التقدير المتين للتباين التقاربي في نموذج الإنحدار البواسوني لبيانات سلاسل زمنية

أ.د. محمد توفيق البلقينى

أستاذ الرياضيات والإحصاء الإكتوارى

كلية التجارة - جامعة المنصورة

أ.د. البيومى عوض طاقية

أستاذ الإحصاء التطبيقى

كلية التجارة - جامعة المنصورة

الباحث / أحمد ابوسليمان العدل الطنطاوى

معيد بقسم الإحصاء التطبيقى

كلية التجارة - جامعة المنصورة

التقدير المتين للتباين التقاربي في نموذج الإنحدار البواسوني لبيانات سلاسل زمنية

الملخص:

يستهدف هذا البحث تقدير التباين التقاربي في نموذج الانحدار البواسوني لتحليل بيانات سلسلة زمنية وذلك في ظل وجود مشكلات التقدير مثل الارتباط الذاتي وعدم ثبات تباين حدود الخطأ heteroskedasticity، ويعد هذا النموذج أحد أنواع نماذج المعلمة المشتقة parameter-driven models، حيث يتم التعبيرعن الارتباط الذاتي بين المشاهدات من خلال إدراج العملية الكامنة latent process في دالة الربط للنموذج، ويتم تقدير متجه معاملات الانحدار β من خلال تعظيم دالة الإمكان الزائفة maximizing the pseudo-likelihood والتي تتجاهل وجود العملية الكامنة. ويكون المقدر الناتج هو مقدر نموذج خطي معمم، وقد تم عرض الإتساق والتقارب الطبيعي لهذا المقدر في دراسة Davis et al (2000). ومن أجل إجراء الاستدلالات الإحصائية بشكل صحيح حول معاملات الانحدار، فإن ذلك يتطلب إيجاد تقدير متسق للتباين التقاربي لـ$\hat{\beta}$، وحيث أن طرق التقدير المتينة تمكننا من الحصول على تقدير متسق للتباين التقاربي في ظل وجود مشكلات التقدير مثل الارتباط الذاتي و عدم ثبات تباين حدود الخطأ فإن هذا البحث سوف يقوم بالمقارنة بين طريقتين من طرق التقدير المتينة للتباين التقاربي وهما؛ طريقة أساس كرنال kernel-based التي تم اقتراحها في دراسة Wu (2012) ومقدر OPG أو خوارزمية BHHH التي نقترح استخدامها في هذا البحث. وبالتطبيق العملي على مجموعتين من البيانات؛ الأعداد الشهرية لحالات شلل الأطفال في الولايات المتحدة الأمريكية في الفترة من ١٩٧٠ إلى ١٩٨٣ والأعداد الشهرية لحالات استقبال الحمى الروماتيزمية في مستشفي أطفال المنصورة خلال الفترة من ٢٠١٠ إلى ٢٠٢١، وقد أوضحت النتائج أن مقدر OPG أو خوارزمية BHHH هي الأفضل في تقدير التباين التقاربي لـ $\hat{\beta}$.

ROBUST ESTIMATION OF ASYMPTOTIC VARIANCE IN A POISSON REGRESSION MODEL FOR TIME SERIES DATA.

Abstract:

This paper aims to estimate asymptotic variance for Poisson regression model for time series of counts in the presence of the estimation problems, such as; autocorrelation and heteroskedasticity. This model represents a parameter-driven model since the autocorrelation among the observations is introduced by incorporating a latent process in the link function of the model. The regression coefficient vector β is estimated by maximizing the pseudo-likelihood that ignores the existence of the latent process. The resulting estimator is a generalized linear model estimator, and its consistency and asymptotic normality have been established by Davis et al (2000). To perform valid statistical inferences about the regression coefficients, it is required to develop a consistent estimation procedure for the asymptotic covariance matrix of $\hat{\beta}$. Since the robust estimation enables us to obtain a consistent estimation for the asymptotic variance in the presence of the estimation problems, such as; autocorrelation and heteroskedasticity. This study will compare between two robust estimation methods for the asymptotic variance; kernel-based method which is suggested by Wu (2012) and OPG estimator or BHHH algorithm which is proposed in this paper. The methods are applied to two data sets; the monthly polio data in the U.S.A from 1970 to 1983 and monthly numbers of rheumatic fever in Mansoura University Children Hospital from 2010 to 2021. The result revealed that OPG estimator or BHHH algorithm is a better estimator for the asymptotic variance of $\hat{\beta}$.

١- المقدمة:

إن تحليل السلاسل الزمنية التي تتكون من قيم معدودة (أي قيم صحيحة غير سالبة) هي محل إهتمام العديد من الباحثين، وذلك لأن لها العديد من التطبيقات في عالم الواقع. يهتم هذا البحث بتقدير التباين التقاربي(أو مصفوفة التغاير التقاربي Asymptotic Covariance Matrix) في فئة النماذج الخطية المعممة لمعلمة مشتقة Parameter-driven Generalized linear model لتحليل بيانات سلاسل زمنية معدودة والتي تعتبر أداة هامة في نمذجة السلاسل الزمنية التي لا تتبع التوزيع الطبيعي. إن تحديد نموذج لبيانات سلاسل زمنية معدودة يتطلب الأخذ في الإعتبار الإرتباط الذاتي بين المشاهدات ، ولقد قام(1981) Cox بتقسيم نماذج السلاسل الزمنية إلى فئتين من النماذج، وذلك بناءاً على كيفية إدخال الإرتباط الذاتي بين البيانات إلى النموذج وهما:

- نماذج المشاهدة المشتقة Observation-driven models.

وفي هذا النوع من النماذج يتم إدخال الإرتباط الذاتي بين البيانات إلى النموذج من خلال جعل المشاهدات الحالية مُعتمدة بشكل واضح وصريح على مُشاهدات الماضي مثل AR وARIMA وGARCH .

- نماذج المعلمة المُشتقة Parameter-driven models.

وفي هذا النوع من النماذج يتم تقديم الإرتباط الذاتي بين البيانات من خلال عملية كامنة latent process تُضاف إلى المتنبأ الخطي. ولذلك، يُمكننا القول أن النماذج الخطية المعممة لبيانات سلاسل زمنية معدودة –والتي هي محل اهتمامنا في هذا البحث- تُعد نوع من أنواع نماذج المعلمة المشتقة Parameter-driven models. بحيث يتم إدخال العملية الكامنة latent process إلى النموذج الخطي المُعمم من خلال إضافتها إلى دالة الربط.

وفي هذا البحث سوف يتم تقدير التباين التقاربي في نموذج إنحدار لبيانات سلاسل زمنية معدودة، وذلك في ظل نموذج خطي معمم لمعلمة مشتقة بحيث يتم التعبيرعن الإرتباط الذاتي بين بيانات السلسلة الزمنية المعدودة من خلال عملية كامنة latent process تُضاف إلي المتنبأ الخطي.

وبفرض أن التوزيع الشرطي لمتغير الإستجابة Y_t بمعلومية العملية الكامنةα_t يتبع توزيع بواسون- أي أن $Y_t|\alpha_t \sim Po(\lambda_t)$.

(1-1) المشكلة:

في هذه الدراسة نحاول الحصول على مُقدّر مُتسق لمصفوفة التغاير التقاربي لـ $\hat{\beta}$ consistent estimator for the asymptotic covariance matrix في نموذج الإنحدار البواسوني لتحليل بيانات سلسلة زمنية معدودة. وحيث أن مصفوفة التغاير التقاربي يُمكن التعبير عنها بحدود دوال التغاير الذاتي(ACVFs) للعملية الكامنة α_t، فقد قام Davis and Wu (2009) بتقديم طريقة مُبتكرة لتقدير مصفوفة التغاير التقاربي في ظل نموذج إنحدار ذو حدين سالب لتحليل سلسلة زمنية معدودة، وذلك من خلال تحديد نموذج للعملية الكامنة α_t وتقدير معلماتها وتغايراتها الذاتية(ACVFs) والحصول على مقدّر مصفوفة التغاير التقاربي من خلال التعويض بمقدّرات التغايرات الذاتية للعملية الكامنة α_t ولكن إتساق هذا المقدّر كان مشكوك فيه، وذلك لعدم إمكانية إثباته. وهذا ما تم تداركه في دراسة Wu (2012) حيث تم تقدير مصفوفة التغاير التقاربي لـ $\hat{\beta}$ بدون تخصيص نموذج للعملية الكامنة α_t ، ولكن تم التعامل معها بشكل لامعلمي. وذلك من خلال استخدام الطرق اللامعلمية لتقدير التباين، وقد استخدم طريقتين هما؛ طريقة أساس كرنال Kernel-based التي قدّمها Newey and West (1987) وطريقة المعاينة الفرعية Subsampling التي قدّمها Carlstein (1986)، وبذلك حصل على مُقدّرين لمصفوفة التغاير التقاربي لـ $\hat{\beta}$ بالإعتماد على هاتين الطريقتين. وقد أثبت الإتساق لكلا المُقدّرين في ظل نموذج إنحدار ذو حدين سالب لتحليل بيانات سلسلة زمنية معدودة. وكان الدافع وراء هذا البحث هو أن دراسة Wu (2012) قد أوضحت أنه يمكننا الحصول على مقدر متسق لمصفوفة التغاير التقاربي لـ $\hat{\beta}$ بإستخدام طريقة أساس كرنال Kernel-based عند استخدام نموذج خطي معمم لمعلمة مشتقة والتي يكون فيها التوزيع الشرطي لـ $Y_t|\alpha_t$ هو توزيع بواسون بدلاً من توزيع ذو الحدين السالب.

وفي هذه الدراسة نحاول الإجابة على السؤال: هل مقدر أساس كرنال Kernel-based يقدم أفضل تقدير متسق لمصفوفة التغاير التقاربي لـ $\hat{\beta}$ في نموذج الإنحدار البواسوني لبيانات سلاسل زمنية ؟ وذلك من خلال مقارنته بمقدر آخر نقترح استخدامه في هذه الدراسة وهو (مقدر حاصل الضرب الخارجي لمتجهات التفاضلات الجزئية من الدرجة الأولي) outer product of gradients (OPG) estimator والذي يسمى أيضاً خوارزمية BHHH والتي قدمها Berdent, Hall, Hall and Hausman (1974) وهي طريقة رقمية تستخدم في البحث التجريبي.

(1-2) الأهمية:

إن أهمية هذه الدراسة تنبع من ضرورة إيجاد مُقدّر متسق لمصفوفة التغاير التقاربي لـ $\hat{\beta}$ ، وذلك بهدف إجراء الإستدلالات الإحصائية الممكنة بالنسبة لـ β من إنشاء فترات ثقة وإختبارات فروض، حيث تتجه الأفكار المستقبلية فى هذا الصدد نحو الإستدلال الإحصائى ،وكخطوة أولية فى هذا الإتجاه نحاول تقديم مقدّر متسق لمصفوفة التغاير التقاربى لـ $\hat{\beta}$ التى حصُلنا عليها من خلال تعظيم

دالة الإمكان اللوغاريتمي الزائفة وذلك في ظل نموذج خطى مُعمّم لمعلمة مُشـــتقة Parameter-driven Generalized linear model مُعتمداً على توزيع بواسون.

(1-3) أهداف البحث:

يهدف هذا البحث إلى استخدام طريقة أساس كرنال Kernel-based ومقدر OPG أو ما يطلق عليه خوارزمية BHHH لتقدير التباين التقاربي لـ $\hat{\beta}$ وذلك في ظل استخدام نموذج الإنحدار البواسوني لبيانات سلاسل زمنية معدودة، وسوف تتم المقارنة بين التقديرات الناتجة بالتطبيق على مجموعتين من البيانات، هما الأعداد الشهرية للإصابة بشلل الأطفال في الولايات المتحدة الأمريكية خلال الفترة من ١٩٧٠ إلى ١٩٨٣والأعداد الشهرية لحالات الحمى الروماتيزمية التي يتم استقبالها بمستشفي أطفال المنصورة خلال الفترة من ١ يناير ٢٠١٠ إلي ٣١ ديسمبر ٢٠٢١ .

٢ ـ النموذج الخطي المعمم لمعلمة مشتقة لبيانات سلسلة زمنية معدودة

يتم تحديد النموذج الخطي المُعمم لمعلمة مشتقة Parameter-driven Generalized linear model من خلال:

<u>المُكوّن العشوائي random component :</u>

وفيه يتم تحديد التوزيع الشرطي لمتغيرات الإستجابة Y_t بمعلومية العملية الكامنة α_t، أي تحديد $p(Y_t|\alpha_t)$ ، وذلك بفرض أن المشاهدات $Y_1, \dots \dots . Y_n$ تكون مستقلة بمعلومية α_t.

<u>المُكوّن المنتظم systematic component :</u>

وفيه يتم تحديد دالة الربط $f(.)$ التي يتحدد من خلالها شكل العلاقة بين μ_t المتوسط الشرطي لـ Y_t والمتنبأ الخطي مُضافاً إليه العملية الكامنة α_t، وبالتالي فإن دالة الربط تأخذ الشكل التالي $f(\mu_t) = X_t^T\beta + \alpha_t$ ، حيث تقوم العملية الكامنة α_t بإدخال الإرتباط الذاتي في النموذج وذلك بفرض أنها تخرج بشكل مستقل من المشاهدات.

وفي هذا البحث سوف نفترض نموذج خطي مُعمم لمعلمة مشتقة وهو نموذج الإنحدار البواسوني لتحليل بيانات سلاسل زمنية معدودة.

(2-1) نموذج الإنحدار البواسوني لبيانات سلسلة زمنية

بفرض أن $\{Y_t: t = 1,2, \dots \dots, n\}$ ترمز لبيانات سلسلة زمنية معدودة، وأن X_t هو مُتجه المتغيرات التفسيرية والتي يُفترض أن تكون غير عشوائية ومُكونها الأول هو الواحد الصحيح، وفي بعض الحالات، قد يعتمد X_t على حجم العينة n، ولذلك سوف يتم كتابتها بدليل سُفلى مزدوج كالتالي X_{nt}. وافترضنا أن المتغيرات العشوائية $Y_1, \dots Y_n$ تكون مُستقلة بمعلومية العملية الكامنة α_t، وأن التوزيع الشرطي للمشاهدة Y_t بمعلومية العملية الكامنة α_t هو توزيع بواسون ؛أي أن:

$$Y_t \Big| \alpha_t \overset{indep}{\sim} po(\lambda_t) \qquad (1)$$

وأن دالة الربط هي:

$$\log \lambda_t = X_{nt}^T \beta + \alpha_t \qquad (2)$$

حيث أن $\beta = (\beta_1, \dots, \beta_k)^T$ هو مُتجه معلمات الإنحدار، وأن α_t هي عملية خطية ساكنة. فبدلاً من التعامل مع العملية الكامنة α_t أثناء التفاضل، فإنه من الأنسب استخدام ϵ_t، حيث أن $\epsilon_t = e^{\alpha_t}$، وهي سلسلة زمنية ساكنة وغير سالبة بمتوسط واحد؛ أي أن $E(\epsilon_t) = 1$. ومن ثم يُمكن صياغة المتوسط الشرطي لـ Y_t بمعلومية α_t على النحو التالي:

$$E(Y_t|\alpha_t) = \lambda_t = \exp(X_{nt}^T \beta + \alpha_t) = \exp(X_{nt}^T \beta)\epsilon_t \qquad (3)$$

وبالتالي، فإن متوسط Y_t هو:

$$\mu_t = E(Y_t) = E[E(Y_t|\alpha_t)] = E[\exp(X_{nt}^T \beta)\epsilon_t] = \exp(X_{nt}^T \beta) \qquad (4)$$

وتباين Y_t هو:

$$\begin{aligned} var(Y_t) &= E[\mathrm{var}(Y_t|\alpha_t)] + \mathrm{var}[E(Y_t|\alpha_t)] \\ &= E[\exp(X_{nt}^T \beta)\epsilon_t] + \mathrm{var}[\exp(X_{nt}^T \beta)\epsilon_t] \\ &= \exp(X_{nt}^T \beta) + \exp(2X_{nt}^T \beta)\mathrm{var}(\epsilon_t) \\ &= \exp(X_{nt}^T \beta)[1 + exp(X_{nt}^T \beta)\gamma_\epsilon(0)] \\ &= \mu_t[1 + \mu_t \gamma_\epsilon(0)] \qquad (5) \end{aligned}$$

ودالة التغاير الذاتي لـ Y_t هي :

$$cov(Y_{t+h}, Y_t) = cov[E(Y_{t+h}|\alpha_{t+h}), E(Y_t|\alpha_t)] + 0$$

$$= cov\left(e^{(X_{nt+h}^T\beta)}\epsilon_{t+h}, e^{(X_{nt}^T\beta)}\epsilon_t\right)$$

$$= E\left[\left(e^{(X_{nt+h}^T\beta)}\epsilon_{t+h} - e^{(X_{nt+h}^T\beta)}\right)\left(e^{(X_{nt}^T\beta)}\epsilon_t - e^{(X_{nt}^T\beta)}\right)\right]$$

$$= e^{(X_{nt+h}^T + X_{nt}^T)\beta}\ E(\epsilon_{t+h}\epsilon_t) - e^{(X_{nt+h}^T + X_{nt}^T)\beta} E(\epsilon_{t+h})$$

$$-e^{(X_{nt+h}^T + X_{nt}^T)\beta}\ E(\epsilon_t) + e^{(X_{nt+h}^T + X_{nt}^T)\beta}$$

وحيث أن:

$$E(\epsilon_{t+h}) = E(\epsilon_t) = 1$$

فإن:

$$cov(Y_{t+h}, Y_t) = e^{(X_{nt+h}^T + X_{nt}^T)\beta}(\gamma_\epsilon(h) - 1)$$

$$= e^{X_{nt+h}^T\beta} e^{X_{nt}^T\beta}(\gamma_\epsilon(h) - 1)$$

$$= \mu_{t+h}\mu_t(\gamma_\epsilon(h) - 1) \qquad (6)$$

عند $h \neq 0$

وبفرض أن مُتجه معلمات الإنحدار الحقيقي هو β_0 ، فإنه يُمكن الحصول على مُقدّر النموذج الخطي المُعمم $\hat{\beta}_n$ من خلال تعظيم دالة الإمكان اللوغاريتمي الزائفة pseudo log-likelihood كالتالي:

$$\ell(\beta) = log\left\{\prod_{t=1}^{n} e^{-\lambda_t}\frac{\lambda_t^{Y_t}}{Y_t!}\right\}$$

$$= \sum_{t=1}^{n} log\, e^{-\lambda_t} + \sum_{t=1}^{n} log\, \lambda_t^{Y_t} - log\left[\prod_{t=1}^{n} Y_t!\right]$$

$$= -\sum_{t=1}^{n} \lambda_t + \sum_{t=1}^{n} Y_t \log \lambda_t - \log\left[\prod_{t=1}^{n} Y_t!\right]$$

وحيث أن $\lambda_t = e^{(X_{nt}^T\beta+\alpha_t)}$ ، وبإهمال وجود العملية الكامنة، فإن $\lambda_t = e^{(X_{nt}^T\beta)}$ وبالتالي نحصل على دالة الإمكان الزائفة التالية:

$$\ell(\beta) = -\sum_{t=1}^{n} e^{(X_{nt}^T\beta)} + \sum_{t=1}^{n} Y_t X_{nt}^T \beta - \log\left[\prod_{t=1}^{n} Y_t!\right] \qquad (7)$$

وفي ظل شروط الاتساق والتقارب الطبيعي الوارد ذكرها في دراسة Davis et al (2000) ، يكون التوزيع التقاربي Asymptotic distribution لـ $\hat{\beta}_n$ عندما $n \to \infty$، كالتالي:

$$M_n^{-1}\left(\hat{\beta}_n - \beta_0\right) \xrightarrow{d} N(0, [\Omega_1(\beta_0)]^{-1}[\Omega_1(\beta_0) + \Omega_{11}(\beta_0)][\Omega_1(\beta_0)]^{-1})$$

بحيث يكون:

$$\Omega_1(\beta_0) = \lim_{n\to\infty} M_n^T \left(\sum_{t=1}^{n} X_{nt} X_{nt}^T e^{X_{nt}^T \beta_0}\right) M_n \qquad (8)$$

$$\Omega_{11}(\beta_0) = \sum_{k=-\infty}^{\infty} W_k \gamma_\epsilon(k) \qquad (9)$$

وذلك عندما تكون $\gamma_\epsilon(k)$ دالة التغاير الذاتي لـ ϵ_t عند فجوة زمنية k . وحيث أن

$$W_k = \lim_{n\to\infty} M_n^T \left(\sum_{t=1}^{n} X_{nt} X_{n,t+k}^T e^{(X_{nt}^T + X_{n,t+k}^T)\beta_0}\right) M_n \qquad (10)$$

وذلك بفرض أن $M_n = n^{\frac{-1}{2}} I_P$ حيث أن I_P هي مصفوفة الوحدة $P \times P$ والتي تجعل النهايات السابقة موجودة، بحيث تكون مصفوفة التغاير التقاربي لـ $\hat{\beta}_n$ لها الصيغة التالية:

$$[\Omega_1(\beta_0)]^{-1}[\Omega_1(\beta_0) + \Omega_{11}(\beta_0)][\Omega_1(\beta_0)]^{-1}$$

3 - تقدير مصفوفة التغاير التقاربي لـ $\hat{\beta}$

(٣-١) مشاكل التقدير:

(٣-١-١) عدم ثبات التباين: Heteroskedasticity

فى حالة أن يكون تباين حد الخطأ ثابت عند المشاهدات المختلفة للمتغير التفسيري فإن هذا يسمى ثبات التباين homoskedasticity بمعنى أن تباين التوزيع الإحتمالي لحد الخطأ (وهو نفسه التباين الشرطي لـ Y_i بمعلومية X_i) يظل ثابت بغض النظر عن القيم التي يأخذها المتغير التفسيري (X) ، ونجد أن عدم ثبات تباين حد الخطأ عند المشاهدات المختلفة للمتغير التفسيري يسمى عدم ثبات التباين heteroskedasticity. وهناك عدة أسباب يمكن أن تؤدي الى عدم ثبات تباين حد الخطأ ومنها ما يلي:

١- طبيعة الظاهرة محل الدراسة، فقد يزيد تباين أو تشتت الأخطاء كلما زادت قيم المتغير التفسيري فعلى سبيل المثال؛ يزيد تباين الإنفاق عند قيم الدخل المرتفعة ويقل تباين الإنفاق عند قيم الدخل المنخفض حيث أن أصحاب الدخول المرتفعة يكون لديهم نطاق أكبر للإختيار بشأن التصرف في دخولهم وبالتالي من المتوقع أن يزيد التباين في الإنفاق كلما زاد الدخل، وقد يحدث العكس بأن يقل تباين أو تشتت الأخطاء كلما زادت قيم المتغير التفسيري مثل الظواهر التي تعتمد على التعلم من الخطأ فنجد أنه كلما زاد تعلم الأشخاص كلما قل خطأهم السلوكي.

٢- يمكن أن يظهر عدم ثبات التباين كنتيجة لوجود القيم المتطرفة outliers في مشاهدات العينة . ونجد أن إستبعاد أو تعديل هذه المشاهدات خاصة إذا كان حجم العينة صغير فإن ذلك قد يؤدي إلى تغير النتائج.

٣- عدم تحديد نموذج الإنحدار بشكل صحيح قد يكون سبب في عدم ثبات تباين حد الخطأ، ففي كثير من الأحيان يظهر عدم ثبات التباين بسبب عدم وجود بعض المتغيرات الهامة في النموذج، وتختفي هذه المشكلة بمجرد إدخال هذه المتغيرات الهامة إلى النموذج.

٤- الإلتواء في توزيع واحد أو أكثر من المتغيرات التفسيرية في النموذج قد يتسبب في عدم ثبات تباين حد الخطأ.

٥- هناك مصادر أخرى لعدم ثبات تباين حد الخطأ مثل إستخدام صيغة خاطئة للدالة التي تمثل النموذج أو إستخدام تحويلات غير صحيحة للبيانات.

وللكشف عن عدم ثبات التباين يمكن إستخدام عدد من الإختبارات الإحصائية؛ وأحد هذه الاختبارات هو إختبار Breusch-Pagan.

(٢-١-٣) مشكلة الإرتباط الذاتي Autocorrelation

في الحالة التي يكون فيها الإرتباط بين حدود الخطأ المتتالية في نموذج الإنحدار مساوي للصفر ، فإن هذا يعني أن حدود الخطأ مستقلة، أما في حالة أن تكون حدود الخطأ في نموذج الإنحدار مرتبطة فإن هذا يطلق عليه مشكلة الإرتباط الذاتي، ونجد أن مشكلة الإرتباط الذاتي عادة ما تظهر في بيانات السلاسل الزمنية حيث يتم جمع البيانات بشكل مرتب خلال الزمن وبالتالي فإن المشاهدات المتتالية عادة ما تكون مرتبطة خاصة إذا كانت الفترات الزمنية بين المشاهدات المتتالية قصيرة ، ونجد أن هناك عدة أسباب لوجود الإرتباط الذاتي بين حدود الخطأ في نموذج الإنحدار ومنها ما يلي:

١- أن يكون هناك ترتيب ما في وحدات المعاينة التي يتم دراستها قد يؤدي إلى وجود الإرتباط الذاتي سواء كانت المشاهدات مرتبة خلال الزمن مثل بيانات السلاسل الزمنية أو أن تكون المشاهدات مرتبة في فضاء مثل البيانات المقطعية cross-section data

٢- إستبعاد بعض المتغيرات الهامة من النموذج قد يكون هو مصدر الإرتباط الذاتي حيث أن ذلك يؤدي إلى وجود نمط ما في قيم البواقي الناتجة من توفيق نموذج الإنحدار، وعند إدخال هذه المتغيرات الهامة إلى النموذج يختفي نمط الإرتباط الذي تم مشاهدته بين البواقي، إلا أن هناك أسباب يمكن أن تؤدي الى حذف أو إهمال بعض المتغيرات الهامة في النموذج ومنها أن يكون المتغير وصفي أو أن مشاهدات متغير ما غير متاحة.

٣- التحديد الخاطئ لصيغة الدالة في نموذج الإنحدار، فعلى سبيل المثال عندما يتم إفتراض أن هناك علاقة خطية بين المتغير التابع والمتغيرات التفسيرية ، في حين أن هناك حدود أسية أو لوغاريتمية في النموذج فإن هذا قد يؤدي إلى ظهور الإرتباط الذاتي بين حدود الخطأ.

٤- إجراء تحويلات على البيانات قد يتسبب في وجود إرتباط ذاتي بين حدود الخطأ.

وللكشف عن وجود الإرتباط الذاتي بين حدود الخطأ في نموذج الإنحدار يمكن إستخدام عدد من الإختبارات وأحد هذه الإختبارات هو إختبار Durbin-Watson .

ونجد أن مشكلتي الإرتباط الذاتي بين حدود الخطأ وعدم ثبات تباين حد الخطأ تؤثر على تباينات معاملات الإنحدار التي نرغب في أن تكون أصغر ما يمكن لكي نحقق أقصى درجة من الدقة، فعلى سبيل المثال عند تقدير معاملات الإنحدار بإستخدام طريقة المربعات الصغرى (OLS) في ظل وجود ارتباط ذاتي بين حدود الخطأ أوعدم ثبات تباين حد الخطأ heteroskedasticity فإن مقدر OLS لن يكون له أقل تباين من بين كل المقدرات غير المتحيزة، وهذا بدوره يؤثر على دقة الإستدلالات الإحصائية الخاصة بمعاملات الإنحدار. مع ملاحظة أن مشكلتي الإرتباط الذاتي بين حدود الخطأ وعدم ثبات تباين حد الخطأ لن تؤثر على خصائص الإتساق وعدم التحيز لمقدرات معاملات الإنحدار. وفي حالة تجاهل وجود مشكلات التقدير، فإن هذا يتطلب استخدام طرق التقدير المتينة robust estimation لتقدير التباينات التقاربية لمقدرات معاملات الإنحدار. ولذلك، في

هذا البحث سوف يتم استخدام طريقتين من طرق التقدير المتينة للتباين التقاربي وهما؛ ؛ طريقة أساس كرنال kernel-based ومقدر OPG أو خوارزمية BHHH.

(٢-٣) تقدير التباين التقاربي لـ $\hat{\beta}$ باستخدام طريقة أساس كرنال Kernel-based

يمكن تقدير مصفوفة التغاير التقاربي في حالة البيانات المرتبطة ذاتيا بإستخدام الطرق اللامعلمية لتقدير التباين ومن بين هذه الطرق طريقة أساس كرنال kernel Based التي اقترحها Newey and West (1987) ونجد أن التقدير المتسق لمصفوفة التغاير التقاربي هو أمرهام لبناء فترات الثقة وإجراء اختبارات الفروض التقاربية وذلك لأن تحقيق الاتساق يجعلنا نقوم بإجراء الاستدلالات الإحصائية بالقرب من القيمة الحقيقية للمعلمة. ويتم استخدام دوال كرنال لتمهيد دالة التغاير الذاتي للعينة وذلك بحيث تقل قيمة الأوزان التي تستخدم لترجيح التغايرات الذاتية للعينة كلما زادت الفجوة الزمنية.

بالنسبة لنموذج انحدار بواسون لبيانات سلسلة زمنية الذي تمت مناقشته في القسم الثاني من هذا البحث، سوف نجد أنه بأخذ تفاضل دالة الإمكان اللوغاريتمي الزائفة the pseudo log-likelihood function بالنسبة لـβ ، فأننا سنحصل علي $\hat{\beta}$ من خلال حل معادلة التقدير المعممة التالية:

$$M_n^T \sum_{t=1}^{n} x_t \left(y_t - e^{x_t^T \beta} \right) = 0 \qquad (11)$$

حيث أن

$$U_t(\beta) = M_n^T x_t \left(y_t - e^{x_t^T \beta} \right)$$

قدمت دراسة Wu (2012) مقدر متسق لـ $\Omega_1^{-1}(\Omega_1 + \Omega_2)\Omega_1^{-1}$ والتي تمثل مصفوفة التغاير التقاربي لـ $\hat{\beta}$ بإستخدام طريقة أساس كرنال Kernel-based حيث أن المقدر المتسق لـΩ_1 نحصل عليه بالتعويض المباشر في الصيغة التالية:

$$\hat{\Gamma}_1 = M_n^T \left[\sum_{t=1}^{n} x_t\, x_t^T\, e^{x_t^T \hat{\beta}} \right] M_n \qquad (12)$$

وأن المقدر المتسق لـ $\Omega_1 + \Omega_2$ هو:

$$\hat{\Gamma}_n = \sum_{h=-n+1}^{n-1} \sum_{t=max(1-h\,,1)}^{min(n-h\,,n)} E\left[U_t(\hat{\beta})\, U_t^T(\hat{\beta})\right]$$

$$= \sum_{t=1}^{n} E\left[U_t(\hat{\beta})\, U_t^T(\hat{\beta})\right] + \sum_{h=1}^{n-1} \sum_{t=1}^{n-h} E\left[U_t(\hat{\beta})\, U_{t+h}^T(\hat{\beta})\right]$$

$$+ \sum_{h=-n+1}^{-1} \sum_{t=1-h}^{n} E\left[U_t(\hat{\beta})\, U_{t+h}^T(\hat{\beta})\right]$$

$$= \sum\nolimits_0 + \sum_{h=1}^{n-1} \left(\sum\nolimits_h + \sum\nolimits_h^T\right)$$

حيث أن:

$$\sum\nolimits_h = \sum_{t=1}^{n-h} E\left[U_t(\hat{\beta})\, U_{t+h}^T(\hat{\beta})\right]$$

وبإستخدام أوزان كرنال سوف نحصل علي المقدر التالي:

$$\hat{\Gamma}_n = \sum\nolimits_0 + \sum_{h=1}^{k-1} w(h/k) \left(\sum\nolimits_h + \sum\nolimits_h^T\right) \qquad (13)$$

حيث أن $w(\cdot)$ هي دالة كرنال و k هي (عرض الحزمة) bandwidth وهذا المقدر قائم علي فكرة بتر أو تقليل وزن الحدود غير المرغوب فيها، وبالتالي فأن $\hat{\Gamma}_1^{-1}\, \hat{\Gamma}_n\, \hat{\Gamma}_1^{-1}$ هو مقدر متسق لـ$\Omega_1^{-1}(\Omega_1 + \Omega_2)\Omega_1^{-1}$ والتي تمثل مصفوفة التغاير التقاربي لـ $\hat{\beta}$.

(٣-٣) تقدير التباين التقاربي لـ $\hat{\beta}$ باستخدام طريقة OPG أو خوارزمية BHHH

مصفوفة التغاير التقاربي لمقدر الإمكان الأعظم $\hat{\beta}$ هي مصفوفة في المعلمات β المجهولة (أى أنها دالة في المعلمات β المجهولة) ويمكن حسابها من خلال معكوس مصفوفة المعلومات كالتالي:

$$\mathrm{Var}(\hat{\beta}) = [I(\beta)]^{-1} \qquad (14)$$

وحيث أن مصفوفة المعلومات هي سالبة القيمة المتوقعة لمصفوفة هيسين Hessian (وهي مصفوفة التفاضلات الجزئية من الدرجة الثانية لدالة الإمكان اللوغاريتمي بالنسبة لـــ β) ، أى أن

$$I(\beta) = -E\big(H(\beta)\big)$$

$$= -E\left(\frac{\partial^2 \ln L\,(\beta)}{\partial\beta\,\partial\beta^T}\right) \tag{15}$$

وبالتالي فإن مصفوفة التغاير التقاربي لمقدر الإمكان الأعظم $\hat{\beta}$ ستكون:

$$Var\big(\hat{\beta}\big) = [I(\beta)]^{-1}$$

$$= \left[-E\left(H(\beta)\right)\right]^{-1}$$

$$= -\left[E\left(\frac{\partial^2 \ln L\,(\beta)}{\partial\beta\,\partial\beta^T}\right)\right]^{-1} \tag{16}$$

وبما أن متجه المعلمات الحقيقية β مجهولة، فإنه يمكننا استبدال β بـ $\hat{\beta}$ وذلك لتقدير مصفوفة التغاير التقاربي لمقدر الإمكان الأعظم، إلا أن هذا المقدر نادرا ما يمكن الحصول عليه وذلك لأن القيمة المتوقعة الدقيقة للتفاضلات الجزئية من الدرجة الثانية لدالة الإمكان اللوغاريتمي ستكون غير معروفة، وهناك بديل آخر لتقدير مصفوفة التغاير التقاربي لمقدر الإمكان الأعظم $\hat{\beta}$ عن طريق حساب سالب مصفوفة التفاضلات الجزئية من الدرجة الثانية لدالة الإمكان اللوغاريتمي بدون استخدام التوقع كالتالي:

$$\hat{I}\big(\hat{\beta}\big) = \left(-\frac{\partial^2 \ln L\,(\beta)}{\partial\beta\,\partial\beta^T}\right) \tag{17}$$

ولكن عيب هذا المقدر هو أن التفاضلات الجزئية من الدرجة الثانية في بعض الحالات يكون من المعقد اشتقاقها وبرمجتها على الكمبيوتر. وهناك بديل آخر لتقدير مصفوفة التغاير التقاربي لـ$\hat{\beta}$ وذلك من خلال استخدام خوارزمية BHHH او ما يطلق عليه outer product of gradients (OPG) estimator(مقدر حاصل الضرب الخارجي لمتجهات التفاضلات الجزئية من الدرجة الأولى) والتي قدمها(Berdent, Hall, Hall and Hausman (1974 والقائمة على أنه عند تحديد النموذج بشكل دقيق وعندما تكون القيمة المتوقعة للتفاضلات الجزئية من الدرجة الأولى مساوية للصفر (أى عند القيمة العظمى لدالة الإمكان) فإن سالب القيمة المتوقعة لمصفوفة هيسين Hessian سيكون مساوي لتباين التفاضلات الجزئية من الدرجة الأولى $var\,(g_i(\beta))$ والذي يسمى (the variance of gradients) في المجتمع حيث أن:

$$g_i(\beta) = \frac{\partial \ln L\,(\beta)}{\partial \beta}$$

والتي تعني أن $g_i(\beta)$ هو متجه kx1 للتفاضلات الجزئية من الدرجة الأولى لدالة الإمكان اللوغاريتمي. وبالنسبة للعينة، نجد أنه عند تحديد النموذج بشكل دقيق، فإن تباين التفاضلات الجزئية من الدرجة الأولى $\mathrm{var}\,(g_i(\hat{\beta}))$ سوف يكون تقريب جيد لـ $-E\,\left(H(\beta)\right)$ كلما زادت حجم العينة؛ أى أن $\mathrm{var}\,\left(g_i(\hat{\beta})\right) \to -E\,\left(H(\beta)\right)$ كلما $n\to\infty$ وذلك عند الحد الأقصى لدالة الإمكان (أى عندما يكون توقع التفاضلات الجزئية من الدرجة الأولى مساويا للصفر)

وبعبارة أخرى، فإن سالب القيمة المتوقعة لمصفوفة التفاضلات الجزئية من الدرجة الثانية يمكن تقديرها من خلال مصفوفة التغاير للتفاضلات الجزئية من الدرجة الأولى عندما تكون n كبيرة بما يكفي، وبالتالي فإن:

$$\begin{aligned}
\widehat{\mathrm{Var}}(\hat{\beta}) &= \left[\hat{I}(\hat{\beta})\right]^{-1} \\
&= \left[-E\,\left(H(\hat{\beta})\right)\right]^{-1} \\
&= \left[\mathrm{var}\,(g_i(\hat{\beta}))\right]^{-1} \\
&= \left[\frac{1}{n}\sum_{i=1}^{n} g_i(\hat{\beta})\, g_i(\hat{\beta})^{T}\right]^{-1} \qquad (18)
\end{aligned}$$

وهذا ما يسمى outer product of gradients (OPG) estimator (مقدر حاصل الضرب الخارجي لمتجهات التفاضلات الجزئية من الدرجة الأولى) أو خوارزمية BHHH.

وهذه الطريقة تستخدم في البحث التجريبي والقائمة على إجراء مجموعة من التكرارات للحصول على تباين التفاضلات الجزئية من الدرجة الأولى عند الحد الأقصى لدالة الإمكان، والتي سبق استخدامها في العديد من الدراسات من بينها دراسة Berliana, et al (2019) ودراسة Aini (2020) ودراسة Wenur and Suharsono (2020).

ونجد أنه عند تحديد النموذج بشكل دقيق فإن توزيع المعاينة للفرق ما بين المقدر والقيمة الحقيقية للمعلمات سوف يكون توزيع طبيعي بشكل تقاربي بمتوسط صفر، وتغاير يساوي معكوس سالب القيمة المتوقعة لمصفوفة هيسين Hessian

$$\sqrt{n}\left(\hat{\beta}-\beta\right) \xrightarrow{d} N\left(0\,,\left[-E\left(H(\beta)\right)\right]^{-1}\right)$$

كلما $n\to\infty$ حيث أن β هو متجه المعلمات الحقيقية و $\hat{\beta}$ هو مقدر الإمكان الأعظم و $\left[-E\left(H(\beta)\right)\right]^{-1}$ هي سالب القيمة المتوقعة لمصفوفة هيسين Hessian في المجتمع. ولكن اذا كان النموذج غير محدد بشكل دقيق misspecified، فإن التغاير التقاربي لـ $\hat{\beta}$ يكون أكثر تعقيدا، فنجد أنه بالنسبة لأي نموذج تكون فيه القيمة المتوقعة للتفاضلات الجزئية من الدرجة الأولى لدالة الإمكان اللوغاريتمي في المجتمع مساوية للصفر فإن:

$$\sqrt{n}\left(\hat{\beta}-\beta\right) \xrightarrow{d} N\left(0\,,\left[-E\left(H(\beta)\right)\right]^{-1} \mathrm{var}\left(g_i(\beta)\right)\left[-E\left(H(\beta)\right)\right]^{-1}\right)$$

حيث أن $\mathrm{var}\left(g_i(\beta)\right)$ هو تباين التفاضلات الجزئية من الدرجة الأولى لدالة الإمكان اللوغاريتمي بالنسبة لـ β في المجتمع.

وعندما يكون النموذج محدد بشكل دقيق، فإن $\mathrm{var}\left(g_i(\beta)\right) = -E\left(H(\beta)\right)$ وبالتالي فإن $\left[-E\left(H(\beta)\right)\right]^{-1} \mathrm{var}\left(g_i(\beta)\right)\left[-E\left(H(\beta)\right)\right]^{-1} = \left[-E\left(H(\beta)\right)\right]^{-1}$ وهذا ما يجعلنا نحصل على نفس التغاير التقاربي في حالة أن يكون النموذج محدد بشكل دقيق. إلا أنه إذا كان النموذج غير محدد بشكل دقيق أو أن هناك مخالفات لفروض النموذج مثل وجود إرتباط ذاتي أو عدم ثبات تباين حدود الخطأ سوف تكون صيغة التغاير التقاربي هي:
$\left[-E\left(H(\beta)\right)\right]^{-1} \mathrm{var}\left(g_i(\beta)\right)\left[-E\left(H(\beta)\right)\right]^{-1}$ ونجد أن هذه المصفوفة سوف تكون صالحة سواء كان النموذج محدد بشكل دقيق أم لا. وهذه الصيغة تسمى مصفوفة التغايرالمتينة Robust covariance matrix، وفي بعض الأحيان تسمي أيضا مقدر "sandwich" للتغاير، حيث أن معكوس سالب القيمة المتوقعة لمصفوفة هيسين Hessian تظهر على كلا الجانبين.

وبالنسبة لنموذج الإنحدار البواسوني الذي هو محل دراستنا، سيكون سالب متوسط مصفوفة هيسين في العينة كالتالي:

$$-\frac{1}{n}H\left(\hat{\beta}\right) = -\frac{1}{n}\left.\frac{\partial^2 \ln L\,(\beta)}{\partial\beta\,\partial\beta^T}\right|_{\hat{\beta}} = -\frac{1}{n}\sum_{t=1}^{n} x_t x_t^T e^{x_t^T\hat{\beta}} \qquad (19)$$

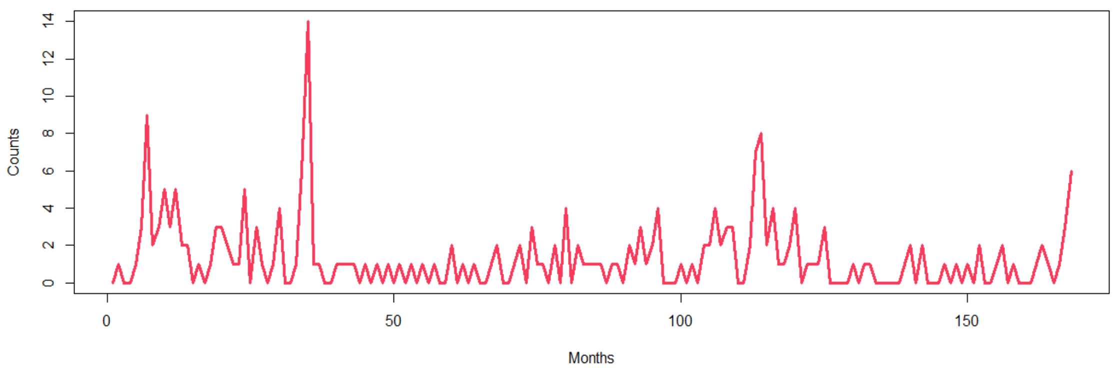
ويكون تباين التفاضلات الجزئية من الدرجة الأولى لدالة الإمكان اللوغاريتمي هي:

$$\mathrm{var}\left(\mathrm{g_i}(\hat{\beta})\right) = \mathrm{var}\left(\frac{\partial \ln \mathrm{L}\,(\beta)}{\partial \beta}\bigg|_{\hat{\beta}}\right)$$
$$= \mathrm{var}\left(\sum_{t=1}^{n} x_t \left(y_t - \mathrm{e}^{x_t^T \hat{\beta}}\right)\right) \qquad (20)$$

4 - الدراسة التطبيقية

(٤-١) التطبيق على الأعداد الشهرية للإصابة بشلل الأطفال:

تمثل هذه البيانات الأعداد الشهرية للإصابة بشلل الأطفال في الولايات المتحدة الأمريكية خلال السنوات من ١٩٧٠ الى ١٩٨٣ كما هو مسجل في مركز مكافحة الأمراض بالولايات المتحدة الأمريكية وبذلك يكون لدينا سلسلة زمنية تتكون من ١٦٨ مشاهدة، والتي سبق دراستها أيضا في العديد من الدراسات، منها ؛ Davis et al (1999) و Davis et al(2000) و Wu (2007) و Safari et al (2017). وتعتبر هذه البيانات مثال نموذجي عند دراسة النماذج الخطية المعممة لمعلمة مشتقة. ونجد أن أول من إستخدم هذه البيانات هي دراسة Zeger (1988) والذي افترض نموذج معلمة مشتقة parameter-driven model بحيث يكون التوزيع الشرطي للمشاهدات بمعلومية العملية الكامنة latent process هو توزيع بواسون، واستخدام نظرية معادلات التقدير من أجل تقدير معلمات النموذج.



الشكل (1-4) يمثل الأعداد الشهرية للإصابة بشلل الأطفال في الولايات المتحدة من ١٩٧٠ الى ١٩٨٣

الشكل (1-4) يوضح أن هناك بعض التغيرات الموسمية في البيانات وأن هناك إحتمالية لوجود اتجاه خطي متناقص خلال الزمن. ونجد أن الكشف عن وجود إتجاه خطي متناقص هو من الأهداف

الرئيسية لدراسة هذه البيانات، وعلى الرغم من أن القيمة المشاهدة في نوفمبر ١٩٧٢ وهي القيمة ١٤ تعتبر قيمة متطرفة outlier، إلا أنه لم يتم إستبعادها او تعديلها عند تحليل هذه البيانات وذلك لأن لها تأثير طفيف على النتائج وذلك وفقا لدراسة Chan and Ledolter (1995) حيث تم تعديل هذه القيمة المتطرفة ووجد أن هناك تغير طفيف في النتائج.

ونجد أنه في Davis et al (2000) تم دراسة هذه البيانات في إطار نموذج المعلمة المشتقة parameter-driven model وقد تم إستخدام طريقة الإمكان الزائف pseudo–likelihood method لتقدير معلمات النموذج. وذلك بفرض أن التوزيع الشرطي للمشاهدات بمعلومية العملية الكامنة latent process هو توزيع بواسون كالتالي:

$$Y_t \left| \alpha_t \overset{indep}{\sim} po(\lambda_t) \right.$$

وهذا هو نفس النموذج الذي سنقوم بتطبيقه والذي سبق توضيحه في القسم الثاني من هذا البحث. وسوف نقوم بإستخدام نفس متغيرات الإنحدار التي تم إستخدامها في الدراستين السابقتين وهذه المتغيرات هي حد التقاطع والإتجاه الخطي وحدود متسلسلة فورير Fourier السنوية والنصف سنوية للتعبير عن النمط الموسمي وبالتالي:

$$x_t = (1, t`/1000\ , \cos\ (2\pi t`/12), \sin(2\pi t`/12), \cos(2\pi t`/6), \sin(2\pi t`/6)\ )^T$$

حيث أن $t` = t - 73$ والمستخدمة في تحديد موقع حد التقاطع عند شهر يناير ١٩٧٦ . وهذا يعني أننا سوف نصل لنفس النتائج التي توصلت إليها دراسة Davis et al (2000) فيما يتعلق بالنموذج وتقدير الإنحدار إلا أننا في هذه الدراسة سنقوم بإستخدام طريقة أساس كرنال kernel-based ومقدرOPG أو خوارزمية BHHH والذي سبق توضيحهم في في القسم الثالث من هذا البحث وذلك لتقدير التباينات التقاربية لـ $\hat{\beta}$.

سوف نقوم بتوفيق النموذج الخطي المعمم البواسوني المعياري The standard Poisson GLM والذي سبق توضيحه في في القسم الثاني من هذا البحث لبيانات الأعداد الشهرية للإصابة بشلل الأطفال في الولايات المتحدة بإستخدام الدالة "glm.po" في البرنامج الإحصائي R وتم عرض النتائج في الجدول (1-4).

الجدول (1-4): تقديرات المعلمات لبيانات شلل الأطفال بالإعتماد على توفيق نموذج خطي معمم بواسوني معياري.

GLM	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	0.20694	0.07508	2.75	0.005849 **
Trend	-4.79866	1.40289	-3.421	0.000625 ***
cos(2πt`/12)	-0.14873	0.09722	-1.53	0.126037
sin(2πt`/12)	-0.53188	0.10904	-4.878	1.07e-06 ***

cos(2πt`/6)	0.1691	0.09881	1.711	0.087013 .
sin(2πt`/6)	-0.43214	0.1008	-4.287	1.81e-05 ***

تظهر تقديرات β في العمود (٢) من الجدول (1-4) وتظهر الأخطاء المعيارية لـ$\hat{\beta}$ في العمود (٣) من من الجدول (1-4) مع ملاحظة أن الأخطاء المعيارية لـ$\hat{\beta}$ المحسوبة في الجدول السابق تتجاهل إمكانية وجود العملية الكامنة latent process وبالتالي فإنه لا يمكن الإعتماد عليها في الإستدلالات الإحصائية الخاصة بمعاملات الإنحدار لأنها قد تؤدي إلى إستنتاجات خاطئة.

سوف يتم إستخدام إختبار Breusch-Pagan للكشف عن عدم ثبات تباين حدود الخطأ Heteroskedasticity ونجد أن إحصاءة الإختبار و P-value لهذا الإختبار هي:

Breusch-Pagan test

BP	Df	p-value
8.2776	**3**	**0.04061**

حيث أن P-value أقل من 0.05، فإن القرار هو رفض الفرض العدمي بأن تباينات الأخطاء متساوية. وهذا يعني أن هناك دليل كافي على عدم ثبات تباين حدود الخطأ في نموذج الإنحدار عند مستوى معنوية 0.05.

سوف يتم إستخدام إختبار Durbin-Watson للكشف عن وجود إرتباط ذاتي بين حدود الخطأ في النموذج. ونجد أن إحصاءة الإختبار و P-value لهذا الإختبار هي:

Durbin-Watson test

DW	p-value
1.5008	**0.0003086**

حيث أن P-value قيمة صغيرة جدا، فإن القرار هو رفض الفرض العدمي بأن الإرتباط الذاتي بين حدود الخطأ مساوي للصفر. وهذا يعني أن هناك دليل كافي على وجود إرتباط ذاتي بين حدود الخطأ في نموذج الإنحدار.

وبناءا على النتائج السابقة، فإنه لا يمكننا الإعتماد على الأخطاء المعيارية لـ$\hat{\beta}$ الناتجة عن توفيق نموذج خطي معمم بواسوني معياري في ظل تجاهل وجود الإرتباط الذاتي وعدم ثبات تباين حدود الخطأ. وللحصول على مقدرات متينة robust estimators للتباينات التقاربية لـ$\hat{\beta}$ والتي تسمح لنا

بالحصول على تقديرات متسقة يمكن استخدامها في الاستدلالات الإحصائية الخاصة بمعاملات الإنحدار، سوف يتم إستخدام طريقة أساس كرنال kernel-based والتي تم توضيحها في في القسم الثالث من هذا البحث، وذلك بالإعتماد على ثلاث دوال كرنال مختلفة وهم؛ Bartlett kernel و Parzen kernel وQuadratic Spectral kernel . كما أننا سنقوم بتقدير التباينات التقاربية لـ$\hat{\beta}$ بإستخدام مقدر OPG أو خوارزمية BHHH والتي سبق توضيحها في القسم الثالث من هذا البحث وتم عرض النتائج في الجدول (2-4) كما يلي:

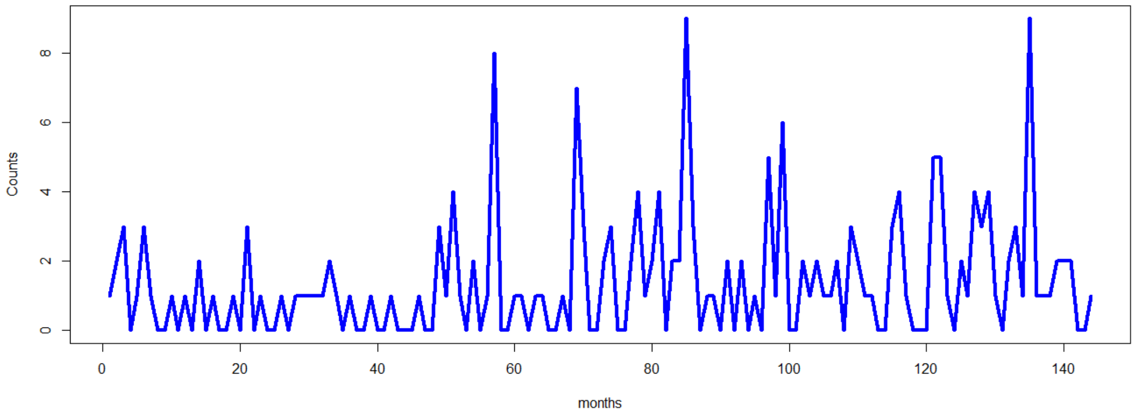
الجدول (2–4): تقديرات التباينات التقاربية لـ $\hat{\beta}$.

	orignal	Bartlett	Parzen	Quadratic Spectral	OPG
Intercept	0.0090	0.0139	0.0143	0.0145	0.0042
Trend	4.6735	8.2279	8.2636	8.3158	1.0058
cos(2πt`/12)	0.0173	0.0199	0.0205	0.0202	0.0060
sin(2πt`/12)	0.0235	0.0433	0.0435	0.0447	0.0079
cos(2πt`/6)	0.0180	0.0209	0.0212	0.0213	0.0058
sin(2πt`/6)	0.0206	0.0218	0.0221	0.0222	0.0064

ومن خلال الجدول السابق نجد أنه لا يوجد اختلافات جوهرية بين تقديرات أساس كرنال للتباينات التقاربية لـ$\hat{\beta}$ بإستخدام ثلاث دوال كرنال مختلفة، والتي تظهر في الأعمدة 2 و3 و4. في حين أن تقديرات OPG للتباينات التقاربية لـ$\hat{\beta}$ و التي تظهر في العمود 5 كانت أقل بفارق كبير عن التباينات التقاربية التي حصلنا عليها بإستخدام طريقة أساس كرنال، ولذلك فإننا نوصي بإستخدام مقدر OPG لتقدير التباينات التقاربية لـ$\hat{\beta}$ في ظل إستخدام نموذج خطي معمم بواسوني لتحليل بيانات سلاسل زمنية وذلك في ظل تجاهل مشكلات التقدير مثل وجود إرتباط ذاتي بين حدود الخطأ أو عدم ثبات تباين حدود الخطأ، حيث أن مقدر OPG يحقق أقل قيم للتباينات التقاربية.

(٤-2) التطبيق على الأعداد الشهرية لحالات الحمى الروماتيزمية:

تمثل هذه البيانات الأعداد الشهرية لحالات الحمى الروماتيزمية التي يتم استقبالها بمستشفى أطفال المنصورة خلال الفترة من ١ يناير ٢٠١٠ إلي ٣١ ديسمبر ٢٠٢١ .



الشكل (4-2) يمثل الأعداد الشهرية لإستقبال حالات الحمى الروماتيزمية في مستشفي أطفال المنصورة خلال الفترة من ١ يناير ٢٠١٠ إلي ٣١ ديسمبر ٢٠٢١ .

الشكل (4-2) يوضح أن هناك بعض التغيرات الموسمية في البيانات وأن هناك إحتمالية لوجود اتجاه خطي متناقص خلال الزمن. وسوف يتم دراسة هذه البيانات من خلال نموذج المعلمة المشتقة parameter-driven model وبفرض أن التوزيع الشرطي للمشاهدات بمعلومية العملية الكامنة latent process هو توزيع بواسون كالتالي:

$$Y_t \,\Big|\, \alpha_t \overset{indep}{\sim} po(\lambda_t)$$

والذي سبق توضيحه في القسم الثاني من هذا البحث ، وقد تم إستخدام طريقة الإمكان الزائف pseudo–likelihood method لتقدير معلمات النموذج. وقد أوضح التحليل المبدئي لهذه البيانات إلي الحاجة بأن يكون النموذج متضمن إتجاه خطي متناقص محتمل خلال الزمن وحدود متسلسلة فورير Fourier للتعبير عن النمط الموسمي في البيانات،ومن خلال اختبار معنوية معاملات الإنحدار وجد أن حد الإتجاه الخطي لم يكن معنوى، ولذلك تم استبعاده من النموذج،

وبالتالي فإن استقبال حالات الحمى الروماتيزمية في مستشفي أطفال المنصورة ليس له اتجاه متناقص معنوي خلال الفترة من ١ يناير ٢٠١٠ إلي ٣١ ديسمبر ٢٠٢١، في حين أن حدود متسلسلة فورير Fourier الربع سنوية والنصف سنوية للتعبير عن النمط الموسمي كان لها معاملات معنوية، وبالتالي فإن متغيرات الإنحدار هي:

$$x_t = (1\,, \cos\,(2\pi t/6), \sin(2\pi t/6), \cos(2\pi t/3), \sin(2\pi t/3)\,)^T$$

وتم توفيق النموذج الخطي المعمم البواسوني المعياري The standard Poisson GLM والذي سبق توضيحه في القسم الثاني من هذا البحث لبيانات الأعداد الشهرية لإستقبال حالات الحمى

الروماتيزمية في مستشفي الأطفال بالمنصورة بإستخدام الدالة "glm.po" في البرنـامج الإحصـائي R وتم عرض النتائج في الجدول (3-4).

الجدول (3-4): تقديرات المعلمات لبيانات الحمى الروماتيزمية بالإعتماد على توفيق نموذج خطي معمم بواسوني معياري.

GLM	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	0.16167	0.08412	1.922	0.05461 .
cos(2πt/6)	−0.34302	0.1035	-3.314	0.00092 ***
sin(2πt/6)	0.66176	0.12664	5.225	1.74e-07 ***
cos(2πt/3)	0.32825	0.10411	3.153	0.00162 **
sin(2πt/3)	0.24431	0.11478	2.128	0.03330 *

تظهر تقديرات β في العمود (٢) من الجدول (3-4) وتظهر الأخطاء المعيارية لـ$\hat{\beta}$ في العمود (٣) من من الجدول (3-4) مع ملاحظة أن الأخطاء المعيارية لـ$\hat{\beta}$ المحسوبة في الجدول السابق تتجاهل إمكانيـة وجـود العمليـة الكامنـة latent process وبالتـالي فإنـه لا يمكـن الإعتمـاد عليهـا فـي الإستدلالات الإحصائية الخاصة بمعاملات الإنحدار لأنها قد تؤدي إلى إستنتاجات خاطئة.

سوف يتم إستخدام إختبار Breusch-Pagan للكشف عن عدم ثبات تباين حدود الخطأ Heteroskedasticity ونجد أن إحصاءة الإختبار و P-value لهذا الإختبار هما:

Breusch-Pagan test

BP	Df	p-value
10.499	4	0.03281

حيث أن P-value أقل من 0.05، فإن القرار هو رفض الفرض العدمي بأن تباينات الأخطاء متساوية. وهذا يعني أن هناك دليل كافي على عدم ثبات تباين حدود الخطأ في نموذج الإنحدار عند مستوى معنوية 0.05.

وسوف يتم إستخدام إختبار Durbin-Watson للكشف عن وجود إرتباط ذاتي بين حدود الخطأ في النموذج. ونجد أن إحصاءة الإختبار و P-value لهذه الإختبار كالتالي:

Durbin-Watson test

DW	p-value
1.9775	0.446

حيث أن P-value أكبر من ٠.٠٥، فإن القرار هو عدم رفض الفرض العدمي بأن الإرتباط الذاتي بين حدود الخطأ مساوي للصفر. وهذا يعني أنه لا يوجد دليل كافي على وجود إرتباط ذاتي بين حدود الخطأ في نموذج الإنحدار.

وبناءا على النتائج السابقة، فإنه لا يمكننا الإعتماد على الأخطاء المعيارية لـ$\hat{\beta}$ الناتجة عن توفيق نموذج خطي معمم بواسوني معياري بسبب وجود إحدى مشكلات التقدير وهي عدم ثبات تباين حدود الخطأ. وللحصول على مقدرات متينة robust estimators للتباينات التقاربية لـ$\hat{\beta}$ والتي تسمح لنا بالحصول على تقديرات متسقة يمكن استخدامها في الإستدلالات الإحصائية الخاصة بمعاملات الإنحدار، سوف يتم إستخدام طريقة أساس كرنال kernel-based والتي تم توضيحها في القسم الثالث من هذا البحث، وذلك بالإعتماد على ثلاث دوال كرنال مختلفة وهم؛ Bartlett kernel و Parzen kernel وQuadratic Spectral kernel . كما أننا سنقوم بتقدير التباينات التقاربية لـ$\hat{\beta}$ بإستخدام مقدر OPG أو خوارزمية BHHH والتي سبق توضيحها في القسم الثالث من هذا البحث وتم عرض النتائج في الجدول (4-4) كما يلي:

الجدول (4–4): تقديرات التباينات التقاربية لـ $\hat{\beta}$.

	orignal	Bartlett	Parzen	Quadratic Spectral	OPG
Intercept	0.0092	0.0099	0.0104	0.0099	0.0058
cos(2πt/6)	0.0174	0.0206	0.0184	0.0183	0.0085
sin(2πt/6)	0.0213	0.0151	0.0148	0.0147	0.0149
cos(2πt/3)	0.0168	0.0184	0.0197	0.0192	0.0095
sin(2πt/3)	0.0208	0.0157	0.0155	0.0154	0.0099

ومن خلال الجدول السابق نجد أنه لا يوجد اختلافات جوهرية بين تقديرات أساس كرنال للتباينات التقاربية لـ$\hat{\beta}$ بإستخدام ثلاث دوال كرنال مختلفة، والتي تظهر في الأعمدة 2 و3 و4. في حين أن تقديرات OPG للتباينات التقاربية لـ$\hat{\beta}$ و التي تظهر في العمود 5 كانت أقل بفارق كبير عن التباينات التقاربية التي حصلنا عليها بإستخدام طريقة أساس كرنال بإستثناء تقدير التباين التقاربي لـ$\hat{\beta}_3$ حيث كان التقدير بإستخدام مقدر OPG له قيمة قريبة إلى حد كبير من القيم التي حصلنا عليها بإستخدام طريقة أساس كرنال ولذلك فإننا نوصي بإستخدام مقدر OPG لتقدير التباينات التقاربية

لـ$\hat{\beta}$ في ظل إستخدام نموذج خطي معمم بواسوني لتحليل بيانات سلاسل زمنية عندما يتم تجاهل وجود إحدى مشكلات التقدير مثل وجود إرتباط ذاتي بين حدود الخطأ او عدم ثبات تباين حدود الخطأ، حيث أن مقدر OPG يحقق أقل قيم للتباينات التقاربية.

٥ – النتائج والتوصيات

(٥-١) النتائج:

قامت هذه الدراسة بالمقارنة بين طريقتين من الطرق المتينة لتقدير التباين التقاربي لـ $\hat{\beta}$ وهما طريقة أساس كرنال Kernel-based التي تم اقتراحها في دراسة Wu(2012) ومقدر OPG أو خوارزمية BHHH التي نقترح إستخدامها في هذه الدراسة وذلك في ظل نموذج الإنحدار البواسوني لتحليل بيانات سلسلة زمنية، ومن خلال التطبيق العملي علي مجموعتين من البيانات توصلت الدراسة إلى النتائج التالية:

- عند استخدام طريقة أساس كرنال Kernel-based لتقدير التباين التقاربي لـ $\hat{\beta}$ بالإعتماد على ثلاث دوال كرنال مختلفة وهم؛ Bartlett kernel و Parzen kernel و Quadratic Spectral kernel، قد توصلت الدراسة إلى أن استخدام أيا من الدوال الثلاثة سوف يعطي نتائج متقاربة إلى حد كبير عند تقدير التباين التقاربي لـ $\hat{\beta}$.

- بمقارنة مقدر OPG أو خوارزمية BHHH التي تم اقتراحها في هذه الدراسة مع طريقة أساس كرنال Kernel-based التي تم اقتراحها في دراسة Wu(2012)، قد تبين أن مقدر OPG أو خوارزمية BHHH هي الأفضل في تقدير التباين التقاربي لـ $\hat{\beta}$، حيث اتضح من خلال التطبيق العملي على مجموعتين من البيانات أن تقديرات التباينات التقاربية لـ $\hat{\beta}$ بإستخدام مقدر OPG كانت أقل بفارق كبير عن تقديرات التباينات التقاربية لـ$\hat{\beta}$ التي حصلنا عليها بإستخدام طريقة أساس كرنال Kernel-based.

(٥-٢) التوصيات:

في هذه الدراسة تم استخدام طريقتين من طرق التقدير المتينة وهما طريقة أساس كرنال Kernel-based ومقدر OPG أو خوارزمية BHHH لتقدير التباين التقاربي في نموذج الإنحدار البواسوني لبيانات سلسلة زمنية وذلك في ظل وجود مخالفات لفروض النموذج مثل الإرتباط الذاتي بين حدود الخطأ وعدم ثبات تباين حدود الخطأ مع التطبيق على مجموعتين من البيانات. ويوصي الباحث استخدام مقدر OPG أو خوارزمية BHHH لتقدير التباين التقاربي لـ $\hat{\beta}$ عند استخدام هذا النوع من النماذج حيث أن هذه الطريقة تحقق أقل قيم لتقدير التباين التقاربي.

المراجع:

1. Aini, Q. "Bivariate zero inflated generalized Poisson regression model in the number of pregnant maternal mortality and the number of postpartum maternal mortality in the Central Java Province in 2017." *In Journal of Physics: Conference Series (Vol. 1511, No. 1, p. 012055). IOP Publishing*, 2020.

2. Berliana, S, Purhad, Sutikno, and S Rahayu. "Multivariate generalized Poisson regression model with exposure and correlation as a function of covariates: Parameter estimation and hypothesis testing." *In AIP Conference Proceedings (Vol. 2192, No. 1, p. 090001).* AIP Publishing LLC, 2019, December.

3. Berndt, E.R., B.H. Hall, R.E. Hal, and Hausman. "Estimation and inference in nonlinear structural models." *In Annals of Economic and Social Measurement, Volume 3, number 4*, 1974: 653-665.

4. Carlstein , E. "The use of subseries values for estimating the variance of a general statistic from a stationary sequence." *Annals of Statistics 14*, 1986: 1171-1179.

5. Chan , K.S., and J. Ledolte. "Monte Carlo EM estimation for time series models involving counts." *Journal of the American Statistical Association, 90(429)*, 1995: 242-252.

6. Cox , D.R. "Statistical analysis of time series: some recent developments." *Scandinavian Journal of Statistics 8*, 1981: 93–115.

7. Davis, R.A. , W.T.M. Dunsmuir, and Y. Wang,. "On autocorrelation in a Poisson regression model." *Biometrika 87*, 2000: 491–505.

8. Davis, R.A., and R. Wu. "A negative binomial model for time series of counts." *Biometrika 96*, 2009: 735–749.

9. Davis, R.A., W.T.M. Dunsmuir, and Y. Wang. "Modeling time series of count data, , 1999, pp." *in: S. Ghosh (Ed.), Asymptotics, Nonparametrics and Time Seriesc, MarcelDekker, New York*, 1999: 63–114.

10. Newey, W.K., and K.D. West. "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix." *Econometrica 55*, 1987: 703–708.

11. Safari, A., R.M. Altman, and B Leroux. "Parameter-driven models for time series of count data.." *arXiv preprint arXiv:1711.02753.*, 2017.

12. Wenur, G.H., and A. Suharsono. "Three-parameter bivariate gamma regression model for analyzing under-five mortality rate and maternal mortality rate." *In Journal of Physics: Conference Series (Vol. 1538, No. 1, p. 012054). IOP Publishing*, 2020.

13. Wu, R. "Estimation for Some Linear and Nonlinear Time Series Models." *ProQuest*, 2007.

14. Wu, R. "On variance estimation in a negative binomial time series regression model ." *Journal of Multivariate Analysis, 112*, 2012: 145-155.

15. Zeger , S.L. "A regression model for time series of counts." *Biometrika 75*, 1988: 621–629.