University Press.
- 11-Filho , C.M. and Yao , F., (2008), " Non parametric regression Estimation with general parametric error covariance ", AMS 2000 , subject classificat-ions.62GO8,62G20.
- 12- Hajargasht, C , (2009) , " Nonparametric Panel Data Models A Penalized Spline Approach " , Marcel Dekker, Inc. New York.
- 13-Hsiao, C., and M. H. Pesaran, (2004) , "Random Coefficient Panel Data Models," IZA Discussion Papers 1236, Institute for the Study of Labor (IZA).
- 14-Huang, J. Z., Wu, C. O. and Zhou, L., (2004), "Polynomial Spline estimation and Inference For Varying Coefficient Models With Longitudinal Data", statistica Sinica, vol. 14, pp. 763-788.
- 15-Lee, Y. and Mukherjee , D . ,(2014) , " Nonparametric Estimation of the Marginal Effect in Fixed-Effect Panel Data Models: An Application on the Environmental Kuznets Curve " ,Vol.11, NO. 4, PP735–757.
- 16-Park , H.M,(2005),” Linear regression models for panel data using SAS STATS,LIMDEP and SSPS” <http://www.indiana.edu/~stat.math>.
- 17-Porter, J. ,(1996) , “Essays in Econometrics,”Ph.D. thesis, Uuiversity of California.
- 18-Poo , J.M.R and sobron , A., (2016) , "Non parametric and semi Parametric panel data models : recent developments " journal of economic surveys , vol.00 , No. 0 , pp . 1-37, john wiley and sons Ltd, 9600 Garsington road . oxford ox4 2Dce , UK and 350 main street , Molden . MA 02148 USA
- 19-SU,L. and ULLAH , A. , (2011) , " Nonparametric and semi parametric panel data Econometric models ‘Estimation and Testing " , Handbook of Empirical Economics and Finance , (pp.455-497). Research collection school of Economics , Available at: <https://ink.library.smv.edu.sg/soe-research/1281>.
- 20-Su , L. and LU, X. , (2013) , "Non parametric Dynamic panel data models : kernel Estimation and specification testing " , journal of

Econometrics , No.176 , vol.2 , pp (112-133).Research collection schod of economics . Available at: https://ink.library.smu.edu.sg/soe_research/.
21-Wang ,Y. , (2012), " Essays on Non parametric and semi parametric model and continuous time models , A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of philosophy in Economic , university of California riverside .