



مقارنة لأساليب التقدير التكرارية للبيانات المرتبطة ذاتيا

م. د. جنان عباس ناصر
معهد الادارة / الرصافة

المستخلص

في هذا البحث نستعرض اساليب الـ $ARIMA(p,d,q)$ ، الـ $EWMA$ والـ DLM (النمذجة الديناميكية الخطية) بأيجاز لتكوين بنية الارتباط الذاتي للبيانات. نعتد خوارزميات التقدير التكرارية والتنبوء المعتمدة على أسلوب بيز وتنقية كالمان لمشاهدات مرتبطة. نتحرى عن تاثير تلك المعالجات لمتوسط مربعات الخطأ (MSE) ونقارن بينهم باستخدام بيانات مولدة.

Abstract:

In this study, we review the $ARIMA(p, d, q)$, the $EWMA$ and the DLM (dynamic linear moodelling) procedures in brief in order to accomdate the ac(autocorrelation) structure of data .We consider the recursive estimation and prediction algorithms based on Bayes and KF (Kalman filtering) techniques for correlated observations.We investigate the effect on the MSE of these procedures and compare them using generated data.

1- المقدمة

من المفيد استخدام النظرية والتطبيقات الخاصة بلوحات السيطرة لتحسين النوعية والانتاجية في العديد من المجالات الصناعية. أذ طورت تلك النظرية على افتراض أساسي وهو كون المشاهدات المستخدمة للوحات السيطرة تكون مستقلة [5]. أذ يمكن افتراض نموذج المشاهدة في هذه الحالة بالصيغة الآتية :

$$X_t = \mu + \varepsilon_t \quad \dots (1.1)$$

عندما تمثل X_t المشاهدة رقم t ($t=1,2,\dots,n$)، ويمثل μ متوسط العملية وتكون الأخطاء

(ε_t) متغيرات عشوائية مستقلة ومتطابقة التوزيع (iid) بمتوسط صفرو تباين ثابت مقداره σ^2 . وفي التطبيقات الصناعية، توجد ثلاثة أنواع من لوحات السيطرة، بالتحديد لوحة شيوارت التي استخدمت من قبل الباحث Shewart [7] عام 1931 عندما يكون المتغير تحت السيطرة يتبع التوزيع الطبيعي للكشف عن الانحرافات الكبيرة في متوسط العملية الانتاجية. ولوحة cusum التي اقترحت من قبل الباحث Page [7] عام 1961 وطورت من قبل الباحث Lucas عام 1973 باستخدام المجموع المتراكم الذي يظهر التغيرات الفجائية في متوسط العملية الانتاجية على شكل اتجاه خطي للقيم المرسومة في لوحة cusum. ولوحة Exponentially Weighted Moving Average (EWMA) التي اقترحت من قبل الباحث Roberts [10,7] عام 1959 وذلك برسم المتوسط المتحرك الموزون اسيا وكافة المشاهدات السابقة والحالية للعملية الانتاجية. أذ تمثل لوحة الـ EWMA دالة التنبؤ بفترة واحدة للامام (the one -step - ahead forecast) المتقدم ذكرها في الصيغة (1.1) تعطى بتحقيق خطة المتوسط المتحرك الموزون اسيا وكما يلي

$$Z_t = \hat{X}_{t,1} = \sum_{j=0}^n \lambda(1-\lambda)^j X_{t-j} \quad \dots (1.2)$$

عندما تكون $0 < \lambda < 1$ أذ تكون المعادلة في الصيغة (1.2) مكافئة للصيغة التكرارية التالية :

$$Z_t = \lambda X_t + (1-\lambda)Z_{t-1} \quad \dots (1.3)$$

أذ يتم اختيار قيمة بداية ملائمة لـ Z_0 . وتكون لوحة Exponentially Weighted Moving Average (EWMA) اسلوب شائع الاستخدام في عملية السيطرة الاحصائية ، حين تكون المشاهدات غير مرتبطة كما مبينة في الصيغة (1.1) وان حدي السيطرة الادني (LCL) و الاعلى (UCL) لتلك اللوحة (تحت شروط حالة الاستقرار) تكون وفق الصيغة الآتية [7,5]:

$$LCL = \bar{X} - K\sigma\sqrt{\lambda/(n(2-\lambda))}$$

$$UCL = \bar{X} + K\sigma\sqrt{\lambda/(n(2-\lambda))}$$

ويمثل \bar{X} متوسط العينة، ونموذجيا تكون قيمة $K=2$ لحدود التحذير و $K=3$ لحالات الخروج عن السيطرة . الا انه بالتطبيق غالبا ما يخالف شرط الاستقلالية للملاحظات. حيث يمكن الكشف عن الارتباط الذاتي من خلال رسوم تشخيصية او من خلال صيغ اختبارات للاستقرارية في السلسلة الزمنية، أذ تستخرج قيم معاملات الارتباط الذاتي الخاصة بالسلسلة فإذا كانت السلسلة مستقرة فإن قيم معاملات الارتباط الذاتي تقترب من الصفر بعد الوقت الثاني او الثالث. بكلام اخر، ان السلسلة الزمنية تبقى ثابتة بمرور الزمن عندما يكون التوزيع المشترك للملاحظات $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ نفس التوزيع المشترك

للملاحظات $X_{t_1+k}, X_{t_2+k}, \dots, X_{t_n+k}$ ولاية قيمة k أي ان

$$F(X_{t_1}, X_{t_2}, \dots, X_{t_n}) = F(X_{t_1+k}, X_{t_2+k}, \dots, X_{t_n+k})$$

وان شروط الاستقرارية هي :

• قيمة الوسط الحسابي ثابتة أي

$$E(X_t) = \mu$$

• قيمة التباين ثابتة أي

$$E(X_t - \mu)^2 = \sigma^2$$

• دالة التغاير (التباين الذاتي المشترك) Autocovariance هي دالة في فرق الزمن أي

$$E((X_t - \mu)(X_{t+k} - \mu)) = \gamma_x(k)$$

اما السلاسل الزمنية غير المستقرة فهي السلاسل التي تتغير خصائصها بتغير الزمن اذ تقترب فيها قيم معاملات الارتباط الذاتي من الصفر بعدد كبير من الفترات الزمنية . وتسمى السلسلة الزمنية بانها مستقرة من الدرجة الثانية، اذا حققت الشرطين الاول والثاني من الشروط اعلاه وباستخدام عامل الفروقات الارتدادية (Backward Difference Operator) ويرمز له بـ ∇ ويكون $\nabla X_t = X_t - X_{t-1}$ وتصبح السلسلة الزمنية مستقرة بعد أخذ d من

$$Y_t = \nabla^d X_t$$

ان رسم بسيط للبواقي (Residuals) ممكن ان يساعد في الكشف عن الارتباط الذاتي فمثلا اذا البواقي المرسومة مقابل الوقت تتغير بسرعة في الإشارة، ربما يبين ذلك وجود ارتباط ذاتي سالب. من ناحية ثانية، وجود عدد كبير غير اعتيادي للبواقي متجمعة معا على نفس الجانب للمتوسط ربما يبين ذلك وجود اخطاء محكومة بالارتباط الذاتي الموجب. ان مخالفة افتراض الاستقلالية للملاحظات نتيجة لوجود ارتباط ذاتي بينهما له تأثير جاد على لوحات السيطرة. وان تجاهل تأثير وجود اخطاء مرتبطة بشكل موجب (اوبشكل سالب) له تاثير على تقدير قيمة التباين الحقيقي σ^2 .

عموما فان ذلك يؤدي الى تزايد التكرار لانذارات زانفة. بعبارة اخرى ، يكون متوسط طول التشغيل (Average Run Length (ARL)) اقصر بكثير منه للحالة التي تكون فيها مشاهدات

العملية غير مرتبطة. لهذا وعلى الاغلب، لا يمكن ان تحدد الحالة الحقيقية للسيطرة على العملية من لوحات السيطرة الاعتيادية .

بناء على ما تقدم، فقد اعتمدت عدة نماذج للبيانات المرتبطة ذاتيا مثل طريقة Exponentially Weighted Moving Average (EWMA)، و النمذجة الديناميكية الخطية (DLM) وأسلوب بيزوتنقية كالمان لتحديد افضل نموذج اعتمادا على قيمة متوسط مربعات الخطاء (MSE) كمعيار مطلق للمفاضلة بين تلك الطرق .

2. هدف البحث

يتمثل بالمقارنة بين اسلوب Exponentially Weighted Moving Average (EWMA)، الـ DLM (النمذجة الديناميكية الخطية)، أسلوب بيزوتنقية كالمان لتحديد أفضل نموذج للبيانات المرتبطة من خلال دراسة تاثير كل اسلوب على قيمة متوسط مربعات الخطاء (MSE) من خلال السلاسل زمنية مرتبطة ذاتيا ومولدة وفقا لتوليفات مختلفة لتباين سلسلة السير العشوائي (Random Walk) ولتباين سلسلة المشاهدات.

3. النماذج لبيانات مرتبطة ذاتيا

في المراجع التاريخية لهذا الموضوع توجد طريقتين رئيسية متاحة لتكوين لوحات السيطرة لبيانات مرتبطة . إذ تعتمد الطريقة الاولى على تقدير نموذج انحدار ذاتي - اوساط متحركة متكامل (ARIMA) (Integrated autoregressive – moving average model) [3,9] للعملية الاصلية لتضمن بنية الارتباط الذاتي، ثم تكوين لوحات السيطرة للبواقي. اما الطريقة الثانية فهي لوحة السيطرة المعتمدة على احصاء الـ EWMA [7,5,10]، كما سيرد شرحها بايجاز.

3.1 نماذج ARIMA (p, d, q)

تكون صيغة النموذج العام نموذج انحدار ذاتي - اوساط متحركة متكامل (ARIMA) من الرتبة (p,d,q) كالآتي [3,9] :

$$\Phi(B)(1-B)^d X_t = \Theta(B)\varepsilon_t \quad \dots(3.1)$$

عندما تكون $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ متعدد حدود من الدرجة p لنموذج انحدار ذاتي (AR(P)) مستقر. وتكون $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ متعدد حدود من الدرجة q لنموذج اوساط متحركة (MA(q)) انعكاسي (invertible) . ويمثل B عامل الارتداد الخلفي ($B^j X_t = X_{t-j}$) والاختفاء (ε_t) متغيرات عشوائية مستقلة ومتطابقة التوزيع (iid)

بمتوسط صفرونتباين σ^2 . لنفرض \hat{X}_t تكون القيمة المتنبأ بها بالوقت t من تقدير نموذج ARIMA في الصيغة (3.1) ، واخطاء التنبوء بفترة واحدة للامام (او البواقي) معطاة بالصيغة الاتية [8] :

$$\varepsilon_t = X_t - \hat{X}_t \quad , \quad t=1,2,\dots$$

أن الاخطاء (ε_t) تكون متغيرات عشوائية مستقلة ومتطابقة التوزيع، إذا كان النموذج المقدر ملائم من حيث الرتبة الموضوعة له. لهذا السبب، لعملية السيطرة الاحصائية ببيانات مرتبطة، يمكن تطبيق لوحات السيطرة للبواقي المقدر من النموذج المقدر ARIMA(p,d,q) [1,7]، ويقال ان العملية X_t (ARIMA) في الصيغة (3.1) خارج السيطرة إذا يوجد انحراف معنوي في

مستوى المتوسط، ويمكن اكتشاف هذا الانحراف من تطبيق لوحة السيطرة للبوافي لانه يفترض تشخيص غير ملائم للنموذج من حيث الرتبة الموضوعه له ليكون منقول للبوافي .

3.2 لوحات الـ EWMA لبيانات مرتبطة ذاتيا

يمكن استخدام خطة الاوساط المتحركة الموزونه اسيا (Exponentially Weighted Moving Average (EWMA)) في الصيغة (1.3) لبيانات مرتبطة ذاتيا. وللتوضيح، افترض لبيانات سلسلة زمنية (متجانسة) غير مستقرة، يمكن نمذجتها بنموذج ARIMA(0,1,1) أو IMA(1,1) وفق الصيغة الاتية [8] :

$$U_t = \varepsilon_t - \theta \varepsilon_{t-1} \quad \dots (3.2)$$

عندما تكون $U_t = X_t - X_{t-1}$ ولنفرض $Z_t = \hat{X}_t(1)$ تمثل التنبؤ بفترة واحدة للامام (بمعني اصغر متوسط مربع) للوقت $t+1$ عملت في الوقت t . دالة التنبؤ المثلى المطابقة تعطى بالمتوسط الشرطي وكمايلي :

$$Z_t = E(X_{t+1} / X_t, X_{t-1}, \dots) = X_t - \theta \varepsilon_t \quad \dots (3.3)$$

أذ تكون الصيغة (3.3) مكافئة للصيغة (1.3) بالمعلمة $\lambda = (1-\theta)$ لانه $\varepsilon_t = X_t - Z_{t-1}$ بتعويض ε_t في الصيغة (3.3) نحصل على $Z_t = X_t - \theta(X_t - Z_{t-1})$ ، وبفتح القوس نحصل على $Z_t = X_t - \theta X_t + \theta Z_{t-1}$ ثم اخراج عامل مشترك X_t من

الحد الاول والثاني نحصل على $Z_t = (1-\theta) X_t + \theta Z_{t-1}$ وهي مكافئة للصيغة (1.3) بالمعلمة $\lambda = (1-\theta)$ وان θ ستكون مساوية لـ $\lambda = (1-\theta)$ في هذه الحالة ، فان اخطاء التنبؤ بفترة واحدة للامام (ε_t) تكون مستقلة بمتوسط صفر وانحراف معياري (σ) ، اذا كان النموذج IMA(1,1) المقدر ملائم من حيث الرتبة الموضوعه له. عموما فان الـ EWMA بـ λ ملائمة ستعطي تنبؤ جيد عند فترة واحدة للامام [7] حتى لو كانت المشاهدات من عملية مرتبطة بشكل موجب او بشكل سالب ومتوسط عملية لاينحرف بسرعة كبيرة .

3.3 النماذج الديناميكية الخطية من الرتبة الاولى (DLM)

أزداد الاهتمام بالنظرية والتطبيق للاساليب التكرارية والتنقية في عملية السيطرة الاحصائية [12,11]. اذا افترضنا النموذج الديناميكي الخطي من الرتبة الاولى (First Order Dynamic Linear Model)، حين يكون متوسط العملية سير عشوائي (Random Walk) يتغير ببطيء كمثل لطريقة تكرارية ثم تبين بان هذا النموذج يمكن ان يمثل خطة الـ EWMA بصورة جيدة وبمعلمات مختارة بصورة ملائمة. لنفرض سلسلة من المشاهدات مولدة بالصيغة الاتية [8] :

$$X_t = \mu_t + v_t \quad \dots (3.4)$$

عندما تكون $v_t \sim N(0, U_t^2)$ ، أي ان الحد v_t يتبع التوزيع الطبيعي بمتوسط صفرو تباين عند الفترة t مقداره U_t^2 ويمثل μ_t مستوى المتوسط عند الوقت t . ولنفرض نشوء الوقت لمستوى المتوسط يتبع السير العشوائي وفق الصيغة الاتية :

$$\mu_t = \mu_{t-1} + w_t \quad \dots (3.5)$$

عندما تكون $w_t \sim N(0, V_t^2)$ ، أي ان الحد w_t يتبع التوزيع الطبيعي بمتوسط صفرو تباين عند الفترة t مقداره V_t^2 . وبافتراض ان السلسلتين $\{v_t\}$ ، $\{w_t\}$ تكونان مستقلتين خلال الوقت، وتكون عملية مستقلة بشكل متعاقب لكل قيم t .

بوضع التباينات $V_t^2 = V^2$ و $U_t^2 = U^2$ كثوابت خلال الوقت يمكن الحصول على الحالة

الخاصة والمهمة للصيغتين (3.4) و (3.5). وان الانحراف النسبي للحالة في الصيغة (3.5) لمعادلة المشاهدة في الصيغة (3.4) في حالة النموذج الثابت الاخير تعطى بـ $r = V^2/U^2$ وتعرف كنسبة تذبذب الاشارة (Single to Noise Ratio). ان الصيغة (3.4) بالوقت (t-1) تكون $X_{t-1} = \mu_{t-1} + v_{t-1}$ وجمع الصيغة (3.4) بالوقت (t) ونفس الصيغة بالوقت (t-1) والتقدم ذكرها نحصل على $X_t - X_{t-1} = \mu_t - \mu_{t-1} + v_t - v_{t-1}$ وحيث ان $W_t = \mu_t - \mu_{t-1}$ من الصيغة (3.5)، وبذلك في حالة ثبات التباين في الصيغتين (3.4) و (3.5) نحصل على

$$X_t - X_{t-1} = w_t + v_t - v_{t-1} \quad \dots(3.6)$$

ويمكن اعادة كتابة الصيغة (3.6) كنموذج IMA(1,1) لمعلمات مختارة بصورة ملائمة لـ

θ و σ^2 وكالاتي

$$X_t - X_{t-1} = \varepsilon_t - \theta\varepsilon_{t-1} \quad \dots(3.7)$$

عندما تكون θ الحل للمعادلة $U^2\theta^2 - \theta(2U^2 + V^2) + U^2 = 0$ عندما

تكون $|\theta| < 1$ ، وتكون الاخطاء (ε_t) مختارة بصورة ملائمة ومستقلة ومتطابقة التوزيع من

التوزيع الطبيعي وتمثل عملية ضوضاء بتباين U^2/θ . وهكذا التنبوء من نموذج ديناميكي خطي بتباين ثابت يحقق الصيغتين (3.4) و (3.5) يمكن ان يكون مستحصل بخوارزمية الـ EWMA المعطاة في المبحث (3.2).

4. التقدير التكراري بطريقة بيز وتنقية كالمان

يتناول هذا المبحث طريقتين تكرارية وطرق التنبوء المعتمدة على خوارزمية بيز (Bayesian) وتنقية كالمان (Kalman Filtering). اولاً، نفترض طريقة بيزين باستخدام النموذج الديناميكي الخطي من الرتبة الاولى والمتقدم ذكرها في المبحث (3.3) وكما يلي:

4.1 خوارزمية بيزين

لنفرض F^t تمثل σ -algebra مولدة بـ X_t, X_{t-1}, \dots, X_1 ولنفرض F^0 تمثل المعلومة الابتدائية المتوفرة حول العملية. وبافتراض معرفة التوزيع الابتدائي $(\mu_0 / F^0) \sim N(m_0, \alpha_0^2)$ والتوزيع اللاحق $(\mu_{t-1} / F^{t-1}) \sim N(m_{t-1}, \alpha_{t-1}^2)$.

وباعتماد النظرية التالية [8]:

$$1. \text{ التوزيع الاولي } (\mu_t / F^t) \sim N(m_{t-1}, \alpha_{t-1}^2 + U^2)$$

$$2. \text{ توزيع التنبوء } (X_t / F^t) \sim N(m_{t-1}, \alpha_{t-1}^2 + U^2 + V^2)$$

$$3. \text{ التوزيع اللاحق } (\mu_t / F^t) \sim N(m_t, \alpha_t^2)$$

عندما تحقق دالة التنبوء بفترة واحدة للامام الصيغة التكرارية التالية [12]:

$$m_t = m_{t-1} + [(\alpha_{t-1}^2 + U^2) / (\alpha_{t-1}^2 + U^2 + V^2)](X_t - m_{t-1})$$

4.2 التقدير التكراري بتنقية كالمان

في هذا المبحث يتم استخدام اسلوب تنقية كالمان لنموذج ARIMA(0,1,1) لغرض المقارنة بين الخوارزميات التكرارية. في التنبوء لا يمكن تجاهل الطبيعة التكرارية للمعادلات الاساسية.

وهناك طريقة شائعة تتضمن تلك الطبيعة التكرارية وتعرف بتقنية كالمان (Kalman Filtering(KF)، المقترحة اصلا من قبل الباحث كالمان [6] عام 1960. وفي السنوات الاخيرة، لعبت تمثيلات فضاء الحالة وخوارزميات KF لعمليات معينة دور مهم في مسائل التقدير والتنبؤ. أذ ان تطوير هذه النظرية يتكون من معادلتين، عرفت كمعادلة المشاهدة ومعادلة الحالة وكما مبين ادناه :

بافتراض ان معادلة المشاهدة في الحالة القياسية او الكمية (the scalar case) تكون

$$Y_t = G_t X_t + v_t \quad \dots(4.1)$$

عندما يكون كل من G_t و X_t متجهات مختارة بصورة ملائمة وان $v_t \sim N(0, U^2)$ نشوء الوقت لمتجه الحالة X_t يتبع الصيغة

$$X_{t+1} = F_t X_t + w_t \quad \dots(4.2)$$

عندما تكون F_t مصفوفة ملائمة و $w_t \sim N(0, V^2)$. وباعتماد نموذج ARIMA(0,1,1) في الصيغة (3.7) ، وللتوضيح تكتب صيغة فضاء الحالة اعلاه باستخدام المتجهات والمصفوفات

$$F_t = (-\theta, 1, 1) = F$$

و

$$X_t = (X_{t-1}, X_t, Y_{t-1})'$$

وبتطبيق خوارزمية تنبؤ كالمان القياسية [4] ، نحصل على الصيغة التكرارية لدالة التنبؤ بفترة واحدة للامام Z_t وكمايلي [8] :

$$Z_t = aX_t + (1-a)Z_{t-1} \quad \dots(4.3)$$

عندما تكون $a = \Omega / (\Omega + U^2)$ و $\Omega = \{V^2 + \sqrt{V^4 + 4U^2V^2}\} / 2$.

5. معيار المقارنة

في حالة الاستقرار لـ t كبيرة، تم افتراض بان الربح في المعلومة من مشاهدة جديدة يكون مساوي بالضبط للخسارة بالمعلومة كلما ننتقل من حالة الى حالة اخرى. ولاغراض المقارنة بين الاساليب المتقدم ذكرها، يتم استخدام المعيار الاحصائي متوسط مربعات الخطاء المحاذي (asymptotic MSE) (تحت شروط حالة الاستقرار) [8] ، ولكل اسلوب كما مبين ادناه :

- اسلوب الـ EWMA ، فان $MSE = (\lambda \sigma^2) / (2 - \lambda)$ عندما تكون $\lambda = (1 - \theta)$.
- اسلوب الـ DLM ، فان $MSE = (A + 1)U^2 + V^2$ عندما تكون $A = \{V(\sqrt{V^2 + 4U^2} - V)\} / 2$.
- اسلوب بيز، فان $MSE = AU^2$ عندما تكون $A = \{V(\sqrt{V^2 + 4U^2} - V)\} / 2$.
- اسلوب تنقيحة كالمان ، $MSE = \Omega + U^2$ عندما تكون $\Omega = \{V^2 + \sqrt{V^4 + 4U^2V^2}\} / 2$.

فضلا عن اعتماد r وتمثل نسبة تذبذب الاشارة المتقدم ذكرها في المبحث (3.3) .

6. الجانب التجريبي

لغرض تحديد تأثير الاساليب المتقدم ذكرها في الجانب النظري والمتمثلة بـ EWMA و DLM و بيز (Bayes) وتنقية كالمان (Kalman) من خلال حساب قيم متوسط مربعات الخطأ (MSE) وفق الصيغ المتقدم ذكرها في المبحث (5) ولكل اسلوب، فضلا عن اعتماد نسبة تذبذب الاشارة المتقدم ذكرها في المبحث (3.3) لغرض المقارنة. فقد تم افتراض عدة قيم للتباينات V^2 و U^2 لتوليد سلسلة السير العشوائي و سلسلة المشاهدات وفقاً للخطوات الاتية :

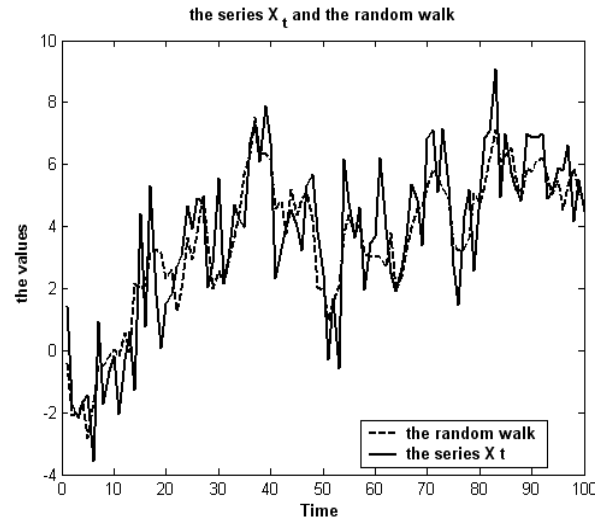
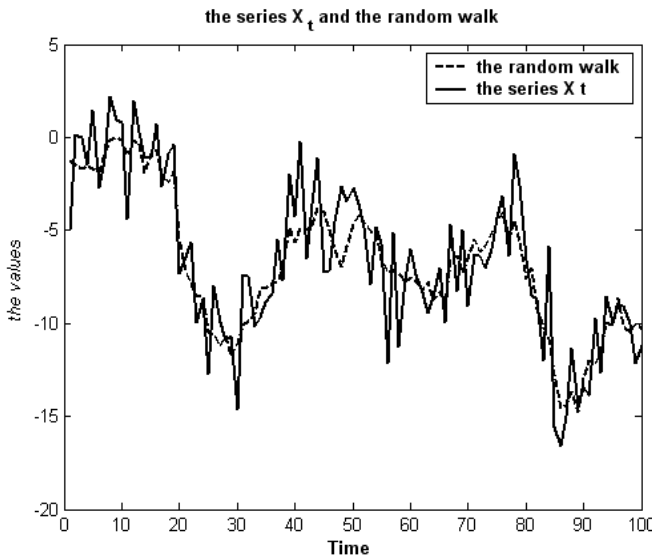
اولا- تم توليد سلسلة السير العشوائي (μ_t) وفق الصيغة (3.5) المتقدم ذكرها في المبحث (3.3) وهي $\mu_t = \mu_{t-1} + w_t$. عندما تكون $w_t \sim N(0, V^2)$ وبقيمة ابتدائية $\mu_0 = 0$ ، ومن خلال افتراض

قيمة V^2 وبحجم منة قيمة .

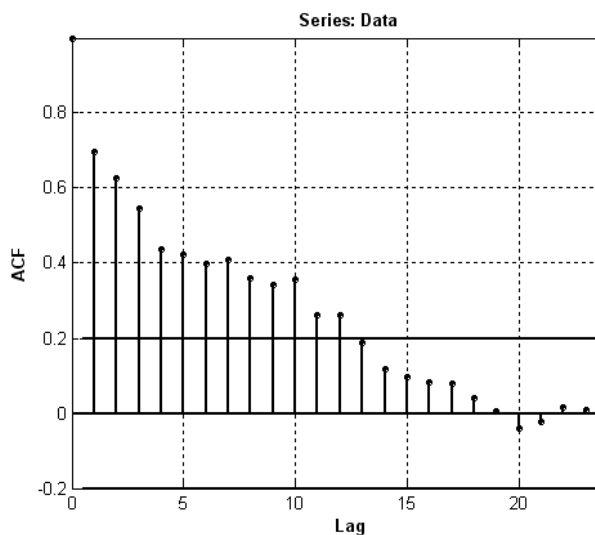
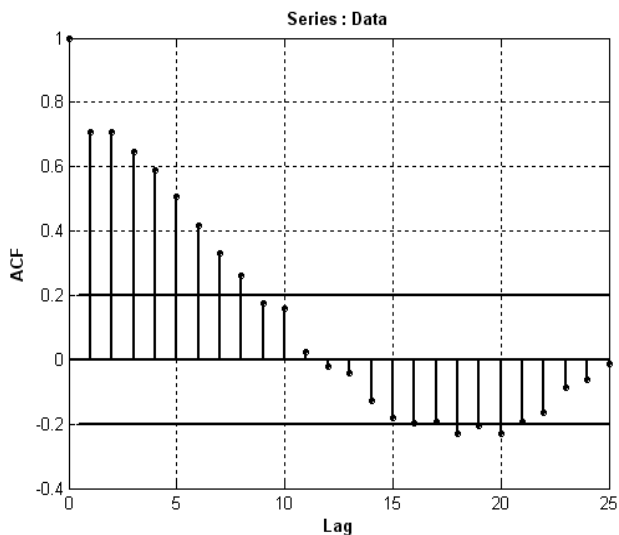
ثانيا - بعد توليد سلسلة μ_t يتم دمجها لتوليد سلسلة المشاهدات X_t وفق الصيغة(3.4) المتقدم ذكرها في المبحث (3.3) وهي $X_t = \mu_t + v_t$. عندما تكون $v_t \sim N(0, U^2)$ ، ومن خلال افتراض قيمة U^2 و بنفس الحجم للسلسلة μ_t .

ثالثا - بعد توليد سلسلة المشاهدات X_t يتم ايجاد سلسلة الفروق $(X_t - X_{t-1})$.

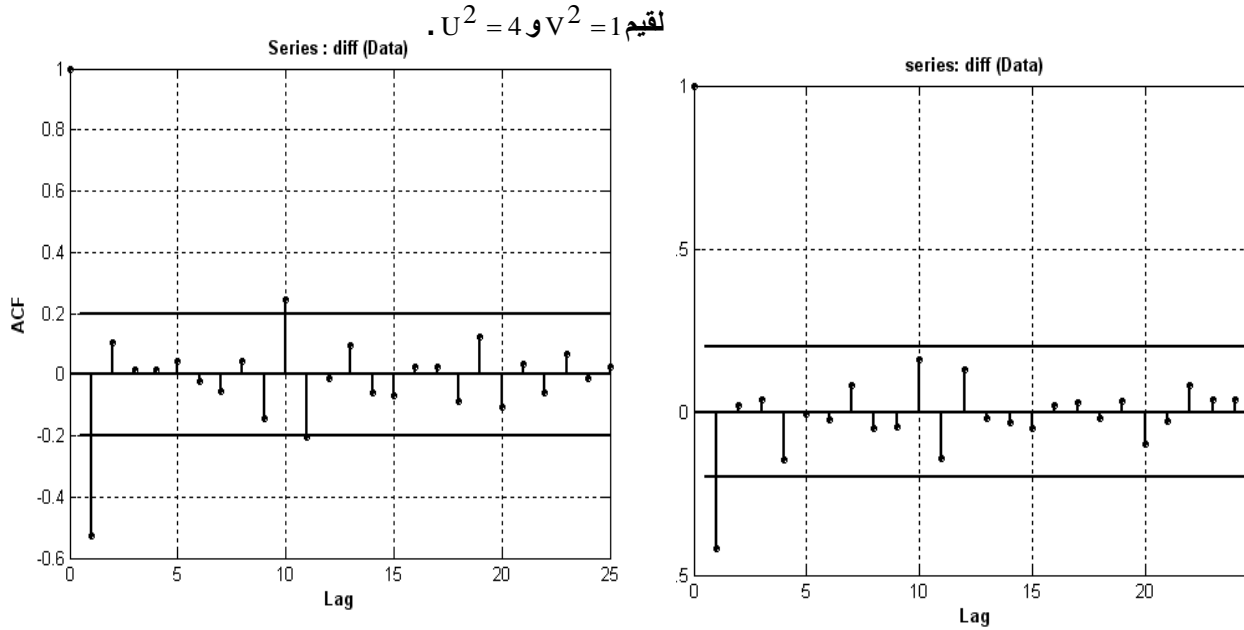
وقد كررت الخطوات اعلاه لتوليد السلسلاسل الزمنية وفقاً لعدة قيم لـ V^2 وهي $V^2 = 1, 2, 3, 4, 5, 6, 7$. وكذلك وفقاً لعدة قيم لـ U^2 وهي $U^2 = 2, 4, 6$ وبحجم منة قيمة . وقد استخدم الـ Matlab لتوليد السلسلاسل الزمنية وفق القيم المفترضة لـ V^2 و U^2 . ورسم سلسلة السير العشوائي μ_t مع سلسلة المشاهدات X_t وفق الصيغتين (3.5) و (3.4) وعلى التوالي ولمنة قيمة مولدة ، فضلا عن حساب سلسلة الفروق $(X_t - X_{t-1})$. نعرض بعض منها ، انظر الشكل (1.1) المرسوم وفقاً لقيم $V^2 = 1$ و $U^2 = 2$. فضلا عن رسم دالة الارتباط الذاتي (ACF) للسلسلة المشاهدات X_t في الشكل (1.2) وكذلك رسم دالة الارتباط الذاتي (ACF) للسلسلة الفروق $(X_t - X_{t-1})$ في الشكل(1.3) . ورسوم مماثلة بالاشكال (2.1) ، (2.2) ، و (2.3) وفقاً لقيم $V^2 = 1$ و $U^2 = 4$.



شكل (1.1) يبين سلسلة X_t و μ_t وفقا لقيم $V^2 = 1$ و $U^2 = 2$. شكل (2.2) يبين
سلسلة X_t و μ_t وفقا لقيم $V^2 = 1$ و $U^2 = 4$.



شكل (1.2) رسم دالة الارتباط الذاتي (ACF) للسلسلة X_t وفقا وشكل (2.2) رسم دالة الارتباط الذاتي (ACF) للسلسلة X_t وفقا لقيم $V^2 = 1$ و $U^2 = 2$.



شكل (1.3) رسم دالة الارتباط الذاتي (ACF) للسلسلة الفروق (ACF) للسلسلة الفروق (2.3) رسم دالة الارتباط الذاتي (ACF) للسلسلة الفروق $(X_t - X_{t-1})$ وفقا لقيم $V^2 = 1$ و $U^2 = 2$.

ولان سلسلة الفروق $(X_t - X_{t-1})$ تتبع عملية $MA(1)$ ، لذا فقد استخدم التطبيق الجاهز Minitab لتقدير النموذج $X_t - X_{t-1} = \varepsilon_t - \theta\varepsilon_{t-1}$ ولكل القيم المفترضة للتباينات V^2 و U^2 . اذ تم الحصول على القيم المقدرة لـ $\hat{\theta}$ و $\hat{\sigma}$ والمستخدم في حساب MSE لاسلوب الـ EWMA بالمعلمة $\hat{\lambda} = 1 - \hat{\theta}$. وبحساب قيم MSE للاساليب المستخدمة في البحث باستخدام الـ matlab لغرض المقارنة فضلا عن حساب نسبة تذبذب الاشارة (r) والمدونة في جدول (1) نلاحظ مايلى :

- تزداد قيم MSE وفقا لاسلوب DLM وبيزوتنقية كالمان بزيادة نسبة تذبذب الاشارة (r) (ولكل القيم المفترضة V^2 بثبات قيمة U^2).
- تزداد قيم MSE وفقا لاسلوب DLM وبيزوتنقية كالمان بتناقص نسبة تذبذب الاشارة ولكل القيم المفترضة U^2 بثبات قيمة V^2 .
- تكون قيم MSE لاسلوب الـ EWMA متذبذبة بزيادة ونقصان بزيادة r للقيم المفترضة لـ $V^2 = 2,3$ عندما تكون $U^2 = 2$. الا انه تزداد قيم MSE لاسلوب نفسه بزيادة r لبقية قيم $V^2 = 1,4,5,6$ عندما تكون $U^2 = 4,6$. وكذلك تناقص قيم MSE بزيادة r لقيم $V^2 = 3,7$ ولكل قيم U^2 .

- تزداد قيم MSE لاسلوب الـ EWMA بتناقص r ولكل القيم المفترضة U^2 بثبات قيمة V^2 ، ولكل قيمة من قيم $V^2 = 3,5,6$ عدا $V^2 = 2$. اذ تكون قيم MSE متذبذبة بزيادة وتناقص ولكل قيم U^2 بثبات قيمة V^2 عند احدى قيم $V^2 = 1,2,4,7$ المفترضة .
- ان قيم MSE لاسلوب الـ EWMA تكون اصغر مقارنة بقيم MSE لاسلوب DLM وتنقية كالمان ولكل القيم المفترضة لـ V^2 و U^2 .
- ان قيم MSE لاسلوب الـ EWMA تكون اصغر مقارنة بقيم MSE لاسلوب بيز ولكل القيم المفترضة لـ U^2 عدا قيمة $U^2 = 2$ ولكل القيم المفترضة لـ V^2 . اذ نلاحظ ان قيمة MSE لاسلوب بيز تكون اصغر من قيم MSE لـ EWMA عند قيمة $U^2 = 2$ ولكل قيم V^2 عدا القيمتين $V^2 = 1,3$ فان قيم MSE لـ EWMA تكون اصغر من قيم MSE لاسلوب بيز ولنفس التوليفتين المتقدم ذكرها . وبشكل عام فان اسلوب بيز يعطي اصغر MSE عندما $\theta \approx 0$ مقارنة بقيمة الاساليب .
- ان قيم MSE لاسلوب بيز تكون اصغر مقارنة بقيم MSE لاسلوب DLM وتنقية كالمان ولكل القيم المفترضة لـ V^2 و U^2 . وان قيم MSE لاسلوب بيز تكون اصغر مقارنة بقيم MSE لاسلوب تنقية كالمان ولكل القيم المفترضة لـ U^2 عدا قيمة $U^2 = 6$ ولكل القيم المفترضة لـ V^2 . اذ نلاحظ ان قيمة MSE لاسلوب كالمان عند تلك القيم ($U^2 = 2$ ولكل قيم V^2) تكون اصغر مقارنة بقيم MSE لاسلوب بيز .

جدول (1) يبين القيم المفترضة للتباينات V^2 و U^2 لسلسلة السير العشوائي وسلسلة المشاهدات

وعلى التوالي ونسبة تذبذب الاشارة (r) والقيم المقدرة (θ و σ) لسلسلة الفروق ($X_t - X_{t-1}$) ، وقيم متوسط مربعات الخطاء المحاذي (asymptotic MSE) لكل اسلوب .

NO.	V^2	U^2	r	$\hat{\theta}$	$\hat{\sigma}$	asymptotic MSE			
						EWMA	DLM	Bayes	Kalman
1	1	2	0.50	0.6316	3.275	0.739	5.000	2.000	4.000
2	1	4	0.25	0.5484	7.788	2.271	11.246	6.246	6.562
3	1	6	0.17	0.8145	7.276	0.744	19.000	12.000	9.000
----	2--	--	----	----	-----	-----	----	----	-----
4	2	2	1.00	0.3616	6.360	2.982	6.472	2.472	5.236
5	2	4	0.50	0.4463	7.382	2.826	14.000	8.000	8.000
6	--	6	0.33	0.6043	9.377	2.313	23.633	15.633	10.606
----	3	--	----	----	-----	-----	----	----	-----
7	3	2	1.50	0.4422	5.996	2.319	7.745	2.745	6.372
8	3	4	0.75	0.4295	8.042	3.209	16.099	9.099	9.275
9	--	6	0.50	0.4813	10.890	3.813	27.000	18.000	12.000
----	4	--	----	----	-----	-----	----	----	-----
10	4	2	2.00	0.288	5.859	3.232	8.928	2.928	7.464
11	4	4	1.00	9	12.720	5.964	17.889	9.889	10.472
12	--	6	0.67	0.3616	12.410	5.381	29.749	19.749	13.292
----	5	--	----	0.3951	-----	-----	----	----	-----
13	5	2	2.50	----	7.291	4.045	10.062	3.062	8.531
14	5	4	1.25	0.2863	10.930	3.841	19.494	10.494	11.623

15	--	6	0.83	0.4799	12.610	5.273	32.125	21.125	14.521
----	6	--	----	0.4103	-----	-----	----	----	-----
16	6	2	3.00	----	8.847	5.923	11.165	3.165	9.583
17	6	4	1.50	0.1980	8.648	5.408	20.978	10.978	12.745
18	--	6	1.00	0.2305	15.730	7.157	34.249	22.249	15.708
----	7	--	----	0.3746	-----	-----	----	----	-----
19	7	2	3.50	----	9.806	5.296	12.247	3.247	10.623
20	7	4	1.75	0.2986	11.930	7.037	22.377	11.377	13.844
21		6	1.17	0.2580	15.430	6.312	36.193	23.193	16.865
				0.4194					

7. الأستنتاجات

أن أهم الاستنتاجات التي تم التوصل اليها من خلال نتائج هذا البحث المتعلقة بمقارنة الاساليب المتعلقة بالتقدير التكرارية للبيانات المرتبطة ذاتيا والمتمثلة بـ EWMA و DLM وبيز وتنقية كالمان من خلال تأثير كل اسلوب على القيمة المحسبة لمتوسط مربعات الخطأ (MSE) نلاحظ الاتي :

- تكون قيم MSE لاسلوب الـ EWMA اصغر من قيم MSE لاسلوب DLM وبيز وكالمان ولكل القيم المفترضة للتباين V^2 عندما تكون $U^2 = 4,6$. فمثلا تكون قيمة MSE=23.633 لاسلوب الـ EWMA وهي اصغر مقارنة بقيمة MSE=15.633 لاسلوب DLM والقيمة MSE=10.606 لاسلوب تنقية كالمان عندما تكون $V^2 = 2$ و $U^2 = 6$. وبذلك فان تقريب البيانات المرتبطة ذاتيا بنموذج IMA (1, 1) يكون ملائم اكثر من بقية الاساليب المستخدمة في البحث .
- ان اسلوب بيز يعطي اصغر القيم لـ MSE مقارنة باسلوب DLM وتنقية كالمان ولكل القيم المفترضة للتباين V^2 وعندما تكون قيمة $U^2 = 2,4$. بالتحديد عندما تقترب القيمة المقدرة لـ θ من الصفر ($\theta \approx 0$) . فمثلا تكون قيمة MSE=6.246 لاسلوب بيز وهي اصغر مقارنة بقيمة MSE=11.246 لاسلوب DLM والقيمة MSE=6.562 لاسلوب تنقية كالمان عندما تكون $V^2 = 1$ و $U^2 = 4$.
- تكون قيم MSE لاسلوب تنقية كالمان اصغر من قيم MSE لاسلوب DLM واسلوب بيز ولكل القيم المفترضة للتباين V^2 وعندما تكون $U^2 = 6$. فمثلا تكون قيمة MSE=13.292 لاسلوب تنقية كالمان وهي اصغر مقارنة بقيمة MSE=29.749 لاسلوب DLM والقيمة MSE=13.292 لاسلوب بيز عندما تكون $V^2 = 4$ و $U^2 = 6$.

References:

1. Alwan, L.C and Roberts, H.V. (1988). Time series modeling for statistical process control, Journal of Business and Economic Statistics, 6, 87-95.
2. Abraham, B., Thavaneswaran, A. and Peiris, S. (1996). On the prediction for a nonlinear time Series models using estimating

- equations. Lecture Notes Monograph Series, VOL.32, 259-267. (I.Basawa and R.L.Taylor Eds).
3. Abraham, B., and J.Ledolter (1986). Forecast functions implied by autoregressive integrated moving average models and other related forecast procedures. *International Statistical Review*, 54(1), 5-66.
 4. Brockwell, P.J. and Davis, R.A. (1996). *An Introduction to time Series and forecasting*, Second Edition. SpriSnger Verlag, New York.
 5. Hunter, J.S. (1986).The Exponentially Weighted Moving Average. *Journal of Quality Technology*, Oct. 1989 VOL.18, No.4, 203-209.
 6. Kalman, R. E. (1960) .Anew approach to linear filtering and prediction problems. *J. Basic Engineering* 82, 35-45.
 7. Montgomery, D .C. and Mastrangelo, C.M. (1991) .Some statistical process control methods for autocorrelated data. *Journal of Quality Technology*, July. 1991, VOL.23, No.3, 179-268.
 8. Mellor, R., Allen, D. peiris, S. Thavaneswaran, A. (2003) .Applications of Recursive Estimation Methods in statistical process control .*Statistical Methohs* , 5(2),156-171.
 9. Peiris, M.S. (2003). Improving the Quality of Forecasting using Generalized ARMA Models: An Application to statistical Quality control. *International Conference on Statistics in Industry and Business*, January 2-4, Cochin University, India.
 10. Roberts, S.W. (1959). Control Chart Tests Based on Geometric Moving Averages. *Technometrics*, 1, 239-251.
 11. Thavaneswaran, A., Macpherson, B.D., Abraham, B.(1998). An Application of Filtering to Statistical Process Control. *Proceeding of the International Conference on Statistics in Industry and Business*, January 1998, Cochin University, India.
 12. West, M and Harrison, J.(1989).*Bayesian Forecasting and Dynamic Modelling*, SpriSnger Verlag, New York.