(مقارنة طريقة تقدير المربعات الصغرى الموضعية الخطية مع طريقة المربعات الصغرى الموضعية الخطية الموزونة للنماذج اللامعلمية للبيانات الطولية)

عادل ماجد حسن
الجامعة المستنصرية /كلية الإدارة والاقتصاد

أ .م .د عدي طه رحيم
الجامعة المستنصرية /كلية الإدارة والاقتصاد

تاريخ القبول: 2021/3/16 تاريخ النشر:2022/1/30

الملخص:[1]

يتلوث الماء نتيجة وجود مواد غريبة فيه سواء كانت عضوية او غير عضوية , تعمل هذه المواد على افساد نوعية المياه وهو مايحصل لنهر دجلة . وتتسبب هذه الفضلات بتلوث مياه نهر دجلة بملوثات كيميائية خطيرة تضر بصحة الانسان مسببة له الامراض , ومن هذه الامراض مرض الاميبا التي اخذت بياناته الممثلة باعداد الاصابات بالمرض لجانبي الكرخ والرصافة في بغداد لعام (2018) كمتغير معتمد (Y) بسلسلة كاملة (12) شهرا", ولمعرفة تاثير تراكيز ملوثات الماء على عدد الاصابات بمرض الاميبا تم الاعتماد على سبعة تراكيز ملوثة وهي تمثل المتغيرات التوضيحية (Xi) وهي تمثل بيانات مقطعية لعشرة محطات واقعة على ضفتي نهر دجلة لجانبي الكرخ والرصافة. وبسبب التعامل مع المقاطع العرضية والسلسلة الزمنية فان البيانات الطولية او ماتسمى بيانات بانل هي افضل تسمية لها , استعملنا نماذج بيانات بانل اللامعلمية بمقدرين لامعلميين وهي المقدر الخطي الموضعي والمقدر الخطي الموضعي الموزون لتحديد اي من المقدرين هو الاكفأ الذي يعطي افضل انموذج,وكانت عدد الاصابات في المستشفيات والمراكز الصحية في جانبي الكرخ والرصافة تعاني من مشكلة عدم تجانس التباين وان بيانات المتغيرات لتراكيز الملوثات السبعة تعاني من مشكلة الارتباط الذاتي.

1-المقدمة

ان النماذج المقدرة لبيانات بانل (*panel data*) تكون تقديراتها اكثر كفاءة ومعنوية من التقديرات التي تحصل عليها باستعمال بيانات السلاسل الزمنية فقط او بيانات المقاطع العرضية فقط.

تمتاز بيانات بانل باعطائها صورة اوضح للظاهرة قيد الدراسة من تلك التي تعتمد على نوع واحد من البيانات ويمكن تمثيل بيانات بانل بنماذج لامعلمية كانموذج التاثير الثابت و العشوائي اللامعلمي وايضا توجد مشكلة الارتباط الذاتي في هذه النماذج اللامعلمية وكذلك مشكلة عدم تجانس التباين وهنالك طرق للمعالجة للتخلص من هذه المشاكل ثم تقدير الانموذج وتحليله,وتناولت الدراسات السابقة المقارنة بين طرق التقدير, ففي عام 2009 قدم الباحثان (*Xu & Zhu*)[21] اختبار حسن المطابقة لانموذج المعاملات المتغيرة زمنيا للبيانات الطولية وفي عام (2016) قدم الباحثان(*Vogt & Linton*)[22] بحثا تناولا تم فيه نماذج البيانات المزدوجة اللامعلمية مع دوال انحدار غير متجانسة لمجموعة متنوعة من التطبيقات.

2- الجانب النظري

(2-1) بيانات بانل

تعرف هذه البيانات بانها قياسات متكررة (*Repeated Measurement*) تاخذ 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 [5][6] , تقرا هذه البيانات لظاهرة معينة على اسلوبين الاول قراءة فترة من فترات السلسلة الزمنية لكل المقاطع العرضية , اما الاسلوب الثاني فتقرأ بيانات مقطع من المقاطع العرضية لكل فترات هذه السلسلة الزمنية , كذلك فانه عند تسجيل هذه البيانات لظاهرة معينة فان هذه الظاهرة ستتغير على مستويين المستوى الاول التغير على المستوى الافقي (العرض) فتكون البيانات عبارة عن بيانات مقطعية *(cross –sections data)* اما المستوى الثاني فهو يمثل التغير على المستوى العمودي (الطول) وعندها تكون البيانات عبارة عن بيانات السلسلة الزمنية (*time serial data*) , ان عملية دمج البيانات المقطعية مع الفترات الزمنية جعل نماذج البيانات الطولية تتمع بالعديد من المزايا التي وضعها الباحث (*Blatagi) (2005*)[4] منها التحكم في عدم تجانس التباين الخاص الذي يظهر في حالة البيانات المقطعية او في حالة البيانات الزمنية وكون البيانات الطولية تعطي الكفاءة الافضل وزيادة في درجات الحرية واقل تعددية

[1] بحث مستل من رسالة ماجستير

خطية بين المتغيرات ومحتوى معلوماتي اكثر اذا ماتم استعمال البيانات المقطعية او البيانات الزمنية,ويمكن التعبير عن هذه البيانات كما في الجدول (*1*) اذ ان العدد الكلي لجميع المشاهدات المعتمدة هو($N=n*T$) وبوجود k من المتغيرات التوضيحية وكما في الجدول ادناه وهو تصميم لمجموعة من البيانات الطولية وكما يأتي:

لجدول (1) تصميم لمجموعة من بيانات بانـــــــل

Subject	Repeated measurement	y_{ij}	χ_{ijr}
1	1	y_{11}	$\chi_{111},\cdots\cdots,\chi_{11d}$
1	2	y_{12}	$.\chi_{121},\cdots\cdots,\chi_{12d}$
1	t_1	y_{1t_1}	$\chi_{1t_11},\cdots\cdots,\chi_{1t_1d}$
2	1	y_{21}	$\chi_{211},\cdots\cdots,\chi_{21d}$
.	2	y_{22}	$\chi_{221},\cdots\cdots,\chi_{22d}$
2	t_2	y_{2t_2}	$\chi_{2t_21},\cdots\cdots,\chi_{2t_2d}$
n	1	y_{n1}	$\chi_{n11},\cdots\cdots,\chi_{n1d}$
.	2	y_{n2}	$\chi_{n21},\cdots\cdots,\chi_{n2d}$
n	t_n	y_{nt_n}	$\chi_{nt_n1},\cdots\cdots,\chi_{nt_nd}$

(2-2) النماذج اللامعلمية للبيانات الطولية (nonparametric panel data models of longitudinal data)

انموذج التأثير العشوائي اللامعلمي للبيانات الطولية *(nonparametric panel data models with random effects)* تكون معلمات انموذج التأثيرات العشوائية غير متحيزة وعادة مايفرض بان تباين الخطا غير ثابت (غير متجانس) لجميع المشاهدات المقطعية وهناك ارتباط ذاتي خلال الزمن بين كل مجموعة من مجاميع المشاهدات المقطعية في فترة زمنية محددة , وان حد الخطا $u_{it} \sim (0,\sigma_\varepsilon^2)$ ، والانموذج يكون بالصيغة الآتية[5] :

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

اذ ان :y_{it}: تمثل قيمة متغير الاستجابة في المشاهدة i عند الفترة الزمنية t، وهو موجه صفي من الدرجة $(nT*1)$. تمثل قيمة المتغير التوضيحي (التفسيري) في المشاهدة i عند الفترة الزمنية t ، وهو موجه صفي من الدرجة $(nT*1)$ ، $x_{it} \in R$ و $m(.)$: دالة التمهيد غير المعرفة ويمكن تقديرها بطرائق التقدير اللامعلمية ، $\mathrm{E}\left[Y_{it} \mid X_{it}=x\right]=m(x)$ ، V_i : التأثيرات العشوائية المفردة ، $v_i=\left(v_1 l_T', v_2 l_T', \ldots\ldots\ldots\ldots, v_n l_T'\right)'$ ، I_T متجه الوحدة بدرجة $(T*1)$ ، ويمثل متغير عشوائيا مستقلا ومتماثل التوزيع $(i.i.d)$ (0 ، σ_v^2) ، وهو حد الخطا في مجموعة البيانات المقطعية .

u_{it} : تمثل قيمة الخطا في المشاهدة i عند الفترة الزمنية t ، وهو متجه صفي من الدرجة $(nT*1)$ ويكون متغيرا عشوائيا مستقلا ومتماثل التوزيع $(i.i.d)$ $(0 \quad , \quad \sigma^2_{\varepsilon})$ ، لكل قيم x_{jt} ، $i=1,2,...,n$ ، $t=1,2,....T$ ويحقق الشرط $E(\varepsilon_{it} \mid x_{it})=0$ ،

$$Var\,(\varepsilon^2_{it} \mid x_{it}) = \sigma^2(x_{it})$$

.

ويطلق على انموذج التأثيرات العشوائية احيانا بانموذج مركبات الخطا *(Error Components Model)* بسبب ان النموذج في الصيغة *(1)* يحتوي مركبتين للخطا في الانموذج

(يكون حد الخطا $\varepsilon_{it} = v_i + u_{it}$, $i=1,2,....,n$, $t=1,2,.....,T$) ، ومن ثم يمكن كتابة الانموذج بالصيغة التالية :

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

وتكتب بشكل مصفوفات كالآتي :

$$Y = l_{nT}\, m(x) + u$$

(2-3) تقدير معلمات النماذج اللامعلمية للبيانات الطولية

تعاني نماذج الانحدار اللامعلمي من مشكلة وجود الارتباط الذاتي ومن مشكلة عدم التجانس , وتوجد عدة طرق لمعالجة هاتين المشكلتين ولكن جميعها تعتمد على تقدير الدالة اللبية والتي تعتمد وتتاثر بعرض الحزمة لان اختيار عرض الحزمة يؤثر بوجود مشكلة الارتباط الذاتي وايضا على وجود مشكلة عدم التجانس لتنقية بيانات العينة من اثر هاتين المشكلتين وقبل ذكر طرق التقدير للنماذج اللامعلمية سيتم عرض الدوال اللبية *(kernel function)* والتي تعتبر من أكثر الدوال الوزنية استعمالا وتعتمد على تهميد البيانات للمتغير المعتمد Y وتمتلك الدالة اللبية الخواص التالية[1][2]:

- الاستمرارية $0 \le K(u) \le \infty$ التماثل $K(u) = K(-u)$

دالة كثافة احتمالية $\int K(u)du = 1$ وتملك تباين معلوم ومتوسط يساوي صفر $Eu = o$

$$\int u\,K(u)du = 0 \qquad \int u^2\,K(u)du = \sigma^2{}_k \qquad E(u^2) =$$

وللدالة اللبية عدة صيغ مختلفة وحسب عدد المتغيرات التوضيحية ففي حالة المتغير الواحد فان الصيغ موضحة[3] في الجدول *(2)* :-

الجدول (2) يبين بعض انواع دوال كيرنل لمتغير واحد

النوع	الدالة اللبية	صيغة الدالة
1	*Uniform Kernel*	$K(u)=\begin{cases}\frac{1}{2} & |u|<1\\ \square & \\ 0 & o.w\end{cases}$
2	*Triangular Kern*	$K(u)=\begin{cases}1-|u| & \sin(u)<1\\ 0 & o.w\end{cases}$
3	*Gaussian Kernel*	$K(u)=\begin{cases}(2\pi)^{\frac{-1}{2}}\,e^{-\frac{u^2}{2}} & 0<u<\infty\\ \square & \\ 0 & o.w\end{cases}$
4	*Epanchnikov ker*	$K(u)=\begin{cases}\frac{3}{4}(1-u^2) & |u|<1\\ 0 & o.w\end{cases}$
5	*Bi Weight Kernel*	$K(u)=\begin{cases}\frac{15}{16}(1-u^2)^2 & |u|<1\\ 0 & o.w\end{cases}$
6	*Tri Weight kerne*	$K(u)=\begin{cases}\frac{35}{32}(1-u^2)^3 & |u|<1\\ 0 & o.w\end{cases}$

$$K(u) = C_a(1-u^2)^a \qquad \text{اذ ان:}$$

$$Ca = \Gamma(2a+2)\,[\Gamma(a+1)]^2 2^{-2a-1}$$

$a = (1,2,3 \dots \dots)$ تمثل درجة الدالة المطلوبة

قد اتجه اغلب الباحثين الى تطبيق دالة $Gaussian$, وعائلة $Epanchnikov$ في مقدرات الانحدار اللامعلمي .

اما في حالة وجود متغيرين فان صيغ الدالة اللبية $(Epanchnikov,Gaussian)$ تكون حسب الجدول (3) .

الجدول (3) يبين بعض انواع دوال كيرنل لمتغيرين

النوع	الدالة اللبية	صيغة الدالة
1	**Bivariate Gaussian**	$K(u_1,u_2) = \begin{cases} (2\pi)^{-1}\, e^{-\frac{{u_1}^2+{u_2}^2}{2}} & 0<u_1,u_2<\infty \\ 0 & o.w \end{cases}$
2	**Bivariate Epanchnikov**	$K(u_1,u_2) = \begin{cases} (2\pi)^{-1}\,(1-{u_1}^2-{u_2}^2), & {u_1}^2+{u_2}^2< \\ 0 & o.w \end{cases}$

وفي حالة متعدد المتغيرات فان صيغة الدالة اللبية $(Epanchnikov,Gaussian)$ [11]تكون بحسب الجدول (4):

الجدول (4) يبين بعض انواع دوال كيرنل في حالة متعدد المتغيرات

$K(u_1,u_2 \dots,u_n) = \begin{cases} (2\pi)^{\frac{-d}{2}}\, e^{-\frac{u^T u}{2}} & 0<u_1,u_2<\infty \\ 0 & o.w \end{cases}$	**Multivarite Gaussian**	1
$K(u_1,u_2 \dots,u_n) = \begin{cases} (2c_d)^{-1}\,(d+2)(I-u^T u) & u^T u<1 \\ 0 & o.w \end{cases}$ (c_d)يكون بالقيم: (c_1=2 ,c_2=π ,c_3=4π/3) اذ ان	**Multivarite Epanchnikov**	2

وعند تطبيق الدوال اللبية في الجداول (2) , (3) , (4) على البيانات لغرض التمهيد فان المتغير u سيكون بالصيغة :-

$$\frac{Z_i - Z_0}{h}\,u \qquad i=1,2,\dots,n \qquad \dots(3)$$

اذ ان : Z_i : يمثل قياسات المتغير التوضيحي و Z_o : يمثل مشاهدة من مشاهدات المتغير التوضيحي

ويشار اليها في مصادر h : اذ يمثل عرض الحزمة (The bandwidth) عديدة باسم معلمة التمهيد (Smoothing Parameter) , وهي الجزء الاهم في الدوال اللبية , ان اختيار (h) هو الجزء المهم في عملية التمهيد , و

هناك عدة اساليب في اختيار قيمة (h) عرض الحزمة منها:- الاسلوب التجريبي واسلوب استعمال طرق تقدير المعلمة (h) ومنها طريقة العبور الشرعي ($(Cross\ Validation$) بالاعتماد على بعض الصيغ الرياضية لتقديرها.

معيار تقاطع الشرعية (العبور الشرعي) **(Leave–One-Out Cross- Validation) (CV)**ان اختيار عرض الحزمة الملائمة (المعلمة التمهيدية) ، مهم لتقريب دالة الانحدار اللامعلمي للدالة الاصلية بحيث يكون الخطا اقل مايمكن وهو دائما نقطة البداية الجيدة ، وان دقة التمهيد تعتمد فقط على عرض الحزمة ، وان الموازنة بين التحيز والتباين تسمى معلمة التمهيد او معلمة عرض الحزمة[1] .
ولكي نتجنب اختيار قيم المعالم بواسطة التجزئة والخطأ ويكون بالصيغة الآتية [3] :

$$LCV(h) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\hat{m}_{(-1)}(x_{it}) - y_{it}\right)^2 \quad ...(4)$$

اذ ان $\hat{m}_{(-1)}(x_{it})$: دالة التقدير التمهيدية لجميع الطرائق اللامعلمية الاعتيادية لانموذج التأثير العشوائي اللامعلمي للبيانات الطولية ، ومن ثم فان قيمة (h) هي التي تقابل اقل$(C\ V)$.

(1-3-2) طريقة مقدر المربعات الصغرى الموضعية الخطية

(Local linear Least squares estimator) LILS

بالاعتماد على متسلسلة تيلر من الدرجة الاولى فتكون كالآتي[20][11]:

$$Yij = m\,i\,(xij) + uij \quad ...(5)$$

$$\approx mi(xi) + (xij - xi)\,mi(1)(xi) + uij \quad (6)$$

$$= (1 \quad (xij - x\,i)') \begin{bmatrix} mi(xi) \\ \vdots \\ mi(1)\ (xi) \end{bmatrix} + uij \quad ...(7)$$

$$yij = zij\,(xi)\ \delta i(xi) + uij \quad ...(8)$$

اذ ان :

$\delta i(xi)$ متجه من درجة $[(qi+1)*1]$

$$\delta i(xi) = (\,mi(xi) \quad mi(1)\,(xi)\,)' \quad ...(9)$$

$$zij\,(xi) = (\,1 \quad (xij - x\,i)'\,)$$

$$zi\,(xi) = (zi1\,(xi), \ \ldots, \ ziN\,(xi))'$$

ولها ابعاد من درجة $N * (qi+1)$

واما ui فهو يمثل $(u_{ij}) = (ui1\ ,\ \ldots,\ uIN)'$

ومن هذه المتجهات لكل انحدار (i) يمكن ان تكتب :

$$Yi \approx zi\,(xi)\,\delta i(xi) + ui \quad ...(10)$$

وبصيغة المصفوفات فان

$$\begin{bmatrix} Y1 \\ \vdots \\ Ym \end{bmatrix} = \begin{bmatrix} Z1(X1) \ldots & 0 \\ \vdots & \\ 0 & Zm(Xm) \end{bmatrix} \begin{bmatrix} \delta 1(X1) \\ \vdots \\ \delta m(Xm) \end{bmatrix} + \begin{bmatrix} U1 \\ \vdots \\ Um \end{bmatrix} \quad ...(11)$$

والذي يمكن كتابتها بالصيغة الآتية:$y = m_{(x)} + u$

$$y \approx z(x)\,\delta(x) + u \quad ...(12)$$

اذ ان ($y = (y'1, ..., y'm$) وهو متجه من درجة $(mN \times 1)$

وهو يمثل $\delta i(xi) = (\delta 1(x1), ..., \delta m(xm))$ ووفقا للمعادلة (12) فان

$$E(u / x) = 0 \qquad mxn$$

وان $\Omega = Var(u / x) = \sum * IN$

اذ ان $\sum$ هي مصفوفة من درجة $(m * m)$ عناصر قطرها تمثل σ_{ii}^2 اما بقية العناصر خارج القطر فتمثل σ'_{ii}

اذ ان $i, i' = 1, ..., m$ وبذلك فان المقدر الموضعي الخطي $\delta(x)$ يمكن ان نحصل عليه عن طريق تصغير الحد $(u'k(x)u)$ وحسب الصيغة الآتية:

$$\hat{\delta}(x) = (z'(x)\,k(x)\,z(x))^{-1}\,z'(x)\,k(x)\,y \quad ...(13)$$

اذ ان :$k(x)$: مصفوفة قطرية من درجة $mN * mN$

وهي:$K(x) = diag(kh1(X1 - x1), ..., khm(Xm - xm))$

وان $khi(Xi - xi)$ هي ايضا مصفوفة قطرية وهي :

$$khi(Xi - xi) = diag(khi(Xi1 - x1), ..., khi(XiN - xi))$$

$$khi(Xi - xi) = (1/hi)\,(xij - xi/hi)$$

وان $u = (u'1, ..., u'm)$ علما ان $Z(x) = diag(Z1(x), ... Zm(x_m))$

وان Z(x) ابعادها هي[mN $* \sum_{i=1}^{m} a\,i$ + m]

اما $\delta(x)$ فهو متجه من درجة [1 $* \sum_{i=1}^{m} a\,i$ + m)](وهو يمثل

$$\delta(x) = (\delta 1(x1), ..., \delta m(xm))$$

علما بان تقدير $\hat{\delta}(x)$ هو يمثل المقدر m^(x) الذي يمثل التقدير m(x) دالة التمهيد.

(2-3-2) طريقة مقدر المربعات الصغرى الخطية الموضعية الموزونة

(Aweighted local linear least squares estimator) (AWLLLS)

لتوضيح عمل المقدر $\delta(x)$وصيغة لايجاد تقدير انموذج التأثيرات العشوائية اللامعلمية للبيانات الطولية نستعمل الصيغة التالية[8][11][20]:

وعند استعمال توسيع تايلور فان:

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

$$y_{it} = z_{it}\,\delta(x) + u_{it} \quad ...(14)$$

ويمكن كتابة (14) بصيغة المصفوفات وكالآتي :

$$Y = Z(x)\,\delta(x) + u \quad ...(15)$$

اذ ان : $Y = \left(y'_{11}, y'_{21}, ..., y'_{nT}\right)'$ متجه بدرجة (nT* 1) .

$Z(x) = \left(Z'_{11}, Z'_{21}, ..., Z'_{nT}\right)'$ وهـــو متجـــه بدرجـــة (nT* 1) .

$\varepsilon = \left(\varepsilon'_{11}, \varepsilon'_{21}, ..., \varepsilon'_{nt}\right)'$ وهو متجه الاخطاء بدرجة (nT* 1).

$\delta(x)$: تشير الى دالة تمهيدية بقيمة واحدة .

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

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

$$\Omega=I_n\otimes V$$

$$l_T\, l_T'=J_T$$

$$\Omega^{-1}=I_n\otimes V^{-1} \qquad ...(16)$$

$$\lambda=\sigma_u^2/\left(\sigma_u^2+T\sigma_v^2\right) \qquad (n * n)$$: مصفوفة لعناصر الوحدة بدرجة I_n

اما V فهي تمثل :

(17) وبتعويض المعادلة (14) بعد اخذ الجذر التربيعي للطرفين سنحصل على $V^{-\frac{1}{2}}=\left[\, I_T-(1-\lambda^{-\frac{1}{2}})\, l_T\, l_T'/T\,\right]$

$$\Omega^{-1/2}=\mathrm{In}\otimes V^{-1/2}$$

وبتعويض قيمة $V^{-1/2}$ في المعادلة (14) تم تعويض المعادلة (16) في المعادلة (17)

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

وبضرب طرفي المعادلة (15) بـ $\Omega^{-1/2}$ سنحصل على: $\Omega^{-\frac{1}{2}}Y=\Omega^{-\frac{1}{2}}Z(x)\,\delta(x)+\Omega^{-\frac{1}{2}}u$

وبجعل $\sum^{*}=\Omega^{-1/2}\sum$ و $Z^{*}(x)=\Omega^{-1/2}Z(x)$ و $y_{*}=\Omega^{-1/2}y$ وينتج لنا الصيغة التالية:

$$Y^{*}=Z^{*}(x)\,\delta(x)+u^{*} \qquad ...(18)$$

اذ ان المعادلة (18) تحقق الشروط التالية :

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

$$y^{*}_{it}=z^{*}_{it}\;\delta(x)+u^{*}_{it}$$

وبالاعتماد على الصيغة (18) و باستعمال طريقة المربعات الصغرى الموزونة نحصل على :

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

وبتعويض

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

وبصيغة المصفوفات واجراء التبسيط ينتج:

$$=(Y^{*}-Z^{*}(x)\,\delta(x))'\,K(x)\,(Y^{*}-Z^{*}(x)\,\delta(x))$$

وباستعمال

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

$$\lambda = \frac{\sigma_u^2}{\left(\sigma_u^2 + T\sigma_v^2\right)} \qquad y^*_{it} = y_{it} - (1-\lambda^{\frac{1}{2}})\,\bar{y}i. \quad ، \quad z^*_{it} = z_{it} - (1-\lambda^{\frac{1}{2}})\,\bar{z}i$$

ومن ثم نحصل على مقدر $\hat{\delta}(x)$ وكالاتي :

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

وبصيغة المصفوفات فان:

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

وبتعويض قيم y^* وZ^* بما يساويها سينتج:

$$\hat{\delta}(x) = \left(Z'(x)\,\Omega^{-\frac{1}{2}}\, K(x)\Omega^{-\frac{1}{2}}\, Z(x)\right)^{-1}.\left(Z'(x)\,\Omega^{-\frac{1}{2}}\, K(x)\,\Omega^{-\frac{1}{2}}\, Y\right) \quad ...(19)$$

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

(4-2) تشخيص الانموذج اللامعلمي للبيانات الطولية

(1-4-2) الاختبار اللامعلمي للكشف عن مشكلة الارتباط الذاتي

(Non-parametric test for the detection of the problem of auto correlation)

المشكلة تكمن في ايجاد اختبار لامعلمي للكشف عن مشكلة الارتباط الذاتي ، وذلك من خلال استعمال الاختبار المعطى من قبل (Su and Lu) (2013)[12] وهو تطوير للاختبار المعطى من قبل (Henderson et al) (2008)[5] ، و لاختبار الفرضية الآتية[8] :[20]

$$H_0 : E(u_{it}\, u_{i,t-1}) = 0$$
$$H_1 : E(u_{it}\, u_{i,t-1}) \neq 0$$

... (20)

فان احصاءة الاختبار (I_N) تحسب كالآتي:

$$\mathbf{I}_N = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} \hat{\boldsymbol{u}}_{it}\, \hat{\boldsymbol{u}}_{i,t-1} \qquad ... \qquad (21)$$

$$\varepsilon_{it} = \left(u_{it} - u_{i,t-1}\right)$$
$$\hat{\varepsilon}_{it} = \left(y_{it} - y_{i,t-1}\right) - \left(\hat{m}(x_{it}) - \hat{m}(x_{i,t-1})\right)$$

فاذا كانت قيمة الاحصاءة اكبر من قيمة χ^2 المجدولة بدرجة حرية k (عدد المتغيرات التوضيحية) نقبل H_1 وهذا يعني ان الانموذج يعاني من مشكلة الارتباط الذاتي وبعكسه نقبل الفرضية H_0[11].

(2-4-2)الاختبار اللامعلمي لكشف مشكلة عدم تجانس التباين للخطا :

توجد العديد من الاختبارات المعلمية المستعملة **(Nonparametric test for the detection of the heterogeneity problem)** للكشف عن مشكلة عدم التجانس ، وللكشف عن مشكلة عدم التجانس ، ولاختبار فرضية العدم $\left(H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_n^2\right)$ ضد الفرضية البديلة $\left(H_1 : \sigma_1^2 \neq \sigma_2^2 \neq ... \neq \sigma_n^2\right)$ على الاقل واحد من فرض ثبات تجانس التباين[5] [12]

$$\frac{1}{(n-1)\sum_{i=1}^{n} R_{i,n}^2} \quad ... (22)$$

$R_{i,n}^2 = \frac{1}{2}\left(Y_{i,n} - Y_{i-1,n}\right)^2$ ، تحت فرضية تحقق $E\left[R_{i,n}^2\right] \approx \sigma^2(x_i, n)$ ، ولاختبار الفرضية $\mathrm{E}(uu') = \sigma_u^2 I_n$) وقد قام العديد من الباحثين بتطوير الاختبار فتم التوصل للصيغة النهائية[12] وكالاتي:

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

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

...(23)

اذ ان :( )K: الدوال اللبية (كيرنل) الموزونة,h: عرض الحزمة (المعلمة التمهيدية) .

$$\overline{R_n^2} = \frac{1}{n-1} \sum_{i=2}^{n} R_{i,n}^2 \quad ، \quad R_{j,n}^2 = \frac{1}{2}\left(Y_{j,n} - Y_{j-1,n}\right)^2$$

ويعتمد القرار النهائي على ايجاد قيمة احصاءة الاختبار فاذا كانت قيمة الاحصاءة اكبر من قيمة χ^2 المجدولة لدرجة حرية k (عدد المتغيرات التوضيحية) وفيها نقبل H_0 وهذا يعني ان الانموذج يعاني من مشكلة عدم تجانس التباين وبعكسه نقبل الفرضية H_1 [11].

(5-2) اختبار الانموذج الملائم

اولا: اختبار مضاعف لاكرانج (Lagrange multiplier Test)

اقتراح كل من *(Breusch-Pagan 1980)* [21] يعتمد على احصاءة تتبع توزيع χ^2 بدرجة حرية واحدة وذلك للمفاضلة بين الانموذج التجميعي والانموذجين الثابت والعشوائي ، وعرف هذا الاختبار بمضاعف لاكرانج *(Lagrange multiplier Test)(LM)*، ويعتمد على الاخطاء المقدرة بطريقة المربعات الصغرى الاعتيادية *(OLS)* لاختبار الفرضية الآتية:

$$H_0: \sigma_u^2 = 0$$

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

انموذج الانحدار التجميعي غير ملائم

فان احصاءة الاختبار للفرضيات هي:

$$LM = \frac{nt}{2(t-1)}\left[\frac{\sum_{i=1}^{n}\left(\sum_{j=1}^{t} u_{ij}\right)^2}{\left(\sum_{i=1}^{n}\sum_{j=1}^{t} u_{ij}\right)^2} - 1\right]^2 \quad ... (24)$$

او $LM = \frac{nt}{2(t-1)}\left[\frac{t^2\,\bar{u}\,\bar{u}'}{uu'} - 1\right]^2 \sim \chi^2_{(1)}$ ... (25) اذ ان:u: متجه بواقي الانموذج التجميعي للبيانات الطولية.

$\overline{u}$: متجه متوسطات المجاميع لبواقي الانموذج التجميعي للبيانات الطولية.

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 انموذجي التأثيرات الثابتة والعشوائية هما الاكثر ملائمة.

كما اقترح *(Honda 1985)* [8]اختبار اخر لمضاعف لاكرانج لتجاوز الشرط الذي يجب توفره وفقا لاسلوب *(Breusch-Pagan)* في الفرضية البديلة وهي ان التباين لايساوي صفر اي بشرط عدم انعدام التباين، فقد قدم *(Honda)* اختبارا احادي الجانب وسهل الاستعمال يتضمن مقارنة الجذر التربيعي لاحصاءة *(Breusch-Pagan)* بقيمة التوزيع الطبيعي القياسي، فان احصاءة اختبار مضاعف لاكرانج باسلوب *(Honda)* تكون:

$$\text{LMH} = \sqrt{\frac{\text{nt}}{2(\text{t}-1)}}\left[\frac{\sum_{\text{i}=1}^{\text{n}}\left(\sum_{\text{j}=1}^{\text{t}}\text{u}_{\text{it}}\right)^2}{\left(\sum_{\text{i}=1}^{\text{n}}\sum_{\text{j}=1}^{\text{t}}\text{u}_{\text{it}}\right)^2}-1\right]^2 \sim \text{N}(0,1) \ldots (26)$$

وذلك لاختبار الفرضية التالية:

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

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

ان رفض فرضية العدم هنا يعني ان الاثار العشوائية بالنسبة للمجاميع معنوية اي ان انموذج الجمع بين البيانات المقطعية والسلاسل الزمنية معنوي، بمعنى اخر فان الانموذج التجميعي غير ملائم للبيانات المدروسة وان انموذجي التأثيرات الثابتة والعشوائية هما الانموذجين الاكثر ملائمة للبيانات المدروسة.

ثانيا: اختبار هاوسمان (Hausman Test) :

وهو اختبار وضع لاختبار فرضية اختيار نوع الانموذج الذي يجب استعماله للبيانات الطولية *(Panel data)* حسب الفرضية الآتية:

Ho: التأثيرات ثابتة *(fixed effects)*

H1: التأثيرات عشوائية *(random effects)*

ويمكن اختبار الفرضية اعلاه باستعمال احصاءة اختبار مربع كاي وكالاتي:

$$X_o^2 = (b_F - b_R)'\Sigma^{-1}(b_F - b_R) \quad \ldots(27)$$

F: مقدار ML لموجه المعلمات للانموذج الثابت, bR: مقدار ML لموجه المعلمات في الانموذج العشوائي.

$$\Sigma^{-1} = \left[(\text{var} \hat{-} \text{cov.}(b_F)) - (\text{var} \hat{-} \text{cov.}(b_R))\right]$$

ويلاحظ خلوها من الحدود الثابتة، وكذلك المتغيرات الوهمية وتقارن χ^2 قيمتها مع قيمة k الجدولية بدرجة حرية فاذا كانت المحتسبة اكبر من الجدولية نقبل الفرضية*H1* وبعكسه نقبل الفرضية*H0.*

(6-2) 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 (Avarage mean square error)

هو عبارة عن معدل متوسط مربعات الاخطاء ورمزه (AMSE) والصيغة الرياضية له هي:

$$\text{AMSE} = \frac{1}{\text{T}}\sum_{i=1}^{T} \text{Mse(i)} \quad \ldots(28)$$

اذ ان Mse(i) تمثل متوسط مربعات الخطا وT: عدد مشاهدات السلسلة الزمنية.

ثانيا":معيار الجذر (N) لمعدل متوسط مربعات الاخطاء(Nth root of the avarge of mean square error)

وهو يمثل الجذر N لمعدل متوسط مربعات الاخطاء ورمزه (RAMSE) وصيغته :

$$\text{RAMSE} = \sqrt[N]{\text{AMSE}} \quad \ldots(29)$$

اذ ان N : يمثل حجم العينة وان AMSE يمثل معدل متوسط مربعات الاخطاء.

ثالثا": معيار متوسط معدل نسبة الاخطاء AMPE :(Mean percentage error)

$$\text{AMPE} = \frac{1}{T}\sum_{i=1}^{T}(100)\text{Mse(i)} \qquad \ldots(30)$$

رابعا":معيار المتوسط النسبي المطلق للاخطاءMAPE : (Mean absolute percentage error)

$$\text{MAPE} = \frac{1}{T}\sum_{i=1}^{T}\frac{|ei|}{yi} \text{ x100} \qquad \ldots(31)$$

وبعض المصادر تتعامل مع هذا المقياس بالصيغة التالية: $$\text{MAPE} = \frac{100}{T}\sum_{i=1}^{T}\left[\frac{|ut|}{|yt|}\right]$$

3- الجانب التطبيقي

(3-1) 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 ولجانبي الكرخ والرصافة لعام *(2018)* وقد استعمل الباحث البرنامج *(R)* في التطبيق .

(3-2) 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* ملغرام/لتر .

BOD5: يدل على الحاجة البيولوجية للاوكسجين(وجود الاوكسجين في الماء) والمحدد البيئي (*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 *(1042-10)* ملغم/لتر وفي الشتاء تتراوح مابين (954-16) *ملغم/لتر.*

Mg: المغنيسيوم وتتراوح قيمته في الصيف مابين *(256-12)* ملغم /لترو الشتاء يتراوح مابين *(386-29)* ملغم /لتر.

So4: الكبريتات المحدد البيئي لهذا المتغير الذي يبلغ اقل من (200) ملغم /لتر.

Po4 : الفوسفات وقيمته يجب ان لا تزيد عن(*4-0*) ملغم / لتر.

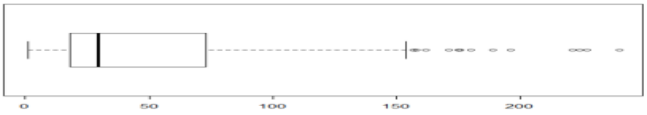
(3-5) عينة البحث والمتغيرات المستعملة

لقد اخذت عينه 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 لجانبي الكرخ والرصافة وهذه المحطات هي *(T17)* جسر المثنى، *(T18)* جسر الائمة، *(T19)* جسر الشهداء، *(T21)* جسر الجادرية، *(T22)* مشروع ماء الرشيد، *(T23)* محطة ماء الزعفرانية، *(T24)* مشروع ماء الدورة، *(D16)* نهر ديالى/ جسر ديالى الجديد، *(ST1)* نهر المصب العام في ابي غريب، *(ST2)* نهر المصب العام في المحمودية وكل محطة مرتبطة بالقطاعات الصحية المكونة من مستشفيات ومراكز صحية تابعة لوزارة الصحة أجري فيها الفحص المختبري لاشخاص بسبب اصابتهم بمرض الاميبا .

(3-4) تحليل البيانات

تم تطبيق طريقة *(Boxplot)* للكشف عن القيم الشاذة في المتغير المعتمد (عدد الإصابات بمرض الاميبا في مدينة بغداد في جانبي الكرخ والرصافة) كماهو مبين في الشكلين (*1*) و (*2*)، وقد تبين وجود قيم شاذة في لكلا الجانبين.

الشكل (1)(الرسم الصندوقي لاعداد المصابين بمرض الاميبا في بغداد / الكرخ لعام 2018)



الشكل (2)(الرسم الصندوقي لاعداد المصابين بمرض الاميبا في بغداد / الرصافة لعام 2018)

ونظرا لعدم توافق وحدات القياس للبيانات المدروسة فقد تم تحويلها الى الصيغة القياسية *(standardization)* $Z = \frac{X-\mu}{\sigma}$.

ولتشخيص نوع إنموذج بيانات البانل تم اجراء اختبارات الكشف عن وجود مشكلة عدم تجانس تباينات المقاطع وذلك من خلال استعمال اختبار(*Breusch-Pagan test*) وذلك لاختبار فرضية العدم التي تنص على عدم وجود مشكلة عدم تجانس تباينات المقاطع العرضية لبيانات البانل، ونتائج الإختبارات المذكورة أعلاه مبينة في الجدولين (5) و (6) أدناه:

جدول (5) إختبار واختبار عدم تجانس التباين لبيانات الكرخ

الإختبار	قيمة إحصاءة الاختبار	(p-value)
إختبار عدم تجانس تباينات المقاطع لبيانات الكرخ	21.408	0.0447

جدول(6) إختبار مضاعف لاكرانج وإختبار هوسمان واختبار عدم تجانس التباين لبيانات الرصافة

الاختبار	قيمة احصاءة الاختبار	(p-value)
إختبار عدم تجانس تباينات المقاطع لبيانات الرصافة	32.417	0.00119

يتضح من الجدول (*5*) ان قيمة *(p-value)* الخاصة باختبار (*Breusch-Pagan test*) لبيانات الكرخ (*0.0447*) اقل من مستوى المعنوية، وهذا يؤدي الى رفض فرضية العدم التي تنص على تجانس في تباينات المقاطع العرضية، ولذا فان بيانات الإصابة بمرض الاميبا لبيانات جانب الكرخ غير متجانسة في تباينات المقاطع العرضية.

ويتضح من خلال الجدول (*6*) الذي يمثل اختبارات بيانات الرصافة ان قيمة *(p-value)* الخاصة باختبار (*Breusch-Pagan test*) لبيانات الرصافة (*0.00119*) اقل من مستوى المعنوية، وهذا يؤدي الى رفض فرضية العدم التي تنص على تجانس تباينات المقاطع العرضية، وبذلك فان بيانات الإصابة بمرض الاميبا لبيانات جانب الرصافة غير متجانسة في تباينات المقاطع العرضية.

(5-3) تقدير الإنموذج:

تم تطبيق طرائق التقدير التي تم تناولها في الجانب النظري على بيانات مرض الاميبا لجانبي الكرخ والرصافة، وتقدير معلمات الانموذج المعلمي العشوائي لجانبي الكرخ والرصافة ووضعت النتائج في الجدول (7).

الجدول (7)(القيم التقديرية لمعلمات الانموذج العشوائي لبيانات الكرخ والرصافة)

المعلمة	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 طرائق التقدير المختلفة وبيان الانموذج الأفضل تم استعمال معايير المقارنة والمتمثلة بمعدل متوسط مربعات الخطا (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 تعريف وتفاصيل المقارنة بين تلك الطرائق في الجدولين(8)و(9)*

جدول (8)النماذج المستعملة في التقدير

	Model Name	Method
1	Non Parametric I	LLR, Gaussian
2	Non Parametric II	LLR, Epanchincov
3	Non Parametric III	WLLR, Gaussian
4	Non Parametric IV	WLLR, Epanchincov

جدول (9)قيم معايير المقارنة بين النماذج المقدرة لبيانات الكرخ

Model	AMSE	AMDAE	AMAE	AMAPE	Best
I	0.7514529	0.4862504	0.6314381	2.283848	3
II	0.7658378	0.4835688	0.6389129	2.219851	4
III	0.112107	0.05722371	0.1634525	0.567051	2
IV	0.047491	0.03255447	0.06784877	0.241275	1

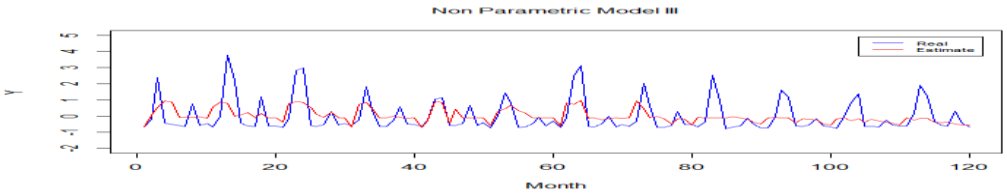
يتضح من خلال الجدول(9) الذي يخص معايير المقارنة المقدرة للنماذج المختلفة لبيانات الكرخ ان هو الأنموذج الاكثر كفاءة من بين النماذج كافة هو الانموذج اللامعلمي(*Non Parametric V*)المقدر وفق طريقة الانحدار الموضعي الخطي الموزون بالدالة اللبية (*Epanechnikov*) حسب قيم معايير المقارنة *(AMSE)* كونه اعطى اقل قيم

جدول (10)قيم معايير المقارنة بين النماذج المقدرة لبيانات الرصافة

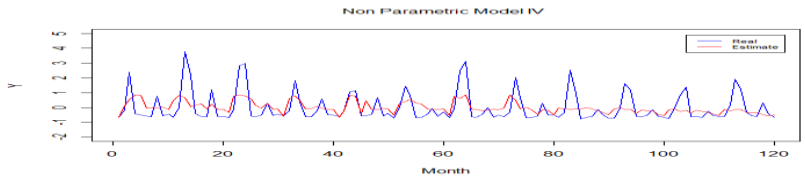
Model	RAMSE	AMDAE	AMAE	AMAPE	Best
I	0.8896763	0.5905003	0.735808	1.094696	3
II	0.8937988	0.5769253	0.7342525	1.073642	4
III	0.0006102	0.0004678	0.0087594	0.0328183	1
IV	0.0029091	0.0001585	0.0123274	0.0247718	2

كما ويتضح من خلال الجدول (10) الذي يخص معايير المقارنة المقدرة للنماذج المختلفة لبيانات الرصافة ان الانموذج اللامعلمي *(Non Parametric III)* المقدر وفق طريقة الانحدار الموضعي الخطي الموزون بالدالة اللبية *(Gaussian)* هو الأنموذج الاكثر كفاءة من بين النماذج كافة كونه اعطى اقل قيم لمعياري المقارنة *(AMSE)* و *(AMAE)* عدا ذلك فقد حل هذا الانموذج ثانيا.

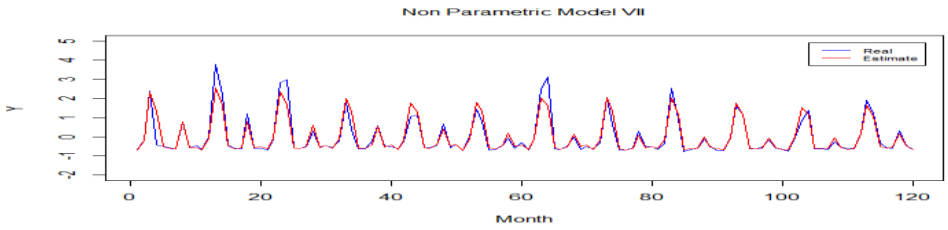
ولتوضيح عملية المقارنة بصورة أكثر دقة تم رسم المنحنيات الحقيقية والمقدرة والخاصة بنماذج بيانات البانل العشوائية لمتغير اعداد الاصابة بمرض الاميبا وكما هو موضح بالاشكال من (*3*) الى (*10*) كما في ادناه.



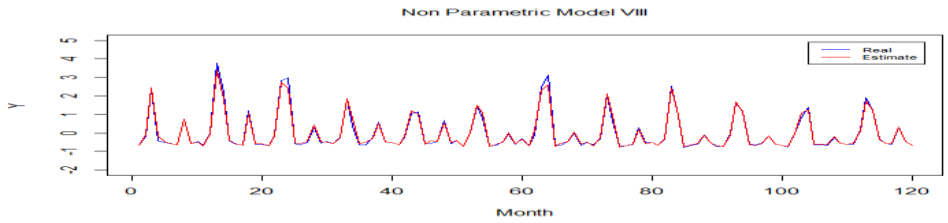
شكل *(3)*منحني بيانات اعداد الاصابة بمرض الاميبا للانموذج اللامعلمي I لبيانات الكرخ



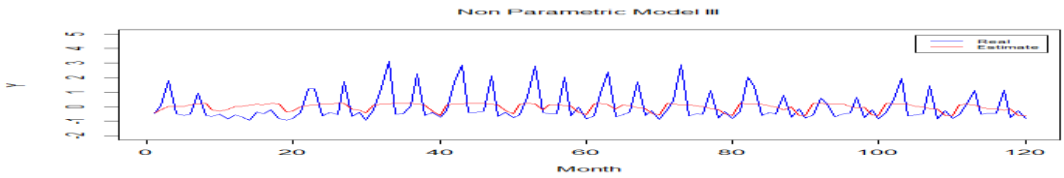
شكل (4)منحني بيانات اعداد الاصابة بمرض الاميبا للانموذج اللامعلمي II لبيانات الكرخ



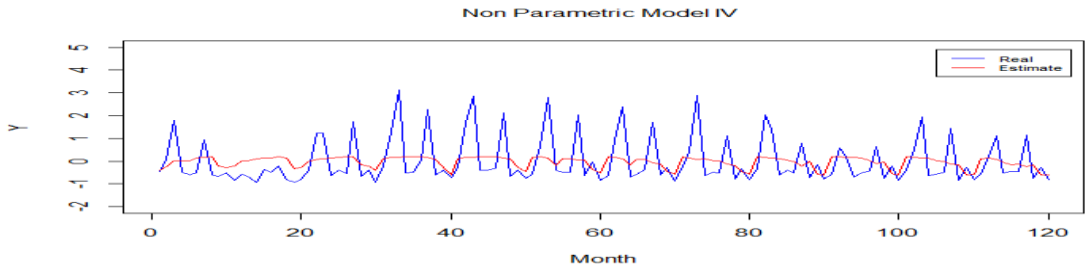
شكل (5)منحني بيانات اعداد الاصابة بمرض الاميبا للانموذج اللامعلمي III لبيانات الكرخ



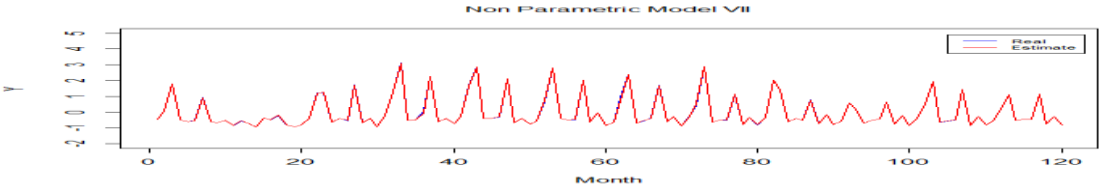
شكل (6)منحني بيانات اعداد الاصابة بمرض الاميبا للانموذج اللامعلمي IV لبيانات الكرخ



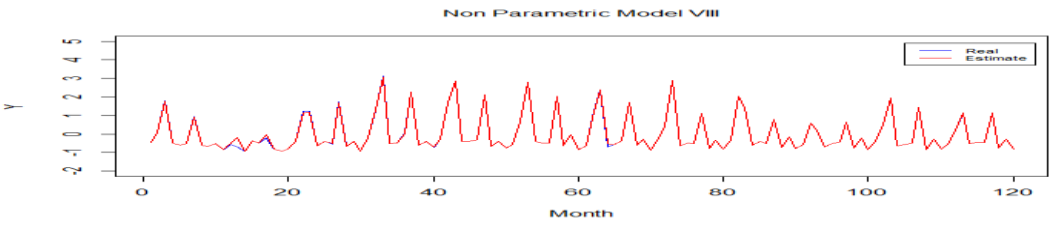
شكل (7)منحني بيانات اعداد الاصابة بمرض الاميبا للانموذج اللامعلمي I لبيانات الرصافة



شكل (8)منحني بيانات اعداد الاصابة بمرض الاميبا للانموذج اللامعلمي II لبيانات الرصافة



شكل (9)منحني بيانات اعداد الاصابة بمرض الاميبا للانموذج اللامعلمي III لبيانات الرصافة



شكل (10)منحني بيانات اعداد الاصابة بمرض الاميبا للانموذج اللامعلمي IV لبيانات الرصافة

ويتبين ان التمهيد وفق الانموذج اللامعلمي (*Non Parametric Model V*) والموضح في الشكل (*6*) لجانب الكرخ الاقرب لمشاهدات المتغير المعتمد الحقيقية، ويمكن القول ان التمهيد وفق هذا الانموذج هو الافضل .

اما التمهيد وفق الانموذج اللامعلمي (*Non Parametric Model* III) في الشكل (*9*) لجانب الرصافةهو الافضل من بين كافة النماذج التي تخص جانب الرصافة، كونه قريب من منحني المتغير المعتمد الحقيقي .

الاستنتاجات

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

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 الخطا العشوائي.
2) تبين ان علاقة المتغيرات التوضيحية التي تمثل خصائص المياه الصالحة للشرب بمتغير اعداد الاصابة في الكرخ كما يأتي: تراكيزالملوثات (PHوBOD5 و.T.H وPO4) في علاقة طردية، بينما ترتبط المتغيرات التوضيحية (CaوMgو SO4) بعلاقة عكسية مع المتغير المعتمد الذي يمثل اعداد الاصابة بالمرض.
3) تبين ان علاقة المتغيرات التوضيحية التي تمثل خصائص المياه الصالحة للشرب بمتغير اعداد الاصابة في الرصافة كما يأتي: تراكيزالملوثات (PHو.T.H وPO4وMg) في علاقة طردية، بينما ترتبط المتغيرات التوضيحية (Ca وBOD5 و SO4) بعلاقة عكسية مع المتغير المعتمد الذي يمثل اعداد الاصابة بالمرض.
4) من خلال قيم معايير المقارنة ومنحنيات التمهيد تبين ان الانموذج الأكثر كفاءة لاعداد الاصابة هو الانموذج المقدر وفق طريقة الانحدار الموضعي الخطي الموزون بالدالة اللبية (Gaussian)،
5) بالنسبة للنماذج اللامعلمية غيرالموزونة فقد كان الانموذج الاكفأ لبيانات اعداد الاصابة لجانبي الكرخ والرصافة هو الانموذج المقدر وفق الانحدار الموضعي الخطي المتعدد بدالة اللب (Gaussian) .
6) الدالة اللبية (Gaussian) اكفأ من الدالة اللبية (Epanchincov) في تقدير انموذج بيانات البانل العشوائي للاصابة بمرض الاميبا لكافة طرائق التقديراللامعلمية الموزونة وغير الموزونة.

التوصيات

1) استعمال طرق تقدير اخرى لامعلمية ومقارنتها مع هذه الطرق لغرض الوصول للانموذج الاكثر جودة الذي يمثل بيانات بانل فضلا" عن استعمال طرق تمهيد ماعدا الطرق اللبية للنماذج اللامعلمية.
2) ضرورة العمل على توفير كادر متقدم مختص باجراء الفحوصات المختبرية لعينات المياه وحساب تراكيز الملوثات عن طريق ارسالهم لبعثات ودورات خارج العراق ليتمكنوا من ادارة الاجهزة المتوفرة لديهم الخاصة بالتحليلات المختبرية فضلا" عن ذلك تجهيز مختبرات الصحة والبيئة باجهزة 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.
3) ضرورة قيام الدولة الممثلة بالحكومة المركزية والحكومات المحلية بتعميم قرارات تمنع فيها مؤسساتها الحكومية اولا من رمي المخلفات السائلة والصلبة في نهر دجلة واعتبارة جريمة يحاسب عليها القانون لما يتسبب من ضرر كبير في المجال الصحي والبيئي وتسببه

بامراض خطيرة وثانيا المؤسسات والشركات والمصانع الاهلية بمحاسبتها قانونيا ويعتبر هذا موازي للضريبة السنوية لهذه الجهات و حسب المحددات البيئية لنظام صيانة الانهار رقم (25) لسنة (1967) في العراق .

4) ضرورة اجراء بحوث احصائية اخرى تهتم بامراض اخرى تسببها هذه الملوثات مثل الامراض السرطانية وغيرها مع توعية الناس بخطورتها ومعرفة ماهية التراكيز الملوثة المسببة لهذه الامراض.

المصادر العربية

- 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.
- طراد,علاء جابر ,(2013), " نماذج الانحدار شبه المعلمي (دراسة مقارنة) ", رسالة ماجستير في الاحصاء ,كلية الادارة والاقتصاد, الجامعة المستنصرية.
- وهيبة ، مقام وسارة ، (2018) ، " أثر الامتثال للسياسات البيئية على القدرة التنافسية لصادرات دول نامية مختارة – دراسة قياسية باستخدام بيانات البانل لعينة من 5 دول نامية (2007 – 2016)", مجلة العلوم الاقتصادية و التسيير والعلوم التجارية ،المجلد .(11) ، العدد (2) ، ص (280 – 291) .

المصادر الاجنبية

- Baltagi ,badi, (2006) “Forcasting with panel data”Econometric studies, series1,NO.25
- Hajargasht, C, , (2009) , " Nonparametric Panel Data Models A Penalized Spline Approach " , Marcel Dekker, Inc. New York.
- Huansha Wang , (2014), " Essays on Semiparametric Ridge-Type Shrinkage Estimation, Model Averaging and Nonparametric Panel Data Model Estimation" A Dissertation Submitted in Partial Satisfaction of The equirements for The Degree of Doctor of Philosophy in Economics , Uuiversity of California.
- Park , H.M,(2005),” Linear regression models for panel data using SAS STATS,LIMDEP and SSPS” http://www.indiana.edu/~stat.math.
- 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, 96009600 Garsington road . oxford ox4 2Dce , UK and 350 main street , Molden . MA 02148 USA
- Qingguo, T. and Longsheng, C., (2008), "M-estimation and B-spline approximation for varying coefficient models with longitudinal data". Journal of nonparametric statistics, vol. 20, no. 7, pp. 611-625.
- Sosa, J. C. and Diaz, L. G., (2010), "Time-varying coefficient model component estimation through generalized estimation equations", Revista colombiana de estadistica, vol. 33, no. 1, pp. 89-109.
- SU,L. and ULLAH , A. , (2011) , " Nonparametric and semi parametric panel data Econometric models of Empirical Economics and Finance , (pp.455-497). Research collection school of Economics , Available at: https://ink.library.smv.edu.sg / soe- research/1281

- 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/.
- Swamy ,P.A.V.B&Mehta,J.S ,(1975), "Bayesian and non Bayesian- Analysis of switching regressions and of random coefficient Regression models" JASA,No.351 ,p.p.593.
- Vidra , H. , Rong , C. and Dag , T. , 2001 , " Non- parametric estimation and testing in panels of inter correlated time series " , support from the Norwegian super computer committal through a grant of computing time . block well publishing Ltd(2014(
- Wacalwiw,Myron.A.&Liang,K-Y.,(1993),"Prediction of random effect in the generalized linear models" ,JASA,Vol.88,No.421,p (171؛ Estimation and Testing " , Handbook
- Ware.James.H ,(1985), "Linear models for the analysis of longitudinal Studies " the American statistician ,Vol.39,No.2 p.p.95.
- Wu, C. O. and Chiang, C. T., (2000), "Kernel smoothing on varying coefficient models with longitudinal dependent variable", statistica sinica, vol. 10, pp. 433-456.
- Wu, C. O. and Yu, K. F., (2002), "Non parametric varying-coefficient models for the analysis of longitudinal data", International statistical review, vol. 70, no. 3, pp. 373-393.
- Wu, W. B. and Pourahmadi, M., (2003), " Non parametric estimation of large covariance matrices of longitudinal data", Biometrika, vol. 90, ISS. 4, pp. 831-844.
- 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.
- Xu,Wil and Zhu, L.X.,(2009)," A goodness –of fit test for varying – coefficient model in longitudinal studies ", Journal of non-parametric statistics,Vol.21 ,no.4,pp.427-440
- Vogt , M . and Linton , O . (2016) ." Classification of Nonparametric Regression Functions in Heterogeneous Panels ", the Canadian journal of statistics, vol.29,pp597_608.

(Comparison of the local linear regression estimation method with the weighted local linear regression method for nonparametric models of longitudinal data)

Abatract

The water is polluted as a result of the presence of foreign substances in it, whether organic or inorganic. These materials corrupt the water quality, which is what happens to the Tigris River. These wastes pollute the waters of the Tigris River with dangerous chemical pollutants that harm human health, causing diseases, and among these diseases is the amoeba disease, whose data represented by numbers of disease cases were taken on the Karkh and Rusafa sides in Baghdad for the year 2018 as a dependent variable (Y) in a whole series of 12 months And to know the effect of water pollutant concentrations on the number of amebiasis infections, seven pollutant concentrations were relied upon,

which represent the explanatory variables (Xi), which represent cross-sectional data for ten stations located on the station of the Tigris River on both Karkh and Rusafa sides. The longitudinal data, or the so-called Panel data, is its best designation. We used nonparametric Panel data models, which are the positioned linear estimator and the weighted local linear estimator, to determine which of the two estimators is the most efficient that gives the best model. From the problem of heterogeneity of variance and that the data of the variables of the concentrations of the seven pollutants suffer from the problem of Auto-correlation, and that the appropriate parameter model is the random effect model of the Panel data according to the (Houseman) test.