Journal of Economics and Administrative Sciences Vol.26 (NO. 120) 2020, pp. 426-439

Journal of Economics and Administrative Sciences (JEAS)

Available online at http://jeasiq.uobaghdad.edu.iq

مقارنة التنبؤ بأنموذج الأوساط المتحركة التكاملية من الرتبة الأولى IMA(1,1) مع طريقة الأوساط المتحركة الموزونة آسيا

م .د. علي سلمان حبيب

جامعة سومر / كلية الادارة والاقتصاد

a.habib@uos.edu.iq

Received: 23/11/2019 **Accepted :7/1/2020** **Published :June / 2020**

مستخلص البحث:

تعـّـد عملية التنبؤ لبعض الظواهر المرتبطة بالزمن وبشكل خاص نماذج الانحدار الذاتي والاوساط المتحركة التكاملية (ARIMA) من المواضع المهمة في تحليل السلاسل الزمنية لعلم الإحصاء التطبيقي وأن اهميتها تكمن في المراحل الأساسية في تحليل هيكليتها أو نمذجتها والشروط الواجب توفرها في العملية العشوائية . البحث تناول طريقتين للتنبؤ الأولى كانت حالة خاصة من نماذج الانحدار الذاتي والاوساط المتحركة التكاملية والتي هي (ARIMA(0,1,1)) و عندما تكون قيمة المعلمة تساوي صفر يمكن اعتباره أنموذج المشي العشوائي (Random Walk) ، والثانية كانت الأوساط المتحركة الموزونة أسياً (Exponential Weighted Moving Average(EWMA)) وتم التطبيق في بيانات الحوادث المرورية الشهرية في محافظة ذي قار للفترة من شهر كانون الثاني لعام 2011 ولغاية شهر آب لعام 2019 .وتبين من خلال البحث أن أنموذج ARIMA(0,1,1) يمثل الحوادث المرورية بشكل جيد ويمكن التنبؤ بحوادث هذه الظاهرة الخطيرة باستعمال هذا الأنموذج والحد من تفاقمها من خلال وضع خطط استراتيجية للطرق .

المصطلحات الرئيسة للبحث: **المشي العشوائي ، عملية الأوساط المتحركة التكاملية ، الأوساط المتحركة الموزونة اسياً ، التنبؤ .**

1.**الجانب النظري** :

1.1 **مشكلة البحث** :

دراسة حالة خاصة من نماذج التنبؤ في نمذجة الانحدار الذاتي والاوساط المتحركة التكاملية (ARIMA(p,d,q والتي هي العملية IMA(1,1) وامكانية تطبيق هذه الحالة الخاصة في بيانات حقيقية وبالتالي التنبؤ الدقيق بالأحداث المستقبلية للظاهرة المدروسة .

1.2 **أهمية البحث** :

تكمن أهمية البحث في استعمال نمذجة السلاسل الزمنية بطريقة (Box&Jenkins) من خلال مراحل التشخيص والتقدير والتنبؤ لبيانات حقيقية ثم مقارنتها مع التنبوءات بإحدى طرق التمهيد الخطي المفرد وهي الأوساط المتحركة الموزونة اسياً .

3.1 **هدف البحث** :

تحليل هيكلية انموذج ARIMA(0,1,1) والذي يعبر عن عملية المشي العشوائي مضاف له حد الخطأ بالمعلمة (θ) في حالة السلسلة الزمنية غير المستقرة من الرتبة الأولى وكيفية تمثيلهُ لأنموذج الاوساط المتحركة الموزون اسياً وتطبيقهُ في بيانات الحوادث المرورية الشهرية في محافظة ذي قار للفترة من شهر كانون الثاني 2011 ولغاية شهر آب 2019 .

2.**تعريف السلسلة الزمنية**: Time Series [1]

تعرف السلسلة الزمنية المفردة بأنها مجموعة من المشاهدات لمتغير واحد فقط مثل (y) وعدد المشاهدات فيها محدد مثل (T) ويرمز لها بالرمز $(y_t \ , t = 1,2,\dots,T)$ وأنموذج السلسلة الزمنية المفردة هو عبارة عن صيغة في حدود القيم الماضية للمتغير (y_t) وهو موجبة بالنسبة للزمن [Box&Jenkins:1976] . وتعرف السلسلة الزمنية ايضاً على أنها عملية عشوائية تمتلك فضاء معلمة وفضاء حالة ومرتبطة بالزمن ، وتكون الفترات زمنية على نوعين الأولى فترات منتظمة وتسمى متقطعة وهي التي نهتم بها وتكون على شكل (أيام ، اسابيع ، أشهر ، .الخ) والأخرى فترات زمنية غير منتظمة وهذه لا يمكن السيطرة عليها بشكل دقيق.

3 . **استقراريه السلسلة الزمنية** Stationary of time series

يعتبر موضوع تحليل السلاسل الزمنية من المواضيع المهمة المستخدمة في أدبيات النظرية الأحصائية لكن ما يعقد تحليل السلسلة الزمنية هو موضوع البحث في استقرارية السلسلة حيث هنالك نوعين منها الأولى سلسلة زمنية مستقرة والثانية سلسلة زمنية غير مستقرة لذلك فأن بيانات الظاهرة لهذا النوع من السلاسل تحتاج معالجة مسبقة لجعلها مستقرة (فضلا عن أن طرق التمهيد الأسي لا تحتاج الى إيجاد الأستقرارية) لذلك يجب عمل بعض المتطلبات منها :

1-رسم السلسلة الزمنية للظاهرة تحت البحث وملاحظة مشاهداتها .
2-حذف بعض المركبات الأساسية (مركبة الاتجاه ، المركبة الموسمية ....الخ) من النموذج .
3-اذا كانت السلسلة غير مستقرة في التباين يمكن أعادة قياس بيانات السلسلة الزمنية من خلال أجراء بعض التحويلات عليها على سبيل المثال أخذ التحويل اللوغارتمي أو التحويل الآسي للبيانات ...الخ.
4-اذا كانت السلسلة غير مستقرة في الوسط الحسابي يمكن أخذ الفروق differences ويرمز لها (d=1,2,…,6) للسلسلة الزمنية لتصبح مستقرة . وهنالك نوعين من الاستقرارية للسلسلة الزمنية وهما :

3.1 **الأستقرارية الضعيفة**: weak stationarity [9,6]

السلسلة الزمنية $\{y_t, \ t \in Z\}$ حيث أن Z هي مجموعة الأعداد الصحيحة ، نقول بأن تكون مستقرة أو تمتلك استقرارية ضعيفة (Stationarity or weakly stationary) إذا توفرت الشروط الآتية :

$$I - E(y_t^2) < \infty \quad \forall t \in Z.$$
$$II - E(y_t) = \mu \quad \forall t \in Z.$$
$$III - \gamma_y(k) = \gamma_y(t, t+k) = Cov(y_t \, , y_{t+k}) \quad \forall \, t,k \in Z.$$

وهذا يعني أن السلسلة الزمنية تمتلك أختلافات محدودة , والعزم الأول يكون ثابت ولا يعتمد على الزمن ، والعزم الثاني (دالة التباين المشترك) تعتمد فقط على الفجوة k) ولا تعتمد على الزمن (t) . وعندما تكون السلسلة مستقرة فأن دالة التباين المشترك تكون دالة زوجية وهذا يعني : $\gamma_y(k) = \gamma_y(-k)$. ويعرف التباين المشترك الذاتي في موضوع تحليل السلاسل الزمنية هو التباين المشترك بين قيم المتغير y_t عند فترة زمنية مثل t ، والمتغير y_{t+k} القيم عند فترة زمنية أخرى مثل t+k ، ويسمى التباين المشترك عند التأخر k (Lag k) . والصيغة الرياضية هي كالآتي[29. Montgomery,et.al:2008,p**]:**

$$\gamma(k) = Cov(y_t, y_{t+k}) = E\,\{(y_t-\mu)(y_{t+k}-\mu)\} \qquad (1)$$

لمجموعة من القيم الى k هذا يعني ... $\gamma(k)\ ,\ k = 0,1,2,$ ، ونلاحظ أن التباين المشترك عند التأخر صفر {Lag k=0} يكون تباين السلسلة الزمنية وهذا يعني $\gamma(0) = \sigma^2_{y_t}$. أن دالة الارتباط الذاتي تقيس الاعتماد الخطي بين y_t و y_{t+k} كذلك أن دالة الارتباط الذاتي عند التأخر K (Lag k) تعطى وفق الصيغة الرياضية الآتية :

$$\rho(k) = \frac{cov(y_t,y_{t+k})}{Var(y_t)} = \frac{E\,\{(y_t-\mu)(y_{t+k}-\mu)\}}{\sqrt{E\,\{(y_t-\mu)^2\}E\{(y_{t+k}-\mu)^2\}}} = \frac{\gamma(k)}{\gamma(0)} \qquad (2)$$

ولمجموعة من قيم k هذا يعني ... $\rho(k)\ ,\ k = 0,1,2,$ وتسمى دالة الارتباط الذاتي ونلاحظ من خلال التعريف أن $\rho_0 = 1$ كذلك فأن دالة الارتباط الذاتي تكون مستقلة عن قياسات السلسلة الزمنية وهي دالة متماثلة حول الصفر وكذلك فأنها تمتلك خواص الدوال الزوجية وهذا يعني $\rho(k) = \rho(-k)$.وبالتالي أذا كانت السلسلة الزمنية تمتلك متوسط وتباين ثابتين وتباين مشترك يعتمد فقط على التأخر الزمني فأنه يمكن القول أنها مستقرة من الرتبة الثانية وهي استقرارية ضعيفة.

3.2 الأستقرارية القوية (Strictly Stationarity)[9,8] :

السلسلة الزمنية $\{y_t,\ t \in Z\}$ نقول بأن تكون مستقرة بقوة (Strict Stationary) إذا توفر الشرط الآتية ، أن التوزيع الاحتمالي المشترك للمتغيرات $(y_{t_1}, y_{t_2}, \dots\dots, y_{t_p})$ يكون نفسهُ للمتغيرات $(y_{t_1+k}, y_{t_2+k}, \dots\dots, y_{t_p+k})$ وهذا يعني أن التوزيع المشترك يعتمد فقط على (K) ولا يعتمد على الفترات الزمنية $(t_1, t_2, \dots.., t_p)$ نلاحظ أن التباين محدود لا يفترض أن يكون معـّرف في الآستقرارية القوية لذلك فان الاستقرارية القوية ربما لا توضح الاستقرارية الضعيفة [25. Ruey S.T.:2005,p**]. إن التوزيع الاحتمالي المشترك لكل المشاهدات عند كل الفترات سيكون توزيع طبيعي وهذا الشرط ضروري لأن تكون السلسلة ذات استقرارية تامة [1]، اما بالنسبة الى معامل الارتباط الذاتي الجزئي فهو يقيس العلاقة بين البواقي لل y_{t+k} والـ y_t . ومن الأساليب المهمة للكشف عن وجود مشكلة عدم الاستقرارية في السلسلة الزمنية هو اختبار جذر الوحدة .**

4. طريقة الاوساط المتحركة الموزونة اسياً Exponential Weighted Moving Average (Method(E.W.M.A

وهي طريقة تتعامل مع البيانات بشكل مباشر من خلال معامل تمهيد وهي لا تحتوي على حد عشوائي وهي أحدى طرائق التمهيد الخطي المفرد وغالباً ما يكون أداءهُ بشكل جيد وأفضل من بقية النماذج الإحصائية عندما تكون حالات التنبؤ لفترات قصيرة نسبياً (Short-Term) كما في [271. Geller & Wilson:1978,P**] ومعادلة التمهيد هي**

$$\hat{y}_{t+1} = \lambda y_t + (1-\lambda)\hat{y}_t \qquad (3)$$

حيث أن : y_t تمثل القيمة المشاهدة عند الفترة الزمنية t، λ ثابت التمهيد وتكون قيمتهُ ($0 \le \lambda \le 1$).
$\hat{y}_t$ تمثل التنبؤ عن عند الفترة الزمنية t، $\hat{y}_{t+1}$ التنبؤ عند الفترة الزمنية . ($t+1$)
كذلك يمكن استعمال نفس الصيغة اعلاه للتنبؤ للفترة t كما في [166. Granger&Newbold:1986,P**]:**

$$\hat{y}_t = \lambda y_t + (1-\lambda)\hat{y}_{t-1} \qquad (4)$$

وعند التعويض المستمر لفترات زمنية أخرى وتعويض الصيغة (4) في (3) تنتج صيغة أخرى لأنموذج الاوساط المتحركة الموزونة اسياً بدلالة المشاهدات السابقة للسلسلة الزمنية الموزونة بمعاملات آسية وكالآتي:

$$\hat{y}_{t+1} = \lambda y_t + \lambda(1-\lambda)y_{t-1} + \lambda(1-\lambda)^2 y_{t-2} + \cdots. \qquad (5)$$

وطرق التمهيد لا تتطلب استقرارية للبيانات . وعندما تكون السلسلة الزمنية مستقرة فأن الوسط الحسابي والتباين المشترك والتباين[415. Cox,D.R.:1961,P**] تكون كالآتي :**

$$E(y_t) = \mu\ \ ,\ \ Cov(y_t, y_{t+k}) = \rho_k\rho^2 \quad and \quad Var(y_t) = \sigma^2$$

[1] ولمعرفة أن دالة التباين المشترك تمتلك الشرط الكافي والضروري لأثبات استقرارية السلسلة الزمنية مراجعة المصدر [1] .

وفي بعض التطبيقات العملية للأنموذج فأن ثابت التمهيد (λ) يسمى عامل التمهيد ويكون عامل تناقصي وبشكل آسي انظر[4.Gabrielsen,A.& et.al:2015,P] ويعبر عن الأنموذج بصيغة حد الخطأ وكالآتي:

$$y_t = +\varepsilon_t \quad , \quad \varepsilon_t \sim iid(0, \sigma^2)$$

وعندما يكون ثابت التمهيد ($(\lambda = (1-\theta)$) يمكن كتابة الأنموذج في الصيغة (5) كما في [414.Cox,D.R.:1961,P] وبشكل عام كالآتي:

$$\hat{y}_t = (1-\theta)\sum_{r=0}^{\infty} \quad \theta^r y_{t-r} \quad . \quad where \ |\theta| < 1 \qquad (6)$$

اما بالنسبة الى الصيغة الرياضية التي تكون قابلة للتطبيق العملي فقد ذكرها [326.Winters,P.R.:1960,P] وهي :

$$\hat{y}_t = (1-\theta)\sum_{r=0}^{M} \quad \theta^r y_{t-r} + \theta^{M+1}\hat{y}_b \qquad (7)$$

حيث أن $\hat{y}_b$: تمثل القيمة الأبتدائية للتنبؤ (Initial value).

وعندما تكون (M) كبيرة الى حد ما في الصيغة (7) وأخذ التوقع الرياضي فأن الحد ($\theta^{M+1}\hat{y}_b$) يقترب من الصفر ومعامل الحد الأخر ($((1-\theta)\sum_{r=0}^{M} \quad \theta^r$) يقترب من الواحد الصحيح ودرجة التقريب تعتمد على قيمتي (M و θ) ليكون($E(\hat{y}_t) \cong E(y_t)$

4. نمذجة السلسلة الزمنية Modeling of time series

تبدأ عملية نمذجة السلسلة الزمنية بعد عملية إيجاد سلسلة مستقرة ويتم خلال هذه العلمية بناء أنموذج رياضي محدد بشكل مضبوط و دقيق من حيث نوع الأنموذج ورتبتهُ وهنالك عدة طرق لتشخيص الأنموذج وتحديد رتبتهُ لعل أهمها دالتي الارتباط الذاتي ((Autocorrelation(ACF) والارتباط الذاتي الجزئي ((Partial Autocorrelation(PACF)انظر [368.Tseng&Tzeng:2002,P]أو استعمال معايير المعلومات مثل ((Akaike's information criterion(AIC) أو (Bayes information criterion ((BIC ....الخ) وبالتالي يكون الأنموذج المشخص يمثل الظاهرة قيد الدراسة وعندها يتم تقدير معالم الأنموذج باحدى طرق التقدير المعروفة ، و المرحلة الأخيرة من النمذجة هي التحقق من دقة الأنموذج من خلال أختبار مربع كاي ولعل أحصاءة [Box&Pierce(1970)] أو أحصاءة [Ljung&Box(1978)] التي تتوزع هي ايضا مربع كاي وعند هذا الحد قد تمت عملية التحقق من الأنموذج بشكل دقيق للسلسلة الزمنية [Ruey:2005]

5. بناء أنموذج ARIMA :

5.1 هيكلية الأنموذج :

أن الصيغة العامة لهذا لأنموذج هي (ARIMA(p,d,q وتعتبر صيغة عامة الى العديد من نماذج تحليل السلاسل الزمنية الخطية حيث يتم من خلاله التعرف على بعض النماذج الخاصة التي لاقت رواجا واسعا عند الباحثين لأنها تمتلك هيكلية رياضية تطابق بعض الظواهر المهمة في مجالات الحياة المختلفة التي نعيشها اليوم . أن التعرف على هيكلية عملية ARIMA يتطلب استراتيجية اختيار الأنموذج بشكل دقيق ومنها التعرف على بعض الطرق المحددة للرتبة وبالتالي معرفة الخصائص الكامنة في الأنموذج التي يمتلكها من الاستقراريه وقابلية الانعكاس ومن ثم استخدامها في عملية التنبؤ بالاعتماد على مجموعة اساسية من النماذج والتي تكتب بالصيغة الآتية [62.Terence:2019,P] :

$$\emptyset(B)\nabla^d y_t = \theta_0 + \theta(B)\varepsilon_t \qquad (8)$$

حيث أن :

$$\emptyset(B) = 1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p \quad and \ \theta(B)$$
$$= 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

وأن (B) تمثل عامل الأزاحة الخلفي ويعرف رياضياً بالشكل التالي ($By_t = y_{t-1}$) ويستعمل هذا العامل في إيجاد الفروق عندما تكون السلسلة الزمنية غير مستقرة وتعاني من مشكلة الأتجاه (Trend) والفرق الأول يكون بالشكل التالي [: Montgomery&et.al.2008, P.37]

$$\nabla y_t = (1-B)y_t = y_t - y_{t-1} \ .$$

ومن بين النماذج الخاصة التي في صدد تحليلها هو النموذج الأتي ARIMA(0,1,1) أنظر[189.Tseng&Adams:1994,P] وتحدث هذه الحالة عندما تكون السلسلة الزمنية غير مستقرة وتعاني من مشكلة الأتجاه وعند أخذ الفرق الأول (Difference) تتحول العملية الى مستقرة حسب درجة الفروق وعندما تكون مستقرة في الفرق الأول هذا يعني (d=1) وفي هذا الأنموذج تكون رتبة الانحدار الذاتي (p=0)

والأوساط المتحركة من الرتبة الأولى (q=1) وبالتالي نحصل على هكذا أنموذج ، وفي كلام أخر سيكون الأنموذج بالشكل التالي IMA(1,1) وهو يمتلك قدرة تمثيلية مفيدة للعديد من الظواهر الحقيقية .ومعادلة الفروق لهُ كالآتي [Box&Jenkins:1976]:

$$\nabla y_t = (1-\theta B)\varepsilon_t \quad for \quad -1<\theta<1 \qquad (9)$$

حيث أن : $\nabla = (1-B)$

ويمكن التعبير عن عملية الأوساط المتحركة (MA) بدلالة سلسلة الأخطاء العشوائية كأنموذج Random Walk كما في [Tsay:2005,p.64].

5.2 خاصية الانعكاس للأنموذج Inverted form of the model

يمكن كتابة الانموذج في المعادلة (9) بالشكل التالي:

$$y_t = \varphi(B)\varepsilon_t \quad \Longrightarrow \varphi^{-1}(B)y_t = \varepsilon_t$$

$$\pi(B)y_t = \varepsilon_t \qquad where: \quad \pi(\beta) = 1 - \sum_{j=1}^{\infty} \pi_j \beta^j$$

$$y_t = \sum_{j=1}^{\infty} \pi_j \beta^j y_t + \varepsilon_t \Longrightarrow \quad y_t = \sum_{j=1}^{\infty} \pi_j y_{t-j} + \varepsilon_t \qquad (10)$$

لذلك فان معاملات y_t هي مجموعة اوزان غير محدودة للقيم السابقة , مضاف اليها حد الاضطراب العشوائي ، وبالتالي يمكن كتابة العملية من الرتبة الاولى بالشكل الآتي :

$$(1-B)y_t = (1-\theta B)\,\varepsilon_t \quad \Longrightarrow \; y_t - y_{t-1} = \varepsilon_t - \theta\varepsilon_{t-1}$$

$$y_t = y_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1} \qquad (11)$$

والصيغة اعلاه هي صيغة (Random Shock) الصدمة العشوائية [Box&Jenkins:1976]:

ويمكن إعادة كتابة الصيغة بدلالة الخطأ العشوائي ε_t من خلال الآتي :

$$\nabla y_t = (1-\theta B)\,\varepsilon_t$$

نلاحظ أن معامل حد الخطأ العشوائي هو $(1-\theta B)$ وبالتالي يمكن إعادة صياغته بالشكل التالي :

$$(1-\theta B) = (1-\theta B + B - B\,)$$

$$(1-\theta B) = (1-\theta)B + (1-B) \Longrightarrow (1-\theta B) = (1-\theta)B + \nabla$$

$$(1-\theta B) = \lambda B + \nabla$$

حيث أن ($\lambda = (1-\theta)$)والعملية تكون قابلة للقياس في حدود λ علماً أن ($0<\lambda<2$)

وهنا يمكن أعادة صياغة العملية (11) لتصبح كالآتي :

$$\nabla y_t = \lambda\varepsilon_{t-1} + \nabla\varepsilon_t$$

وبضرب الطرفين بــ ($\nabla^{-1} = S$) على العكس من عملية الفروق ينتج الآتي :

$$y_t = \lambda\nabla^{-1}\varepsilon_{t-1} + \varepsilon_t \qquad \Longrightarrow y_t = \lambda\, S\varepsilon_{t-1} + \varepsilon_t$$

يمكن التعويض عن الصيغة الأخيرة بصيغة عامة وكالاتي :

$$y_t = \sum_{j=0}^{\infty} \omega_j \varepsilon_{t-j}\;. \quad for \;\; \omega_0 = 1 \;\; and \;\; \omega_j = \lambda \quad ; \forall\, j \geq 1 \qquad (12)$$

حيث أن (ω_j) تمثل الأوزان .

$$y_t = \omega_0\varepsilon_{t-0} + \omega_1\varepsilon_{t-1} + \omega_2\varepsilon_{t-2} + \cdots \ldots \ldots$$

$$y_t = \varepsilon_t + (\omega_1 B^1 + \omega_2 B^2 + \cdots \ldots)\varepsilon_t$$

وباستخدام خاصية التقريب في المتسلسلات يمكن إعادة كتابة الصيغة أعلاه بالشكل التالي :

$$y_t = \left(\sum_{i=1}^{\infty} \omega_i B^i\right)\varepsilon_t \quad \Longrightarrow \quad y_t = \left(1 - \sum_{i=1}^{\infty} \omega_i \beta^i\right)\varepsilon_t$$

$$y_t = \omega(B)\varepsilon_t \quad \Longrightarrow \;\; \omega(B)^{-1}y_t = \varepsilon_t$$

$$\pi(B)y_t = \varepsilon_t \qquad where \quad \pi(B) = \omega(B)^{-1} = 1 - \sum_{i=1}^{\infty} \pi_i \beta^i$$

حيث ان $\sum_{i=1}^{\infty} |\pi_i| < \infty$.وبالتالي فأن الأوساط المتحركة لرتبة محدودة تمثل عملية انحدار ذاتي لرتبة غير محدودة وهذا هو تعريف خاصية قابلية الانعكاس .

6. التنبؤ Forecasts:

1.6. التنبؤ باستعمال العملية : [1] IMA(0,1,1)

يمكن استعمال العملية IMA(0,1,1) للتنبؤ بالقيم المستقبلية للظاهرة تحت الدراسة وذلك من خلال التحليل الاتي أذ أن اسلوب معادلة الفروق عند الزمن المستقبلي $t+l$ للأنموذج يكتب بالصيغة[Box,G.E.P& Jenkins;1976,p.144]:

$$y_{t+l}=y_{t+l-1}+\varepsilon_{t+l}-\theta\varepsilon_{t+l-1} \qquad (13)$$

وبأخذ التوقع الشرطي للعملية اعلاه ينتج :

$$\hat{y}_t(1)=y_t-\theta\varepsilon_t \qquad (14)$$

لكل الخطوات المستقبلية للتنبؤ للزمن t

$$\hat{y}_t(l)=\hat{y}_t(l-1) \quad l>2 \qquad (15)$$

وبالتالي فان الانموذج التنبؤي سيكون:

$$\hat{y}_t(l)=\hat{y}_{t-1}(l)+\lambda\varepsilon_t \qquad (16)$$

حيث ان $\lambda=1-\theta$

والصيغة (16) تعتمد في حسابها على القيمة التنبؤية السابقة بالإضافة الى الكمية المصححة المعرفة بالكمية $(\lambda\varepsilon_t)$ حيث أن (λ) تقيس الصدمة العشوائية ، وهنالك صيغة أخرى لكتابة الصيغة التنبؤية وهي كالأتي [Box&Jenkins:1976]:

$$\hat{y}_t(l)=\lambda y_t+(1-\lambda)\hat{y}_{t-1}(l)$$

والتي تعتمد على تركيبة خطية للتنبؤ القديم والمشاهدة الجديدة وهي ايضاً تعتمد على المعلمة (λ)

2.6. التنبؤ باستعمال الاوساط المتحركة الموزونة اسياً :

يـــــّعد التنبؤ باستعمال طريقة الاوساط المتحركة الموزونة اسياً واحدة من أهم الطرائق الشائعة الاستعمال في تحليل السلاسل الزمنية الخطية والتي يرمز لها (Exponential Weighted Moving Average((EWMA)) وبعض الاحيان تسمى هذه الطريقة بالتمهيد الآسي وتكون الأوزان في هذه الطريقة متغيرة حسب أهمية المشاهدات المستعملة في معادلة التمهيد ويمكن اشتقاق معادلة التنبؤ عند التعويض عن $(\lambda=1-\theta)$ في الصيغة (3) لنحصل على:

$$\hat{y}_t(1)=(1-\theta)y_t+\theta\hat{y}_{t-1}(1) \qquad (17)$$

ومن العلاقة (5) كذلك يمكن الحصول على الصيغة الآتية :

$$\hat{y}_t(1)=(1-\theta)[y_t+\theta y_{t-1}+\theta^2 y_{t-2}+\theta^3 y_{t-3}-\cdots\ldots] \qquad (18)$$

والصيغة أعلاه تمثل أنموذج انحدار ذاتي غير محدودة ويمكن تفسير ذلك اذا كانت $(0<\theta<1)$ فأن الأوزان تضمحل اسياً ، وبالتالي ستكون المعادلة المستعملة في التنبؤ باستعمال حد الاضطراب (random shock) بالشكل التالي :

$$\hat{y}_t(1)=[\pi_1 y_t+\pi_2 y_{t-1}+\pi_3 y_{t-2}+\pi_4 y_{t-3}\ldots\ldots]=\sum_{j=1}^{\infty}\pi_j y_{t+1-j} \qquad (19)$$

بحيث أن : $\pi_j=(1-\theta)\theta^{j-1} \quad , \quad j=1,2,\ldots.$

وعندما تكون $(\theta=1-\lambda)$ في معادلة رقم (17) ينتج الآتي :

$$\hat{y}_t(1)=\lambda y_t+(1-\lambda)\hat{y}_{t-1}(1) \qquad (20)$$

$$\hat{y}_t(1)=\lambda y_t+\hat{y}_{t-1}(1)-\lambda\hat{y}_{t-1}(1)=\hat{y}_{t-1}(1)+\lambda(y_t-\hat{y}_{t-1}(1))$$

$$\hat{y}_t(1)=\hat{y}_{t-1}(1)+\lambda\varepsilon_t \qquad (21)$$

حيث أن : λ تمثل ثابت التمهيد والخطأ يساوي $\varepsilon_t=(y_t-\hat{y}_{t-1}(1))$

وهنا يمكن أن تكون الأوزان متساوية على الجانبين بشكل تقاربي [Box&Jenkins:1976, p.106]:

$$\sum_{j=1}^{t-1}\pi_j=\lambda\sum_{j=1}^{t-1}(1-\lambda)^j$$

وبالتالي نلاحظ ان المعادلة (16) في التنبؤ باستعمال العملية IMA(0,1,1) هي ذاتها المعادلة (21) لفترة واحدة في الأوساط المتحركة الموزونة أسياً مع معلمة التمهيد ذاتها[Box&Jenkins:1976, p.145].

وبالتالي خطأ التنبؤ سيكون بالصيغة الأتية :

$$\hat{\varepsilon}_t=y_t-\hat{y}_{t-1}(1) \quad ,t=2,3,\ldots,T$$

وأن مجموع مربعات خطأ التنبؤ بخطوة واحدة الى الأمام سيكون بالصيغة :

$$SSE = \sum_{t=1}^{T} \varepsilon_t^2 \quad , t = 2,3, \dots , T$$

حيث أن القيم الابتدائية للتنبؤ يمكن فرضها المشاهدة الأولى للبيانات .($\hat{y}_2 = y_1$)

3. الجانب العملي :

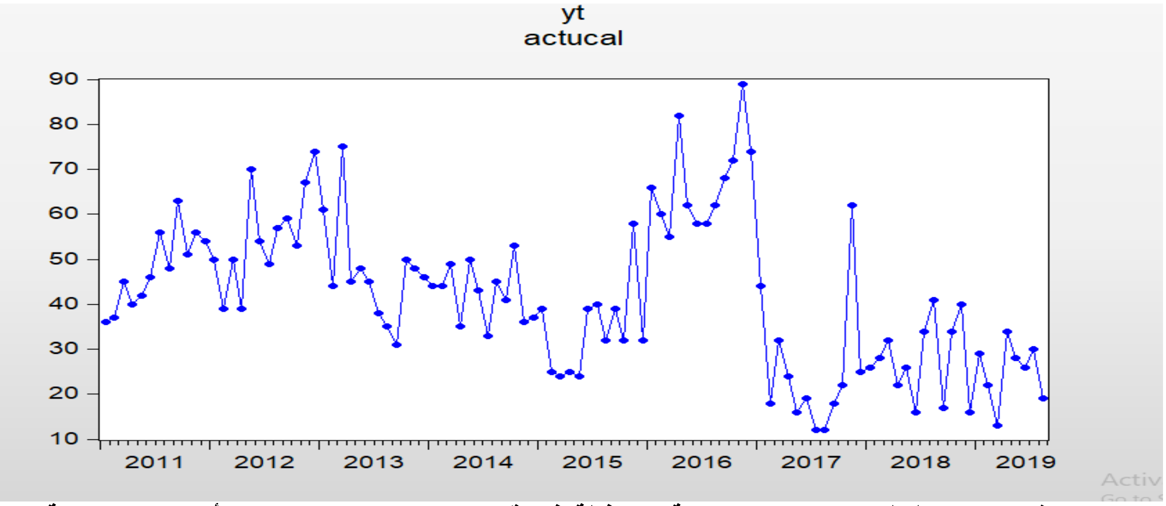
3.1 التحليل الوصفي وتشخيص الأنموذج الأفضل :

لأجل تطبيق أنموذج السير العشوائي على بيانات حقيقية فقد تم أخذ بيانات الحوادث المرورية الشهرية لحدود محافظة ذي قار على الطرق الخارجية التي تربط المحافظة بالمحافظات المجاورة لها والتي كانت مسجلة فقط في دائرة المرور العامة للمحافظة للفترة من شهر كانون الأول 2011 ولغاية آب 2019 وقد استعمال برنامج Eviewa 10في تحليل البيانات وفي إدنا جدول لعرض الإحصاءات الوصفية للظاهرة المعنية والتي سيرمز لها بالرمز y_t بالجدول الاتي :

جدول(1) يمثل الإحصاءات الوصفية للعملية y_t التي تمثل الحوادث المرورية في محافظة ذي قار

عدد المشاهدات	المتوسط	الوسيط	المنوال	الانحراف المعياري	التباين	أقل قيمة	أكبر قيمة
104	41.95	40.50	39	16.79	281.9041	12	89

نلاحظ من الجدول (1) أن عدد المشاهدات كانت (104) مشاهدة شهرية حيث نلاحظ أن معدل عدد الحوادث في الشهر كان (41.95) حادث وهذا المعدل يبين أهمية هذه الظاهرة ومدى خطورتها ، ايضا يمكن ملاحظة أن أعلى عدد حوادث كان (89) حادث في شهر نيسان من عام 2016 وأقل عدد حوادث كان (12) حادث لشهري تموز وآب لعام 2017 والشكل (1) يوضح الرسم البياني لحوادث الطرق في محافظة ذي قار للفترة المدروسة .



شكل(1) يمثل شكل تخطيطي للحوادث المرورية لمحافظة ذي قار للفترة من شهر كانون الأول2011 ولغاية شهر آب2019

ولأجل تشخيص وتحديد رتبة الأنموذج للسلسلة المعروضة في الشكل (1) نحسب قيمتي معامل الارتباط الذاتي (ACF) و معامل الارتباط الذاتي الجزئي (PACF) ومقارنته مع آلية التشخيص المقدمة من قبل [175.Box&Jenkins:1976, P] والجدول (2) يوضح أقيام الدالتين مع درجة التأخر المقابلة .

جدول (2) يمثل أقيام دالتي الارتباط الذاتي(AC) والارتباط الذاتي الجزئي(PAC) وكذلك الشكل التوضيحي للدالة (Autocorrelation) و (Partial Correlation) فضلاً عن قيم إحصاءة (Q-statistic)

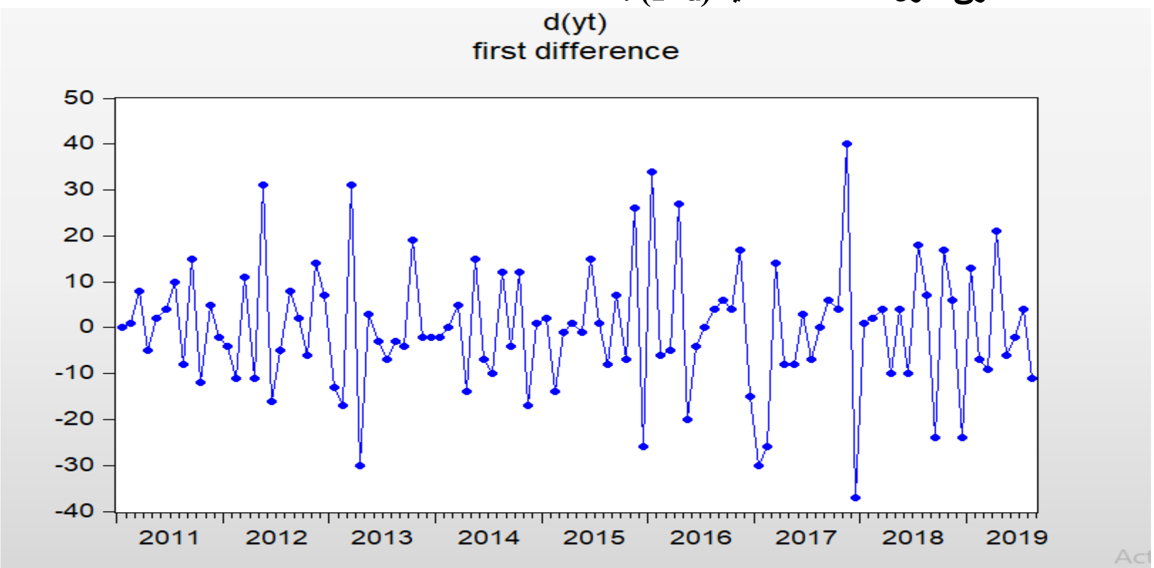
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.559	0.559	23.466	0.000
		2	0.553	0.349	46.716	0.000
		3	0.464	0.111	63.317	0.000
		4	0.322	-0.104	71.433	0.000
		5	0.281	-0.013	77.694	0.000
		6	0.213	0.003	81.369	0.000
		7	0.246	0.141	86.347	0.000
		8	0.108	-0.139	87.321	0.000
		9	0.002	-0.228	87.321	0.000
		10	0.073	0.128	87.784	0.000
		11	-0.116	-0.138	88.952	0.000
		12	-0.093	-0.028	89.714	0.000
		13	-0.168	-0.118	92.276	0.000
		14	-0.178	-0.013	95.187	0.000
		15	-0.147	0.106	97.218	0.000
		16	-0.217	-0.053	101.71	0.000

من خلال ملاحظة الرسم البياني للمشاهدات في الشكل (1) نلاحظ أن السلسلة غير مستقرة في الاتجاه ,وعند اجراء أختبار ديكي فولر الموسع (Augmented Dickey-fuller test) يتبين لنا أن السلسلة غير مستقرة ايضاً وحسب الاختبار في الجدول أدناه :

جدول(2) يمثل نتائج اختبار ديكي- فولر الموسع لبيان استقراريه السلسلة الزمنية

	t-Statistic	Prob.
Augmented Dickey-Fuller test statistic	-2.804027	0.0612
Test critical values : 1% level	-3.495677	
5% level	-2.890037	
10% level	-2.582041	

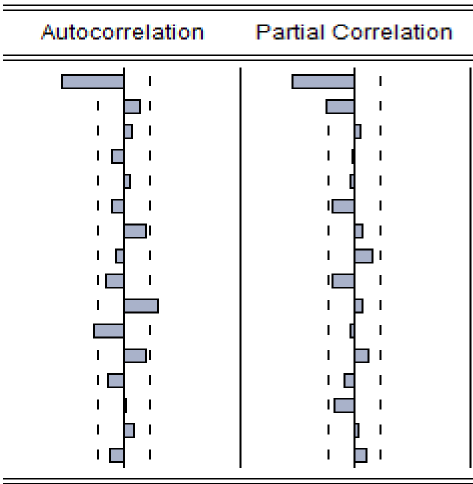
من خلال الاختبار في الجدول (2) نلاحظ أن القيمة الاحتمالية للاختبار كانت (0.0612) وهي أكبر من مستوى المعنوية (0.05) وبالتالي فأن السلسلة غير مستقرة أي أنها تحتوي على جذر الوحدة ولأجل التخلص من هذه المشكلة نأخذ الفرق الأول للسلسلة الأصلية (d=1) :



شكل (2) يمثل رسم سلسلة الفرق الأول للعملية $D(y_t)$))

من خلال الشكل (2) نلاحظ أن السلسلة أصبحت مستقرة في الفرق الأول وهذا ما أكدهُ أختبار جذر الوحدة لديكي-فولر الموسع ، ولأجل تشخيص الأنموذج وتحديد الرتبة بشكل صحيح نحسب أقيام وشكل الرسم لدالتي الارتباط الذاتي (ACF) والارتباط الذاتي الجزئي (PACF).

جدول (3) يمثل أقيام دالتي الارتباط الذاتي(AC) والارتباط الذاتي الجزئي(PAC) لسلسلة الفرق الأول ($D(y_t)$) وكذلك الشكل التوضيحي للدالة (Autocorrelation) و (Partial Correlation) فضلاً عن قيم إحصاءة (Q-statistic)



	AC	PAC	Q-Stat	Prob
1	-0.562	-0.562	23.353	0.000
2	0.145	-0.249	24.932	0.000
3	0.066	0.043	25.263	0.000
4	-0.108	-0.010	26.164	0.000
5	0.044	-0.042	26.317	0.000
6	-0.111	-0.193	27.297	0.000
7	0.190	0.078	30.207	0.000
8	-0.066	0.168	30.568	0.000
9	-0.172	-0.195	33.033	0.000
10	0.311	0.071	41.257	0.000
11	-0.274	-0.030	47.732	0.000
12	0.190	0.120	50.911	0.000
13	-0.151	-0.087	52.942	0.000
14	0.021	-0.181	52.981	0.000
15	0.094	0.029	53.801	0.000
16	-0.123	0.109	55.237	0.000

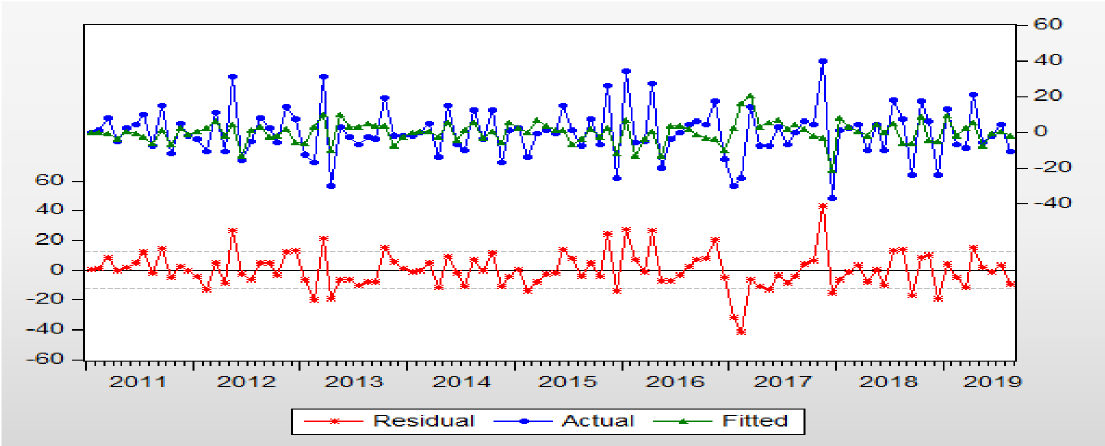
3.2 مطابقة أنموذج ARIMA:

لأجل التحقق من مطابقة الانموذج للبيانات الماخوذه للظاهرة المدروسة بعد عملية المطابقة سوف نعتمد على مجموعة من المعايير في تحديد الانموذج الافضل وحسب ملائمة هذه المعايير لعملها في الجانب النظري ولعل اهمها معيار المعلومات اكاكي (AIC) وعنما يكون حجم العينة اكبر من 60 مشاهدة سيكون البديل لهذا المعيار هو معيار المعلومات البيزي (BIC) حيث سيتم اختيار الأنموذج المقابل الذي يقابل أقل قيمة للمعيار (BIC) والجدول التالي يعرض النماذج المقترحة وما يقابلها من معايير محسوبة وهذا ما يسمى بالنمذجة في تحليل السلاسل الزمنية [75.Harvey:1990,p] .

جدول (4) يمثل النماذج المقترحة وقيم المعايير التي تبين افضل أنموذج

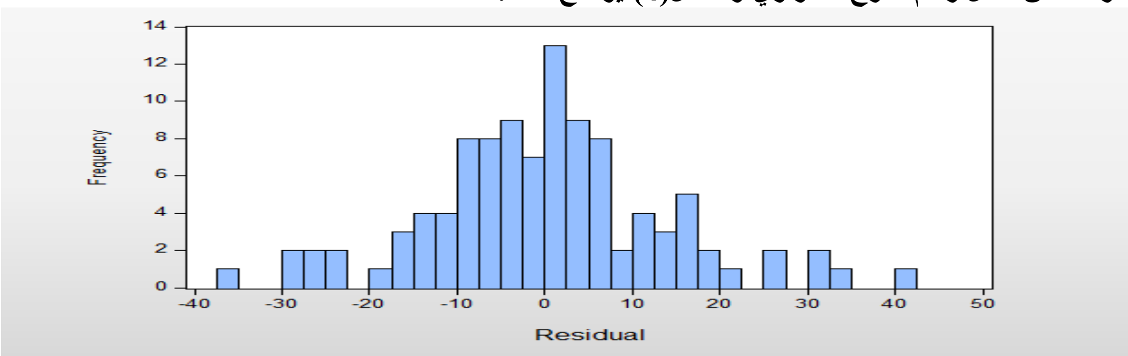
Models	AIC	BIC
ARIMA(0,1,1)	7.624757	7.720303
ARIMA(0,1)	7.9760	8.07094
ARIMA(1,1,0)	7.624737	7.720343
ARIMA(1,1,1)	7.605	7.733323
ARIMA(0,1,2)	7.603101	7.730576
ARIMA(1,1,2)	7.626858	7.786202
ARIMA(2,1,1)	7.616651	7.775995
ARIMA(2,1,2)	7.640682	7.831894
ARIMA(1,1)	7.646600	7.773082
ARIMA(2,2)	7.629318	7.819040

من خلال الجدول (4) نلاحظ ان افضل انموذجين يمتلكان أفضل مطابقة لبيانات ظاهرة الحوادث المرورية للفترة المدروسة هو أنموذج (BIC=IMA(1,1)) من خلال ملاحظة المؤشرات التي يمتلكها من حيث أقل قيمة لمعيار المعلومات البيزي (BIC) والتي كانت (7.720303) للأنموذج المقترح الأول ، وعند الاعتماد على معيار المعلومات أكايكي فأن الأنموذج المقترح الذي يمتلك أقل قيمة هو (AIC=IMA(1,2)) ، ولكن من حيث المعلومات المتوفرة لدينا في مرحلة التشخيص للأنموذج فان الأنموذج الأول والذي يرمز لهُ بعض الأحيان بالرمزIMA(1,1) وهو الأفضل في تمثل الظاهرة والشكل(3) يوضح جودة عملية المطابقة و البواقي:.



شكل (3) يمثل المطابقة في عملية التقدير لعملية الأوساط المتحركة التكاملية IMA(0,1,1) مقارنة مع البيانات الحقيقية.

وايضاً يمكن التحقق من اختبار التوزيع الطبيعي للبواقي وبشكل تقاربي والمسمى (Whiteness)انظر[111.Fan&Yao:2003,P] أو من خلال رسم التوزيع الاحتمالي للبواقي والذي يمكن عرضهُ من خلال رسم مدرج التكراري والشكل(4) يوضح ذلك :



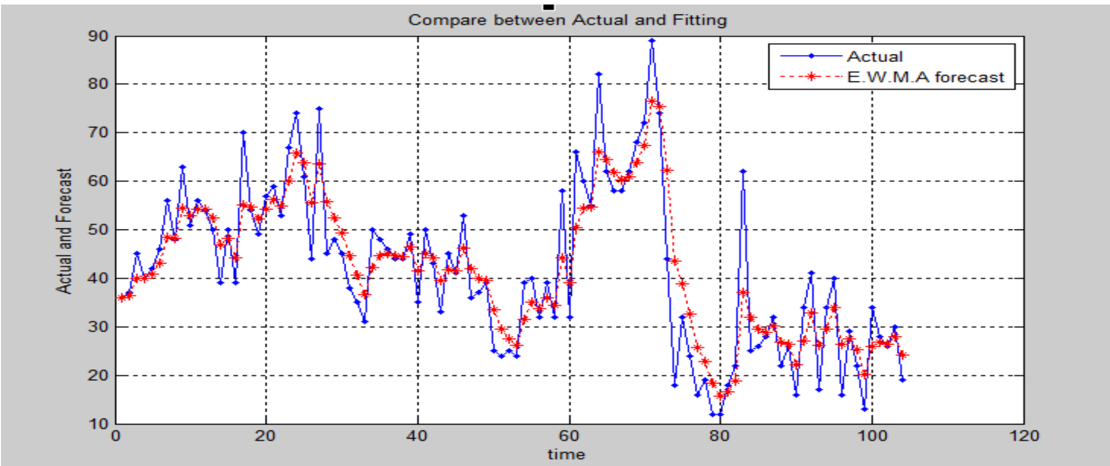
شكل (4) يمثل التوزيع الأحتمالي للبواقي من خلال المدرج التكراري للبواقي.

أما فيما يخص طريقة الأوساط المتحركة الموزونة آسياً (EWMA) التي هي أحدى طرق التمهيد الآسي البسيط فان التقدير سيكون كما في الجدول (5) الذي يوضح أفضل معلمة تمهيد مختارة لهذه الطريقة باستعمال البيانات الحقيقية لظاهرة الحوادث المرورية

جدول (5) يمثل التقدير باستعمال طريقة الأوساط المتحركة الموزونة آسياً ((EWMA

Parameters:	Alpha	0.4200
	Beta	0.0000
Sum of Squared Residuals		7828.651
Root Mean Squared Error		10.42743
End of Period Levels:	Mean	75.70730
	Trend	0.222222

من خلال الجدول (5) نلاحظ أن أفضل قيمة لمعلمة التمهيد كانت $(Alpha = \lambda = 0.42)$ والتي أعطت أقل جذر لمجموع مربعات للخطأ وكان (10.42743=RMAE) وأفضل مجموع مربعات للبواقي والذي مقداره (7828.651=SSR) .وعند استعمال أفضل قيمة للمعلمة في حساب معادلة التمهيد والتي يمكن تمثيلها بيانياً بالشكل (5) :



شكل (5) يمثل التقدير (المطابقة) باستعمال طريقة الأوساط المتحركة الموزونة اسياً مقارنة مع البيانات لحقيقية

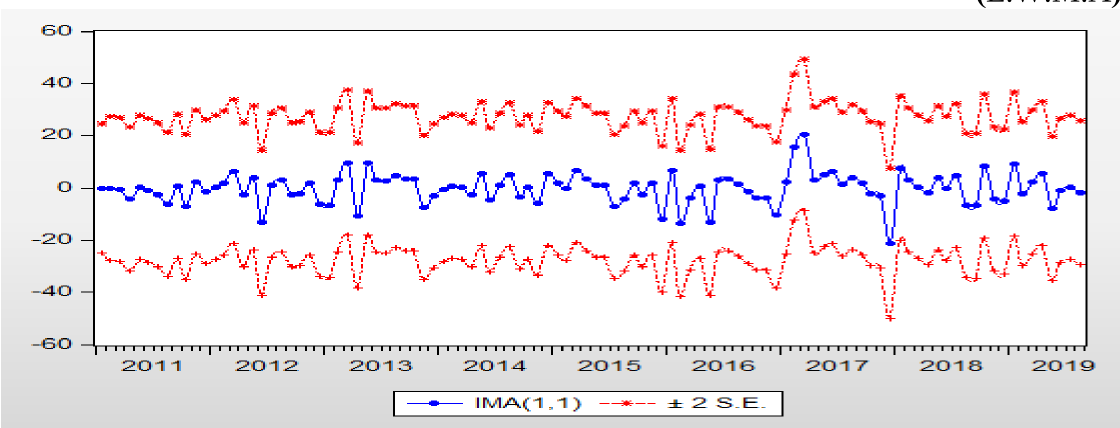
3.3 التنبؤ :

تعد مرحلة التنبؤ من المراحل الأكثر فائدة في عملية تحليل السلسلة الزمنية حيث تأتي بعد مرحلة السيطرة على الأنموذج وهي تقدم فكرة عن سلوك الظاهرة في المستقبل لأجل تهيئة كافة الاحتياطات اللازمة واتخاذ القرارات المناسبة لتفاد أو تقليل من الخسائر المتوقعة الناتجة عن الظاهرة المدروسة ، وهنا فان التنبؤ بظاهرة الحوادث المرورية في الفترة المستقبلية يكون مفيد جداً لنا لتقديمة فكرة واضحة عن الخسائر التي تنتج عنها مما يتطلب تهيئة كافة الاحتياطات المناسبة واتخاذ الإجراءات والقرارات الكفيلة بتقليل تلك الخسائر الناتجة عن تلك الحوادث سواء كانت خسائر في الموارد البشرية أو الموارد المادية وفي هذا البحث تم التنبؤ باستعمال الطريقتين المقدمتين في الجانب النظري والمقارنة بينهما وهما الأوساط المتحركة الموزونة آسياً وأنموذج الأوساط المتحركة التكاملية IMA(0,1,1) . والجدول التالي يعرض معايير أفضل تنبؤ بالاعتماد على خطأ التنبؤ وهما المعدل المطلق لخطأ التنبؤ (Mean Absolute Forecast Error(MAE)) ومتوسط مربعات خطأ التنبؤ (Mean Square Forecast Error(MSE)) :

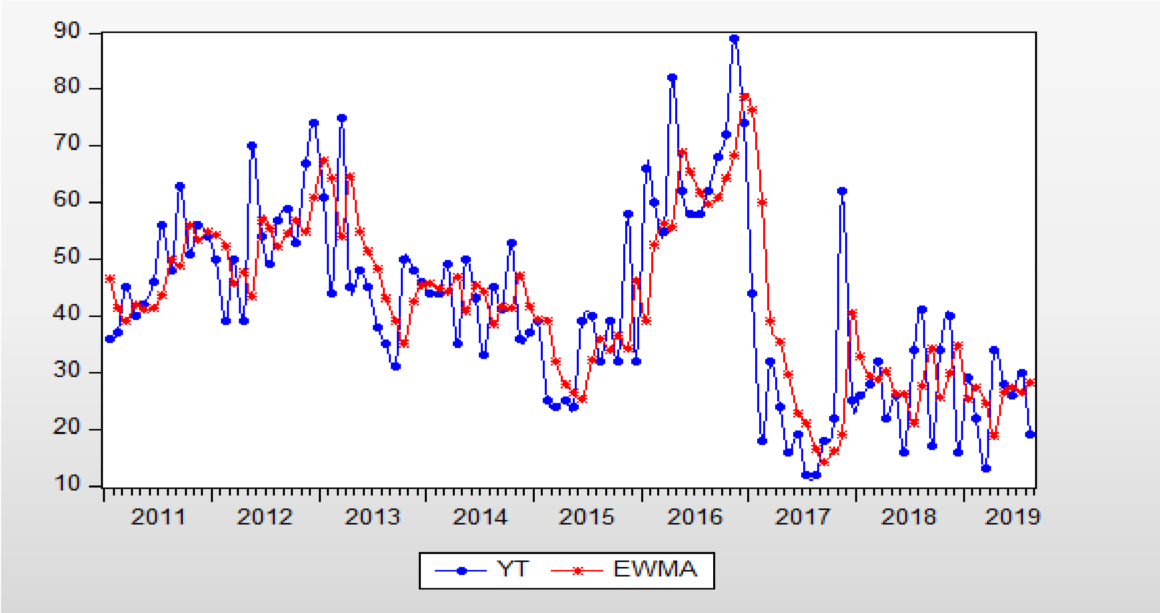
جدول (4) يمثل معايير دقة التنبؤ باستعمال طريقتي الأوساط المتحركة الموزونة آسياً وأنموذج IMA(1,1).

MAE	MSE	Forecast Accuracy Forecast Methods
9.0372	148.9558	E.W.M.A. forecast method
9.0022	147.2733	IMA(1,1) forecast method

والشكل(6) يبين دقة التنبؤ باستعمال أنموذج IMA(1,1) والشكل(7) يبين دقة التنبؤ باستعمال طريقة (E.W.M.A)



شكل (6) يمثل سلسلة التنبؤات باستعمال طريقة عملية الأوساط المتحركة التكاملية . IMA(0,1,1)



شكل (7) يمثل سلسلة التنبوءات باستعمال طريقة الأوساط المتحركة الموزونة آسياً (EWMA) مع البيانات الحقيقية.

4.الاستنتاجات :

1-أن قيمة معلمة التمهيد مهمة جدا في تحديد دقة التنبوءات باستعمال طريقة الأوزان المتحركة الموزونة اسياً حيث أن الاوزان تتناقص اسياً في صيغة متعدد الحدود التنبوئية .

2-أن اسلوب النمذجة هو الاختبار الافضل في تحديد الأنموذج المناسب للبيانات المتوفرة (الحوادث المرورية في محافظة ذي قار) وبالتالي فأن التنبوءات التي تنتج عن هذا الأنموذج تكون أكثر دقة من غيرها .

3-يمكن السيطرة على ظاهرة الحوادث المرورية من خلال أنموذج محدد وحسب البيانات المتوفرة وبالتالي التنبؤ بعدد الحوادث لأزمنة مستقبلية وهذا يتيح لأصحاب القرار أتخاذ الاجراءات المناسب للحد من الحوادث

5.التوصيات :

1-أهمية التحليل الرياضي لأنموذج الأوساط المتحركة التكاملية لأنه يمثل العديد من الظواهر في الواقع .

2-في عملية النمذجة الاعتماد على معيار المعلومات البيزي في مرحلة التحقق من الأنموذج فضلاً عن حساب بقية المؤشرات الاحصائية لمطابقة الأنموذج الصحيح .

3-يعد من الجيد استعمال النمذجة باستعمال عملية أنموذج الانحدار الذاتي والأوساط المتحركة التكاملية في عملية التنبؤ بالقيم المستقبلية .

4-يمكن تطبيق الأسلوب ذاته لبقية المحافظات للسيطرة على هذه الظاهرة .

المصادر :

[1]Box,G.E.P. and Jenkins,G.M.,(1976)" Time Series Analysis Forecasting and Control," ,California, Holden-Day,Oakland.

[2]Cox,D.R.,(1961),"Prediction by Exponentially Weighted Moving Averages and Related Methods," Royal Statistical Societly, No.2,PP.414-422.

[3]Fan,J. and Yao,Q.,(2003),"Nonlinear Time Series Nonparametric and Parametric Methods," New York, Springer-Verlag New York,Inc.

[4]Gabrielsen,A.and et.al,(2015),"Forecasting Value-at-Risk with Time-VaRying Variance,Skewness and Kurtosis in An Exponential Weighted Moving Average Framework," Annals of Financial Economics,Vol.10,No.1,PP.1550005-29.

[5]Geller,E. and Wilson,R.D.,(1978)"Exponentially Wieghted Moving Averages Forecasting Program with Alpha-Level Optimizer,"Marketing Research,Vol.15,No.2,PP.271-272.

[6]Granger,C.W.J. and Newbold,P.(1986)"Forecasting Economic Time Series," 2nd Edition London,Academic Press Inc.
[7]Harvey,A.C.,(1990),"Forecasting Structural Time Series Models and the Kalman Filter," New York, Cambridge University Press.
[8]Montgomery,D.C. , Jennings,C.L. and Kulahci,M.,(2008),"Introduction to Time Series Analysis and Forecasting," New Jersey, John Wiley&Sons,Inc.
[9]Ruey S.T.,(2005),"analysis of Financial Time Series,"2nd Edition,New Jersey, John wiley&Sons,Publication,Hoboken.
[10]Terence C.M.(2019)," Applied Time Series Analysis," ,Loughborough,United Kingdom, Academic Press,An Imprint of Elsevier.
[11]Tsay,R.C.,(2005),"Analysis of Financial Time Series ,"2nd Edition,New Jersy, John Wiley&Sons,Inc.
[12]Tseng,S. and Adams,B.M.,(1994),"Monitoring Autocorrelated Processes with An Exponentially Weighted Moving Avarage Forecasting,"J.Statist.Comput.Simul.,Vol.50,PP.187-195.
[13]Tseng,F.M. and Tzeng,G.H.,(2002)"A Fuzzy Seasonal ARIMA Model for Forecasting," Fuzzy Sets and Systems,126 367-376.
[14]Winters,P.R.,(1960)"Forecasting Sales by Exponentially Weighted Moving Average,"Management Science,Vol.6,No.3,pp.324-342.

Compare Prediction by Autoregressive Integrated Moving Average Model from first order with Exponential Weighted Moving Average

Ali Salman Habeeb
University of Sumer
a.habib@uos.edu.iq

Received: 23/11/2019 **Accepted :7/1/2020** **Published :June / 2020**

Abstract:

The prediction process of time series for some time related phenomena in particular the autoregressive integrated moving average(ARIMA) models is one of the important topics in the theory of time series analysis in applied statistic. Perhaps its importance lies in the basic stages in analyzing of the structure or modeling and the conditions that must be provided in the stochastic process. This paper deals with two methods of predicting the first was a special case of autoregressive integrated moving average which is ARIMA (0,1,1) if the value of the parameter equal to zero, then it is called Random Walk model, the second was the exponential weighted moving average (EWMA). It was implemented in the data of the monthly traffic accidents in the province of Dhi Qar Governorate for the period from (Jan. 2011) to (Aug. 2019). It was found through the research that the model studied is well of the traffic accident, we can predict dangerous traffic accident using this model and reduce the aggravation through Develop plans strategic of the roads.

Keyword: Random walk, Moving Average Integrated Process, Exponential Weighted Moving Average, Prediction.