التنبؤ بمبيعات السكر في معمل الموصل باستخدام معيار اكياكي

مناهل دانيال عبد الأحد
قسم الرياضيات / كلية التربية
جامعة الموصل

القبول	الاستلام
2011 / 01 / 05	2010 / 11 / 01

ABSTRACT:

In this paper we display and compare several multivariate models for some multiple time series data. The models all come from the vector autoregressive moving average class so that comparisons between them can easily be made using criteria such as Akaike's information criterion. Which is case study model specification with multivariate time series: vector autoregressive moving average model.

الخلاصة:

في هذا البحث حاولنا استعراض ومقارنة عدة نماذج لبيانات سلاسل زمنية متعـددة. حيث ان هذه النماذج جاءت من متجهات الانحدار الذاتي والمتوسطات المتحركة وتم المقارنة بينهم بسهولة باستخدام معيار معلومات اكياكي. حيث تم دراسة حالة النموذج باستخدام سلاسل زمنية متعددة الذي يعنى خصوصاً بمتجه الانحدار الذاتي والمتوسطات المتحركة.

نبذة تاريخية:

لقد اهتم علماء كثيرون بدراسة السلاسل الزمنية[3] المتعددة نـذكر مـنهم & chan Wallis (1978) وكذلك Jenkis & Mcleod (1982) Heyse & Wei (1985) وهنـاك عدة بحوث تضمنته أمثلة عديـدة ,Tiao & Box (1981) Tiao & Tsay (1983). ومـن الجدير بالذكر ان العالم[4] Sims (1977) اقترح طريقة مختلفة لبناء نماذج عملية بالتعامـل مع جميع المتغيرات بالاعتماد على مطابقة متجهات الارتباط الذاتي غير المقيدة.

$$X_t = \sum_{t=1}^{M} \Psi_t X_{t-k} + U_t$$

X_t هو متجه لـ r من المتغيرات العشوائية

U_t هو متجه لـ r×1 من الاخطاء العشوائية

وأوضح Tsay (1984) [3] ان التوزيع المحاذي للرتبة التي تقابل قيمـة AIC قـائم عندما تكون السلسلة غير مستقرة ولذلك يمكن تطبيق الانحدار الذاتي VAR بدون اخذ الفرق للبيانات.

وكان Shibata (1976) [5] قد برهن كفاءة التنبؤ من خلال AIC. بمعنى ان هذا المعيـار يخفض متوسط مربعات خطا التنبؤ. وأوضح ان أختبارAIC يعتمد على دالة الامكان الأعظم التي تكون مؤثرة وغير متحيزة وعليه فان هذا الاختبار يعطي نتائج دقيقة تماما لـ $n \geq 30$ وان اختبار هذا المعيار اظهر نجـاح كبيـر $n \geq 20$ لـ .

واستخدم العالم Tong (1975)[9] هذا المعيـار لإيجـاد رتبـة سلاسـل مـاركوف (Markov chains) وفي تقدير عدد العوامل المؤثرة في التحليـل عنـد اسـتخدام التحليـل العاملي.

الجانب النظري:

نفرض لدينا المعادلة الخطية التالية التي تمثل سلسلة زمنية معينة [5]

$$X_n = a_1 X_{n-1} + a_2 X_{n-2} + \ldots\ldots.. a_k X_{n-k} + \varepsilon_n \qquad n=(\ldots,-1,0,1,\ldots..)$$

حيث ان

$a_1, a_2 \ldots\ldots a_k$ تمثل معلمات حقيقية لكي يتم حل المعادلات بها

$$x^k = \sum_{i=1}^{k} a_i x^{k-i}$$

والتي جميعها اقل من الواحد الصحيح في حالة اخذ القيم المطلقـة وان (ε_n) تمثـل سلسلة المتغيرات العشوائية المستقلة لنفس التوزيع الطبيعي $N(0, \sigma^2_\varepsilon)$ نطلق على النمـوذج اعلاه النموذج غير العام (المنمط) لرتبة kth اذا كان $a_k \neq 0$ ورتبة الصفر اذا كانت

$X_n = \varepsilon_n$ $(n=\ldots,-1,0,1,\ldots))$

ولنفرض ان رتبة الانحدار الممكنة محدودة لبعض الرتب k والتي دائمـا تكـون معلومة وان قيود او حدود الانحدار هي موجودة داخل حدود الـK. والمـشاهدات المعطـاة $(X_1,\ldots..,X_N)$ تقترب من مقدرات الامكان الاعظم لـ $\hat{a}_i(k)$ a_i $(i= 1,\ldots.. k)$ معرفـة بالدالة الخطية التالية

$$\begin{bmatrix} \hat{R}(1,1)............\hat{R}(1,k) \\ . \\ . \\ \hat{R}(k,1)...........\hat{R}(k,k) \end{bmatrix} \begin{bmatrix} \hat{a}_i(k) \\ . \\ . \\ \hat{a}_k(k) \end{bmatrix} = \begin{bmatrix} \hat{R}(0,1) \\ . \\ . \\ \hat{R}(0,k) \end{bmatrix}$$

عندما

$$\hat{R}(i,j)=\frac{1}{N}\sum_{n=k+1}^{N} X_{n-i}X_{n-j}$$

وان متوسط مربعات الخطأ

$$\hat{\sigma}_{\varepsilon}^{2}(k)=\frac{1}{N}\sum_{n=K+1}^{N}\{X_n-\hat{a}_1(k)X_{n-1}-........-\hat{a}_k(k)X_{n-k}\}^2$$

هي أيضا مقدر الإمكان الأعظم التقريبي لـ $\sigma^2{}_\varepsilon$.

معيار اكياكي Akaike's Information Criterion

بعد تكوين معادلة السلسلة الزمنية بالاعتماد على نموذج بوكس جينكـز فـي تكـوين المعادلة المطلوبة ستكون هذه المعادلة مكونة من k من المعلمات التي تمثل بيانـات ظـاهرة معنية استناداً الى خواص النموذج الملائم وهناك عدة معايير للارتباط الذاتي فقد اقترح العالم Akaike (1973) [3] معيار سمي باسمه لمعرفة افضل نموذج

$$AIC = nLn(\sigma^2{}_\varepsilon) + 2k$$

K: عدد المعلمات في النموذج

$\sigma^2{}_\varepsilon$: مقدار التباين للخطأ

n: عدد المشاهدات

ان النموذج الذي يمتلك اقل قيمة للمعيار Aic (k) يعتبر افضل نموذج طالما انه يقلل الفرق من الانموذج المعطى الى الانموذج الصحيح. وقد اعتمد هذا المعيار لكونـه أكثـر وأوسـع المعايير استخداما وملائمة للبيانات.

$$AIC = N\ Log\ |\varepsilon_a| + 2(number\ of\ parmeter)$$

وفي معظم التطبيقات الاحصائية[5] يتم استخدام نماذج الانحدار الذاتي للسلسلة الزمنية ذات الرتب المنتهية لانها تعطي نتائج واقعية عندما نقوم بعملية التنبؤ. الا ان صعوبة التحليل الاحصائي في تقدير الرتبة للنموذج وخاصة اذا كانت الرتبة التي تم تحديدها اقل من الرتبـة الحقيقية للمسالة يؤدي ذلك الى ان تقدير المعلمة لن يكون دقيق واذا كان اختيار الرتبة اعلـى من الرتبة الحقيقية سيؤدي الى ازدياد التباين. لذلك فان معرفة افضل نموذج هـو مـن اهـم مراحل التحليل وما تم في هذا البحث اننا استخدمنا معيار اكياكي لمعرفة افضل نموذج

نفـــرض 0= $\hat{a}_i$ (k) لأي $k< i \le K$, وتكتب بالصيغة التالية: $\hat{\sigma}^2_\varepsilon(0) = \hat{R}(0,0)$

وكمـا $(k= 0,1,\ldots K)$ K كمتجهات ذات بعد $\hat{a}'(k)= \{\hat{a}_1 (k) \ldots\ldots \hat{a}_k (k),0,\ldots,0\}$

هو معروف في (Akaike,1973) فان â (k) هي تقدير امثل لـ â (K) ذات بعد k

$$\|X\|^2 = X'\hat{R}X = \sum_{i=1}^{K}\sum_{j=1}^{K} x_i x_j \hat{R}(i,j)$$

$$< x, y >= x'\hat{R}y, \quad \hat{R} = \{\hat{R}(i,j), 1 \le i, j \le K\}$$

هي مصفوفة التباين المشترك وان x,y متجهات ذات بعدk عندئذ الفرق بـين $\hat{a}(k)$ و$\hat{a}(q)$ يعطي متوسط مربعات الخطأ

$$\|\hat{a}(k) - \hat{a}(q)\|^2 = \hat{\sigma}^2_\varepsilon(q) - \hat{\sigma}^2_\varepsilon(k)$$

وان العملية الحقيقية $(X_n \ ,n=1,\ldots, N)$ هي عامة وذات نموذج برتبة k_0th وبمعلمات σ^2_ε, $a' = (a_1,\ldots,a_{k0},0\ldots,0)$, علما ان $\hat{\sigma}^2_\varepsilon(k)$ $\hat{a}(k)$ هي ثابتة وباســتخدام طريقـة اكياكي نحدد الرتبة

$$AIC(m) = NLog\ \hat{\sigma}^2_\varepsilon(m) + 2m \qquad (m = 0,1,2\ldots\ldots\ldots\ldots, K)$$

والانجاز ان يكون في حده الادنى كدالة لـ m

ومثلما يتم استخدام الطريقة الاعتيادية القديمة لإيجاد التنبؤ للسلسلة الزمنية (Box,Jen) هناك عدة طرق لبناء نموذج سلسلة زمنية متعددة[3] $z_{it} = (z_{1t}, z_{2t}, z_{3t})'$

وان

Z_{it} تمثل سلسلة زمنية احادية وبدمج نموذج الانحدار الذاتي والمتوسطات المتحركة يتكـون لنا نموذج ARIMA:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots\ldots\ldots \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots\ldots\ldots\ldots - \theta_q a_{t-q}$$

حيث ϕ_i, θ_j هي معلمات للمصفوفة (k×k)

a_t سلسلة متغيرات عشوائية لتوزيع *(IID)* متماثلة متطابقة قطرية $N_k(0, \varepsilon_a)$ وان ε_a ليس بالضرورة ان تكون قطرية وهذا ما يسمى بنموذج *VARMA (p,q)* وتكتب أيضاً بالـصورة التالية

$$\phi(\beta)\, z_t = \theta(\beta)\, a_t$$

حيث

$$\phi(\beta) = 1 - \phi_1\beta - \phi_2\beta^2 - \ldots\ldots\ldots - \phi_p\beta^p$$

$$\theta(\beta) = 1 - \theta_1\beta - \theta_2\beta^2 - \ldots\ldots\ldots - \theta q\beta^q$$

متعددة الحدود لعامل الإزاحة الخلفي β

$\beta z_t = z_{t-1}$

وهناك عدة شروط لتحقيق استقرارية السلسلة متعددة الحدود[3] [1] على سبيل المثـال ان z_t تكون مستقرة عندما تقع الجـذور لـ ـ $|\phi(B)| = 0$ خارج دائرة الوحدة (|A| تمثل المصفوفة المحددة A. لإنشاء نموذج للسلسلة الزمنية z_t يجب ان تحدد الرتبة (p,q) ثم بعد ذلك يتم التقـدير للمعلمـات $(p+q)^{k2}$ ويفـضل تقليـل العـدد فـي معامـل المـصفوفات (ϕ_i, θ_j) و $1/2k(k+1)$ في ε_a باستخدام بعض البرامج. هنـاك مجموعـة بـرامج عدديـة (NAG) لإيجاد مقدرات الإمكان الأعظم التامة وكذلك برامج MTS التـي تطبـق طريقـة المربعات الصغرى ولكن عند إنشاء نماذج VARMA نواجه مشكلتين الأولى عنـد تحديـد الرتب حيث معظم التقنيـات تـستخدم دالـة الارتبـاط الـذاتي cross Autocorrelation function (CCF) التي تقيس العلاقة بين كل مكونات السلـسلة ودالـة الارتبـاط الـذاتي الجزئي(PCCF) (partial cross Autocorrelation function) تقيس نفس العلاقة عنـد الازاحة h ولكن باستبعاد او إهمال اي تأثير يعود الى ما بـين الفتـرات المتخلفـة زمنيـا. والمشكلة الأخرى ترجع الى العدد الكبير للمعلمات الموجودة في نمـاذج VARMA والتـي معظمها تكون زائدة واي تقدير لها يعطي ارتباط عالي يقود الى تقديرات غير صحيحة.

مقارنة نماذج VARMA مع ARIMA

ان تحليل السلاسل الزمنية المتعددة [8] [3] تكون مهمة لبناء نماذج احادية لكل سلـسلة بدائية ولرؤية التطورات لاي نموذج متعدد على فرض اعتماد نمـاذج ARIMA الاحاديـة ومقارنها مع VARMA (Z_{1t},Z_{2t},Z_{3t}). عند بناء نمـاذج VARMA يواجهنـا مـشكلتين الاولى تحديد الرتبة (p,q) (تشخيص النموذج) مقارنة النموذج Model comparison يمكن مقارنة نماذج VARMA الكاملة للبيانات باستخدام معيار اكيـاكي (AIC) يمكـن مقارنـة النموذج الكامل VARMA لكل البيانات الموجودة لدينا باستخدام معيار اكياكي.

$$AIc = -2\,Log\,(\max imized\ Likelihood) + 2\,(Number\ of\ Parameters)$$

هذا المقياس[3] بكافة معلماته يستخدم لتحديد مدى ملائمة أي نموذج ويفضل النموذج الذي لـه اقل قيمة لهذا المعيار نموذج VAR(2) كانت كفوءة لقياس نموذج ولكن فقط لعدد معين مـن المعالم (حيث اعطينا لكل معلمة وزن (2) كما عرفناها في المعادلة السابقة)

أي عدد المعلمات الممكن اخذها لتكوين النموذج بالحدود المقيـدة + تلـك الموجـودة فـي مصفوفة التحويل وكذلك يمكن مقارنة النموذج الثالث والسادس مع النماذج الاحادية الثلاث.

الجانب التطبيقي

الجداول 3،2،1 تمثل سلسلة الإدخال إنتاج السكر وسلسلة الاخراج مبيعات السكر في معمل الموصل للفترة (1979-1972) حيث ان سلسلة الادخال تتوافق مع سلـسلة الاخـراج لتكون سلاسل زمنية متعددة حيث ان السلاسل كانت مختلفة في التباين ولها تذبذبات قوية تـم معالجتها باخذ اللوغاريتم الطبيعي ليتم تحقيق الاستقرارية في التباين امـا الاسـتقرارية فـي المتوسط فتم تحقيقها من خلال رسم معاملات دالة الارتباط الذاتي الذي يتبع التوزيع الطبيعي بمتوسط (صفر) وانحراف معياري ($\frac{1}{\sqrt{n}}$) عندما تكون السلسلة ذات طبيعـة عـشوائية وتـم الوصول الى استقرارية السلسلة (الادخال) بعد اخذ الفرق الاول لتكون جميع المعاملات ضمن حدود الثقة.

وبالرغم من اخذ الفرق الثاني عشر لبيانات سلسلة الادخال واستخراج معامل الارتباط الذاتي لها الا اننا لم نستطيع تحديد رتبة للنموذج.

وبالنظر لصعوبة تحديد النموذج المناسب فقد تم اعتماد أسلوب آخر في تحديد النموذج وذلك بتوفيق نماذج ملائمة وقريبة من سلوك دالة الارتباط الذاتي حيث تم توفير ستة نماذج

ARIMA (1,1,0), ARIMA (0,1,1), ARIMA (1,0,1)
MRIMA (0,1,2) , ARIMA (2,1,0),
SARIMA $(0,1,1) \times (0,1,1)_{12}$

وكان رسم السلسلة مقارب جدا للسلسلة الأصلية وتم استخراج قيم متوسط مربعـات الخطـا MSE ومتوسط نسب الاخطاء المطلقة MPE حيـث لاحظنـا ان النمـوذج الملائـم هـو ARIMA (0,1,2) حيث اعطى اقل قيمة لـ MPE, MSE

$$MSE = \frac{\sum (x_t - F_t)^2}{n}$$

$$|MPE| = \frac{\dfrac{\sum \left| x_t - F_t \right|}{x_t}}{n}$$

وبما ان Y_t هي قيم لفروق السلسلة $L_n x_t$ يمكن اعتبار النموذج الملائم لـ X_t هـو MA(2) بمعلمتين $\Theta_1=0.7$ ،$\Theta_2=0.1$ كمرحلة اولى.

وبعد ذلك تم إجراء عدد من التكرارات للحصول على معالم أكثـر دقـة وحـساب متوسـط مربعات الخطا عند كل تكرار الى ان وصلنا الى $\Theta_1=0.651, \Theta_2=0.136$.

والنموذج كان كالآتي:

$$L_n x_t = (1- 0.651\beta - 0.136\beta^2)\alpha_t$$

or

$$Y_t = \alpha_t - 0.651\alpha_{t-1} - 0.136\alpha_{t-2}$$

وتم إجراء سلسة من الاختيارات لمعرفة كفاءة النموذج. وإجراء نفس الخطوات السابقة على سلسلة الإخراج اي طبق نموذج ARIMA(0,1,2) على بيانات سلسلة الإخراج

$(1-\beta)\ L_nY_t = (1-0.651\beta - 0.136\ \beta^2)\ \beta_t$

$Y_t = \beta_t - 0.651\ \beta_{t-1} - 0.136\ \beta_{t-2}$

$\beta_t = Y_t - 0.651\ \beta_{t-1} + 0.136\ \beta_{t-2}$

$\beta_2 = -0.19244$

$\beta_3 = 0.18333$

-

-

-

$\beta_{44} = 0.20108$

نستطيع مقارنة جميع نماذج VARIMA باستخدام معيار Akaike

AIC= -2log (maximized likelihood) +2 (number of parameters)

وهذا هو افضل مؤشر يزودنا بعدد المعلمات التي استخدمناها ويفضل النموذج الذي له اقل قيمة لهذا المعيار وباستخدام هذا المعيار لوحظ ان نموذج VAR(2) هو نموذج كفوء للبيانات.

Aic for (likelihood form)

سلسلة الإدخال

Model	LoG- likelihood	no. of parameter	AIC
ARIMA (0,1,2)	688	14	-1.37
ARIMA (1,1,0)	672	13	-1.366
ARIMA (0,1,1)	680	6	-1.38
ARIMA (1,1,1)	679	9	-1.39
ARIMA (2,1,0)	681	10	-1.3
SARIMA (0,1,1)	678	8	-1.28

نلاحظ ان نموذج (0,1,2) ARIMA هو أحسن نموذج باستخدام معيار (AIC) واستخدمنا أيضا هذا المعيار لقياس نموذج VAR.

Model	Nlog	Parameter	AIC
VAR (2,0)	-2267	14	-2246
VAR (1,3)	-2237	6	-2234
Transfer function (2)	-2220	1	-2216
Univariate (1)	-1792	2	-1788

AIC= N Log $|\varepsilon_a|$ *+ 2 (number of parameters)*

حيث $|\varepsilon_a|$ تم وضعها بدلا من الامكان الاعظم كمقياس للكفاءة ووجد ان VAR(2) هو الافضل.

جدول (1): يبين قيم السلسلة الزمنية الخاصة بالمبيعات الشهرية للسكر الأبيض (بالطن)

شهر / سنة	ك2	شباط	اذار	نيسان	مايس	حزيران	تموز	اب	ايلول	ت1	ت2	ك1
1972	1979	3336	2752	3747	4183	3848	3907	8394	3916	5885	5314	4963
1973	4589	5931	4707	6798	6324	7861	5889	6341	5938	6533	5703	5122
1974	5341	5713	5206	6717	6995	4165	3622	4699	4219	5906	6261	5744
1975	8450	8075	6311	5255	2085	3838	6822	6020	3353	2797	3529	6225
1976	8896	6909	10197	9828	11956	6576	4277	4348	3716	2917	5325	4424
1977	4784	9087	8590	11833	9709	6057	9243	9054	7899	3373	5036	5021
1978	7239	7803	8863	9064	9011	5643	8282	11654	7667	10243	5096	5827
1979	6443	8318	7968	9866	8864	10375	6242	6197	5495	2145	3231	5932

جدول (2): يبين قيم السلسلة الزمنية الخاصة بالإنتاج الشهري للسكر الأبيض (بالطن)

شهر / سنة	ك2	شباط	اذار	نيسان	مايس	حزيران	تموز	اب	ايلول	ت1	ت2	ك1
1972	2681	1807	4063	4588	4192	4409	7164	6424	6337	6730	6497	5724
1973	6659	3234	4554	8264	7563	7239	7760	7475	8110	6336	7307	5736
1974	6719	7078	6693	4887	5474	6532	5943	6017	5806	3040	5854	6153
1975	8720	6768	5927	6930	5931	6114	6582	5355	6017	4536	8914	9492
1976	8511	7487	7085	6274	9158	6684	8195	5899	6772	6955	5228	9407
1977	10116	8022	8691	8469	8302	8718	8014	8418	4104	6845	6388	9739
1978	8497	7616	8250	10276	9395	6818	8060	9314	7805	8660	8593	7318
1979	7833	6833	6568	9812	9564	9007	7597	5300	7301	4727	8384	6278

جدول (3): يبين قيم السلسلة الزمنية الخاصة بالانتاج الشهري (للمولاس النهائي)

شهر / سنة	ك2	شباط	اذار	نيسان	مايس	حزيران	تموز	اب	ايلول	ت1	ت2	ك1
1972	623	1055	2047	1252	1133	575	606	454	519	203	392	650
1973	2339	1256	774	373	551	482	560	663	702	707	691	1092
1974	813	365	1458	1747	1322	384	486	423	922	737	694	331
1975	1267	1092	1705	2286	2147	3098	448	539	1362	147	727	854
1976	848	1837	1690	1602	1926	510	673	247	960	627	432	2043
1977	1241	863	854	2008	930	350	520	611	1042	724	199	854
1978	2489	2690	2053	3171	2351	1790	629	882	1793	1754	1476	116
1979	1516	1913	2643	2940	3105	1836	765	1600	1400	640	680	1156

الاستنتاجات :

نستنتج مما سبق اعلاه ان نماذج بوكس جينكز وبالرغم من كفائتها العالية في تحديد رتبة النموذج تمهيدا للتنبؤ لاي سلسلة احادية او متعددة الا انها تلاقي صعوبة فـي تحديـد رتبـة النموذج الاكفأ او رتبة الارتباط الذاتي الخاص بالنموذج الذي تم اختياره حيث تـم اسـتخدام نموذج بوكس جينكز لاستخراج رتبة النماذج وعند المقارنة بين النتائج كان التنبؤ باسـتخدام معيار معلومات اكياكي هو الافضل. نوصي بإتباع معايير تكميلية لاختبار افضل نموذج لـه ارتباط ذاتي خاص باي سلسلة ليتم التنبؤ على اسس صحيحة وهذا ما تم فـي بحثنـا هـذا باستخدام معيار اكياكي كمعيار مكمل او ضروري عوضا عن استخدام نموذج بوكس جينكز.

REFERENCE:

1) Box, G. E. P, and HAUGH, L. (1977), "Identification of Dynamic Regression models connecting Two Time series," Journal of the American statistical association, 72, (121-130).

2) Cavanaugh, J. E. (2000), "Unifying the derivations for the Akiake and corrected Akaike information criterian "Department of statistics, university of Missouri,

3) GRUBBI, H. (1992), "A Multivariate Time series Analysis of some flour price Data", Applied statistics, 41,pp. (95- 107)

4) IAO, C. (1979), "Autoregressive Modeling of Canadian Money and Income Data" Journal of the American Statistical Association, 74, (553-560).

5) Shibata, R. (1976), "Selection of the order of an Autoregressive model be Akaike's criterion information" Biometrika, 63, pp.117-126.

6) Shuha, H. (2007), "Akaike information criterion" center for Research in Sceintific computation, north Carolina state University Raleigh, NC.

7) Taylor, C. (1987), "Akaike information criterion and the histogram" Biometrica, 3, 636- 639

8) TIAO, G. C. and Box, G. E. P. (1981), "Modeling Multiple time series with applications" J. AM. Statist. 76 (802- 816)

9) Tong, H. (1975), "Determination of the order of a markov chain by Akaike's information criterion" J. Appl. Prob,12, pp. 480- 497.