

استخدام سلاسل ARIMA وتقنية الـ PCA في التنبؤ لإنتاج ومساحة الحمضيات في سورية

ابتسام جاسم⁽¹⁾ * ومحمد طاهر عنان⁽²⁾

(1) الهيئة العامة للبحوث العلمية الزراعية ، دمشق، سورية.

(2) قسم الإحصاء الرياضي، كلية العلوم، جامعة حلب، حلب ، سورية.

(*للمراسلة: دابتسام جاسم. البريد الإلكتروني: e_sam_0@hotmail.com)

تاريخ القبول: 2020/05/31

تاريخ الاستلام: 2020/05/16

الملخص

يعتبر التنبؤ بقيم المتغيرات الاقتصادية من أهم الأهداف الأساسية للدراسات الاقتصادية الكمية في المستقبل ، من أجل التخطيط ورسم سياسات الإنتاج والأمن الغذائي، وقد تستخدم طرائق متعددة لغرض الحصول على التنبؤات الاقتصادية تم في هذه الدراسة استخدام نماذج بوكس جينكينز ARIMA التي تجمع بين أسلوب الانحدار الذاتي والمتوسط المتحرك للسلسلة الزمنية مع تقنية الـ PCA من أجل التنبؤ بإنتاج الحمضيات ومساحتها في سورية، حيث يمتاز هذا النموذج بدقة عالية في تحليل السلاسل الزمنية ، واستخدمت في هذا البحث بيانات سنوية للإنتاج والمساحة في سوريا ، للفترة (1970-2018)، وتبين أن النموذج الملائم للإنتاج ARIMA(5,2,0) هو افضل نموذج للتنبؤ بإنتاج الحمضيات حتى عام 2021 وفقاً للاختبارات الاحصائية، وكانت قيم دقة النموذج للإنتاج بدقة تساوي % 85.3 للبرتقال و% 88.3 للليمون وقد تفوق النموذج المقترح على كل النماذج التقليدية. بينما كان أفضل نموذج للتنبؤ بالمساحة هو: ARIMA(1,1,0) اعطى دقة نموذج مقدارها % 92.7 للمساحة المزروعة بالبرتقال و% 94.2 للمساحة المزروعة بالليمون وتفوق النموذج على كل النماذج التقليدية.

الكلمات المفتاحية: البرتقال، الليمون، المكونات الأساسية، التنبؤ، بوكس جينكينز.

المقدمة

تشكل التوقعات الاقتصادية عنصراً هاماً في المجالات التخطيطية لإمكانية تخصيص الموارد المتاحة، ورسم السياسات لتوفير الإنتاج الكافي لمواجهة الاحتياجات البشرية من الغذاء ، اهتمت سورية بزراعة الحمضيات والأشجار المثمرة منذ فترة طويلة، وتتصدر زراعة الحمضيات (البرتقال والليمون) مجموعة الاشجار المثمرة ما عدا الزيتون، وتهتم الحكومة بتحسين إنتاجية هذا النوع من الفاكهة لأغراض الأمن الغذائي.

بلغت مساحة البرتقال والليمون المزروعة في سورية 33 ألف هكتار للعام 2018، وشكلت حوالي %76.7 من إجمالي المساحة المزروعة بالحمضيات في سوريا، وحوالي %3.2 من إجمالي المساحة المزروعة بالأشجار المثمرة في سوريا،

وحوالي 0.58% من إجمالي المساحة المستثمرة في سوريا (وزارة الزراعة والإصلاح الزراعي، 2018)، ويتم زراعة الحمضيات في المنطقة الساحلية في كل من محافظتي (طرطوس واللاذقية).

عملت الحكومة السورية في كل الاوقات على اتباع استراتيجية كاملة للتنمية الزراعية تهدف إلى زيادة الانتاج والإنتاجية، من خلال اعتماد التنمية الرأسية في الإنتاج وذلك بتحسين ظروف إدارة العمليات الزراعية وتطوير واستخدام المدخلات في عملية الإنتاج للمساهمة في زيادة إنتاجية وحدة المساحة، وتطوير واستخدام الأصناف والأصول عالية الإنتاجية والمقاومة للأمراض.

قامت دراسة أجراها (جماح، 1985) بعنوان " تطبيق الاستشعار عن بعد في رصد وتقييم التغيرات بساحل الحديدية- الجمهورية اليمنية " على استخدام بيانات الاستشعار عن بعد من خلال تطبيق تحليل المركبات الاساسية PCA على القمر الصناعي LANDSAT في دراسة ورصد الموارد الطبيعية في ساحل الحديدية الواقع في الساحل الغربي للبحر الأحمر باليمن، والحصول على الوضع الحديث لحالة الموارد الطبيعية فيه، تقييم حالة الغطاء الارضي، دراسة التغيرات التي حدثت بالمنطقة بين عامي 1989 و 2002، دراسة التغير المتوقع في استخدام الاراضي وتأثيره على الغطاء الارضي هناك. وقد أوضحت نتائج هذه الدراسة أن المركبتان الاساسيتان الاولى PC1 والثانية PC2 عادة ما تحتويان على معظم البيانات، أما المركبتين الثالثة PC3 والرابعة PC4 تحتويان على التغيرات الطفيفة المتبقية وأغلب الضوضاء. كما استخدم (عنان، 2009) في دراسة بعنوان " استخدام القيم الذاتية والأشعة الذاتية في تنبؤ مؤشرات الانتاج النباتي " تقنية تحليل المركبات الاساسية PCA لإنتاجية خمس محاصيل حيث تم اختصار هذه المتغيرات الخمس إلى اثنتين فقط، ثم بناء نموذج التنبؤ لهاتين المركبتين وتم التنبؤ بالإنتاجية بشكل أوتوماتيكي عن طريق الأشعة الذاتية والقيم الذاتية.

وفي دراسة اجرتها (عنبر، 2011) بعنوان " النمذجة الرياضية لتطورات الحالة السكانية في سورية الواقع وأفاق المستقبل، تم فيها استخدام سلاسل ARIMA في نمذجة إهدار سكان المحافظات السورية والتنبؤ المستقبلي لها. وفي بحث (ابراهيم وعصمان، 2013) بعنوان " تحليل المركبات الاساسية في نظام كشف التطفل الشبكي باستخدام بيانات NSL-KDD". حيث قام البحث على تصميم نظام كشف تطفل شبكي يعتمد على الخوارزمية الجينية وذلك بسبب زيادة التهديدات الامنية والتطفلات غير المعرفة، حيث تم استخدام تحليل المركبات الاساسية الـ PCA للتخلص من الميزات الفائضة وقليلة الفائدة بالاعتماد على مجموعة بيانات NSL-KDD وأظهرت النتائج انجازاً جيداً لطريقة الـ PCA لاختيارها ثلاث ميزات فقط ما يقلل تعقيد العمليات الحسابية ويقلل وقت التدريب والاختبار.

مشكلة البحث

تتمثل إحدى التحديات الكبرى التي تواجه الحكومات تلبية الطلب المتزايد على أغذية كافية ومغذية، وذلك نظراً للنمو السكاني، مما يتطلب البحث والتنبؤ بإنتاج الأشجار المثمرة ومنها الحمضيات للمساهمة في رسم السياسات الزراعية، وبالتالي مدى مساهمة الإنتاج في الأمن الغذائي على المدى الطويل.

هدف البحث يهدف البحث إلى التوصل لأفضل نموذج للتنبؤ بإنتاج البرتقال والليمون والمساحة المزروعة بهما، وذلك بتطبيق منهجية ARIMA وتقنية الـ PCA والتي تعطي نتائج تنبؤ ذات دقة عالية، وحتى في الحالات التي تكون النماذج الأخرى التقليدية غير فعالة في التنبؤ للمساهمة في رسم سياسات الإنتاج بناء على معلومات تنبؤية دقيقة.

مواد البحث وطرائقه

- **البيانات:** اعتمدت الدراسة على البيانات الثانوية السنوية من المجموعات الإحصائية الزراعية السنوية الصادرة عن وزارة الزراعة والإصلاح الزراعي بدءاً من عام 1970 وحتى عام 2018، وتم استخدام البرامج الإحصائية SPSS و EXCEL لتحليل البيانات وضغط السلاسل للتنبؤ.

- **أسلوب التحليل:** اعتمد البحث على المنهج التحليلي لتحليل السلاسل الزمنية بغرض التنبؤ بتطبيق نماذج الانحدار الذاتي المتكاملة مع المتوسطات المتحركة (ARIMA) Autoregressive Integrated Moving Average والذي طبقها كل من George Box و Jenkins Gwilym على السلاسل الزمنية عام 1970، والذي سميت باسميهما (نموذج Box-Jenkins) حيث يجمع منهجين مختلفين في معادلة واحدة، تتمثل المنهجية الأولى بنموذج الانحدار الذاتي Autoregressive (AR) حيث يعبر عن المتغير التابع (Yt) كدالة في القيم الماضية لنفس المتغير التابع (Yt-1) بالدالة التالية (عنان، 2014):

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \delta + e_t \quad (1)$$

حيث:

Yt: المتغير التابع عند الزمن t.

Yt-1, Yt-2, ..., Yt-p: القيم الماضية لنفس المتغير التابع عند الزمن t-1, t-2, ..., t-p على التوالي.

P: رتبة الانحدار الذاتي وتعبر عن عدد القيم الماضية المستخدمة أو تعني فترات التباطؤ.

$\Phi_1, \Phi_2, \dots, \Phi_p$: معالم الانحدار الذاتي المقدر.

δ : الحد الثابت.

et: الخطأ عند الزمن t.

أما المنهجية الثانية فهي نموذج المتوسط المتحرك (MA) Moving Average حيث يتم التعبير عن المتغير التابع (Yt-1) كدالة في قيم حدود الخطأ السابقة:

$$Y_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (2)$$

حيث:

μ : المتوسط.

et: حد الخطأ عند الزمن t.

et-1, et-2, ..., et-q: أخطاء الفترات السابقة المتعلقة بالمتغير (Yt).

q: رتبة المتوسط المتحرك وتشير إلى عدد قيم حد الخطأ الماضية المستخدمة في النموذج.

$\theta_1, \theta_2, \dots, \theta_q$: معالم المتوسطات المتحركة المقدر.

ومن المعادلتين (1) و (2) يتم صياغة نموذج الانحدار الذاتي مع المتوسطات المتحركة ARMA كالآتي:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \delta + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} \quad (3)$$

ويرمز للنموذج (3) بالصيغة الآتية: ARMA(p,q)

حيث تشير p إلى رتبة الانحدار الذاتي، و q إلى رتبة المتوسط المتحرك، وقد تم تطوير نظرية تحليل السلاسل الزمنية باستخدام معامل التأخير (B). فموجب طريقة (Box-Jenkins) فإن استخدام معامل التأخير (B) لفترة واحدة تكون العلاقة التالية (Mandal, 2005) و (Mandal, 1977):

$$B Y_t = Y_t - Y_{t-1}$$

وفي حال التأخير لفترتين سابقتين تعطى العلاقة التالية:

$$B(B Y_t) = B Y_t - Y_{t-1} - Y_{t-2}$$

وفي حالة التأخير لفترات متعددة فإن (B) يعطى بالعلاقة كالتالي:

$$B^j Y_t = Y_t - Y_{t-j}$$

يتطلب تقدير نموذج ARIMA أن تكون السلسلة الزمنية مستقرة Stationary، ويقصد بذلك أن المتغير التابع له متوسط وتباين ثابتين خلال الفترة الزمنية موضع الدراسة أما السلسلة الزمنية غير المستقرة Non stationary فتعرف بأنها السلسلة التي ليس لها تباين ثابت ولها اتجاه غير ثابت، أي متزايد أو متناقص (بري، 2002)، ومن المعلوم أن معظم السلاسل الاقتصادية غير مستقرة، فإذا تبين أن السلسلة غير مستقرة فيجب تحويلها لسلسلة مستقرة من خلال إيجاد الفرق الأول للمتغير كالتالي:

$$Y_t = \Delta Y = Y_t - Y_{t-1} \quad (4)$$

وإذا لم ينتج من الفروق الأولى سلسلة مستقرة فيمكن تكرار عملية الفروق حتى يتم الحصول على سلسلة مستقرة، ويعبر عن عدد مرات الفروق المطلوبة لتحويل السلسلة إلى سلسلة مستقرة بدرجة التكامل (Integrated) فيقال أن السلسلة متكاملة من الدرجة (d) وعليه يصبح نموذج (ARIMA)، والذي يتصف بثلاث رتب هي رتبة الانحدار الذاتي (p) ورتبة التكامل (d) ورتبة المتوسط المتحرك (q) ويرمز له بالرمز ARIMA(p,d,q).

المكونات الأساسية

تعتبر المكونات الأساسية متغيرات تمثل بتراكيب خطية غير مرتبطة من المتغيرات الأصلية X_1, X_2, \dots, X_p ومرتبطة بحسب نسبة تفسيرها للتباين الكلي "نسبة المعلومات التي تعطيها عن العينة" فالمكون الأول (First Principal Component) هو التركيب الخطي الذي يؤدي إلى تفسير أكبر نسبة من التباين، والمكون الثاني (Second Principal Component) يعبر عن العلاقة الخطية التي تؤدي إلى تفسير أكبر كمية تباين بعد المكون الأول بشرط عدم ارتباطه بالأول ويعبر المكون الثالث عن العلاقة الخطية التي تؤدي إلى تفسير أكبر نسبة تباين بعد المكونين الأول والثاني وهكذا... (Rencher, 2003).

وغالباً ما يكون إيجاد المكونات الأساسية ليس هدفاً بحد ذاته بل وسيلة لإجراء تحليل آخر كتحليل الانحدار مثلاً (وشرن، 1998)، ويمكننا إيجاد المكونات الأساسية بطريقتين مختلفتين:

باستخدام مصفوفة التباين Σ .

باستخدام مصفوفة الارتباط R .

الشكل الجبري:

ليكن لدينا المتجه العشوائي $X' = [X_1, X_2, \dots, X_p]$ له مصفوفة تغاير Σ ولتكن $A = (a_1, a_2, \dots, a_p)$ متجه من العوامل المساعدة بحيث يحقق:

$$\sum a_i a_j = \begin{cases} 1 & ; i = j \\ 0 & ; i \neq j \end{cases}$$

عندئذ يمكننا كتابة $Y = a'X$

بحيث تعطي التراكيب الخطية التالية

$$Y_1 = a'_1 X = a_{11} X_1 + a_{12} X_2 + \dots + a_{1p} X_p$$

ويأخذ $Var(Y_1)$ أعظم قيمة.

$$Y_2 = a'_2 X = a_{21} X_1 + a_{22} X_2 + \dots + a_{2p} X_p$$

ويأخذ $Var(Y_2)$ ثاني أعظم قيمة.

$$Y_i = a'_i X = a_{i1} X_1 + a_{i2} X_2 + \dots + a_{ip} X_p$$

ويأخذ $Var(Y_i)$ أعظم قيمة بعد $i-1$ قيمة.

$$Cov(Y_i, Y_j) = Cov(a'_i X, a'_j X) = 0 \quad ; i \neq j$$

وبأخذ العوامل المساعدة $a_i ; i = 1, 2, \dots, p$ على شكل متجهات ذاتية لمصفوفة التغاير Σ أو مصفوفة الارتباط R نحصل على ما يسمى المكونات الأساسية.

النتائج والمناقشة

في هذا البحث سوف نستخدم تقنية جديدة تتلخص في الخطوات التالية:

1. محاولة ضغط السلاسل (سلاسل الإنتاج والمساحة) والاستعاضة عنها بسلسلة واحدة إن امكن ذلك باستخدام تقنية المكونات الرئيسية ((Principal Component Analysis (PCA)).
2. بناء نماذج ARIMA على سلاسل المكونات الرئيسية.
3. استنتاج قيم سلاسل الإنتاج والمساحة عن طريق الأشعة الذاتية والقيم الذاتية التي هي أساس طريقة (PCA)، ويعود سبب اختيارنا لهذه التقنية أن سلاسل الحمضيات ترتبط ببعضها بشكل قوي مما يوفر لنا إمكانية ضغط السلاسل وبالتالي اختصار الزمن والحسابات وتجنب الكثير من الأخطاء.

أولاً: إنتاج الليمون والبرتقال

إذا رمزنا لإنتاج البرتقال بالرمز y_1 وإنتاج الليمون بالرمز y_2 ، وتطبيق الـ PCA على هاتين السلسلتين من خلال مصفوفة الارتباط بين إنتاج الليمون وإنتاج البرتقال نلاحظ أن المكون الأول يفسر 99.321% من تباين إنتاج البرتقال والليمون، بينما المكون الثاني يفسر فقط 0.679% وبحسب مقياس كايسر ومقياس نسبة التباين نهمل الثاني و نأخذ الأول فقط، الجدول (1).

الجدول 1. يبين مواصفات المكونات الأول والثاني لسلاسل الانتاج.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.986	99.321	99.321	1.986	99.321	99.321
2	.014	.679	100.000	.014	.679	100.000

المصدر: مخرجات برنامج SPSS.

و فيما يلي الاشعة الذاتية التي تربط المكونات بالسلسلتين y_1 و y_2 كما يظهر بالجدول (2).

الجدول 2. الاشعة الذاتية للمركبات.

	Component	
	1	2
برتقال	.997	.082
ليمون	.997	-.082

المصدر: مخرجات برنامج SPSS.

أي أن المكون الأول :

$$Pc1=0.997 y_{11} + 0.997 y_{22}$$

والمكون الثاني :

$$Pc2=0.082 y_{11} - 0.082 y_{22}$$

y_{11}, y_{22} هي السلاسل المعيارية من السلاسل y_1, y_2

ترتبط السلسلتان y_{11}, y_{22} بالمكون الأول بالعلاقات:

$$Y_{11}=0.52*pc1$$

$$Y_{22}=0.52*pc1$$

ومنه فإن السلاسل الاصلية:

$$Y_1=y_{11} * s(y_1) + m(y_1)$$

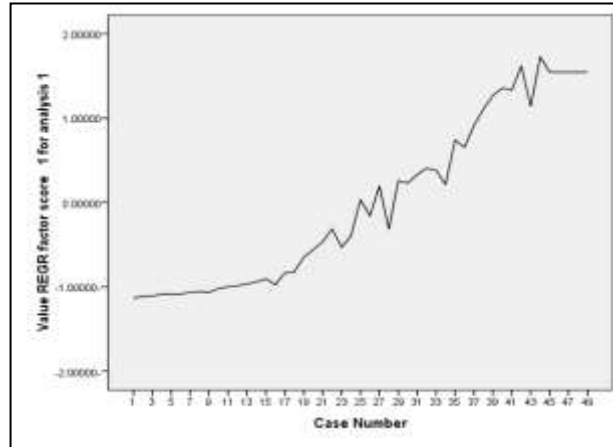
$$Y_2=y_{22} * s(y_2) + m(y_2)$$

حيث:

$m(y_1), m(y_2)$ - متوسط السلسلتين y_1 & y_2 على الترتيب.

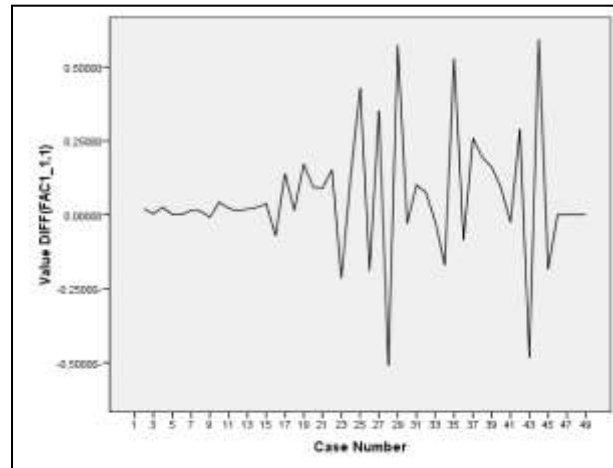
$S(y_1), s(y_2)$ الانحراف المعياري للسلسلتين y_1 & y_2 على الترتيب.

نقوم ببناء نموذج ARIMA للسلسلة $pc1$ ومن ثم نستنتج منها قيم السلاسل $y1$ & $y2$ كما هو موضح أعلاه. ويوضح الشكل رقم (1) الشكل العام لسلسلة المكون الأول.



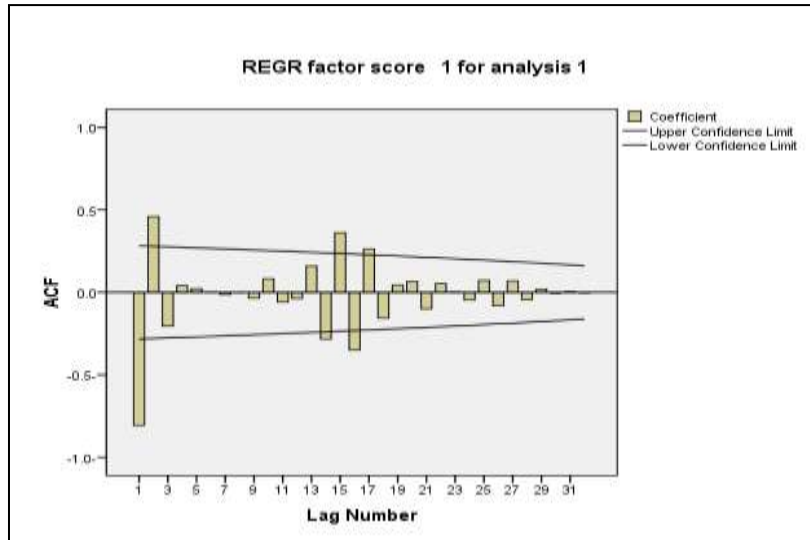
الشكل 1. الشكل العام لسلسلة المكون الأول.

يتضح أن السلسلة تملك اتجاه عام في الزيادة وبالتالي هي غير مستقرة وبأخذ الفروق الثانية نجد أنها أصبحت مستقرة، شكل (2).

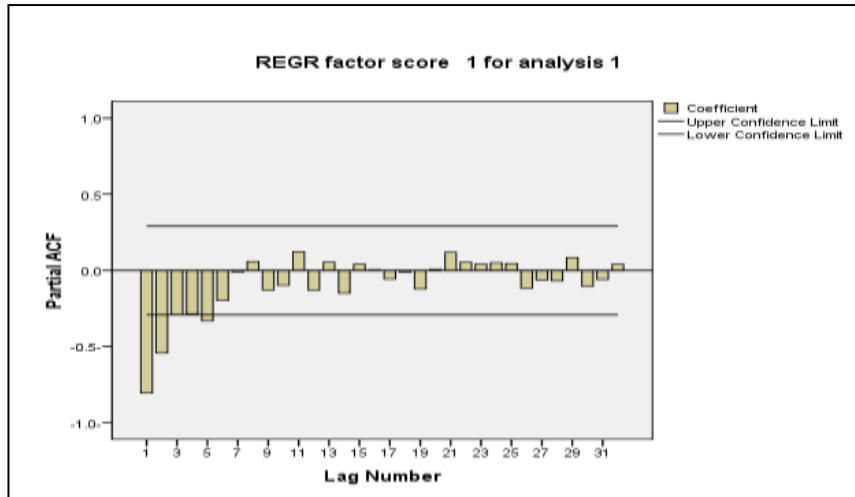


الشكل 2. الشكل العام لسلسلة الفروق الثانية للمكون الأول.

والشكلين (3 ، 4) يبينان معاملات الارتباط الذاتي والذاتي الجزئي.



الشكل 3 . معاملات الارتباط الذاتي لسلسلة الفروق الثانية للمكون الأول.



الشكل 4. معاملات الارتباط الذاتي الجزئية لسلسلة الفروق الثانية للمكون الأول.

من الشكل الأول نجد أن معاملات الانحدار الذاتي تتحد نحو الصفر بشكل جيبي متناوب مما يؤكد فعلا أن السلسلة أصبحت مستقرة بعد الفروق الثانية، ومن الشكل الثاني أن هناك معاملين خارج الحدود ثم اثنان على الحدود تماماً ومن ثم المعامل الخامس خارج الحدود لذلك سنفترض أن هناك انحدار ذاتي من الدرجة الخامسة (5) AR بينما لا يوجد متوسط متحرك لعدم وجود أي معامل متأخر خارج الحدود، بالتالي (0) MA وعليه أصبح النموذج المقترح هو $ARIMA(5,2,0)$ ، ويوضح الجدول (3) تقدير معاملات النموذج، ومنه نجد أن جميع معاملات النموذج معنوية وأيضا لدينا $R\text{-square}=0.97$.

الجدول 3. معاملات نموذج $ARIMA(5,2,0)$.

ARIMA Model Parameters					
		Estimate	SE	t	Sig.
Constant		0.000	0.004	-0.002	0.998
	Lag 1	-1.557	0.147	-10.579	0.000
	Lag 2	-1.356	0.261	-5.200	0.000
	Lag 3	-1.048	0.298	-3.515	0.001
	Lag 4	-0.774	0.273	-2.832	0.007

	Lag 5	-0.347	0.164	-2.113	0.041
Difference		2			

المصدر: مخرجات برنامج SPSS.

نجرب نموذج أقل من السابق وليكن: ARIMA(4,2,0) فنحصل على:

الجدول 4. معاملات نموذج ARIMA(4,2,0).

		Estimate	SE	t	Sig.	
No Transformation	Constant	.000	.006	-.084	.933	
	AR	Lag 1	-1.456	.148	-9.838	.000
		Lag 2	-1.120	.252	-4.449	.000
		Lag 3	-.657	.253	-2.597	.013
		Lag 4	-.269	.150	-1.802	.079
	Difference	2				

المصدر: مخرجات برنامج SPSS.

نلاحظ أن النموذج غير صالح لأن معامل الانحدار الذاتي الرابع غير معنوي.

نجرب نموذج أعلى من المقترح وليكن: ARIMA(5,2,1) والجدول رقم (5) يوضح تقدير معاملات النموذج،

ويتضح أن النموذج غير صالح لوجود عدة معاملات غير معنوية.

الجدول 5. معاملات نموذج ARIMA(5,2,1)

		Estimate	SE	t	Sig.	
No Transformation	Constant	9.725E-5	.004	.027	.978	
	AR	Lag 1	-1.208	.383	-3.155	.003
		Lag 2	-.848	.577	-1.471	.149
		Lag 3	-.658	.509	-1.292	.204
		Lag 4	-.546	.392	-1.391	.172
		Lag 5	-.252	.225	-1.124	.268
Difference	2					
MA	Lag 1	.409	.387	1.057	.297	

المصدر: مخرجات برنامج SPSS.

وبالتالي فإن النموذج المقترح ARIMA(5,2,0) هو الذي سيتم اعتماده في التنبؤ وبالتعويض في النموذج

المقترح والتنبؤ لثلاث سنوات قادمة (2019,2020,2021) كما هو موضح بالجدول (6).

الجدول 6. نتائج التنبؤ.

العام	Pc1	Y1	Y2
2019	1.67	750979.05	164266.33
2020	1.54	715392.55	156445.19
2021	1.64	741939.00	162279.52

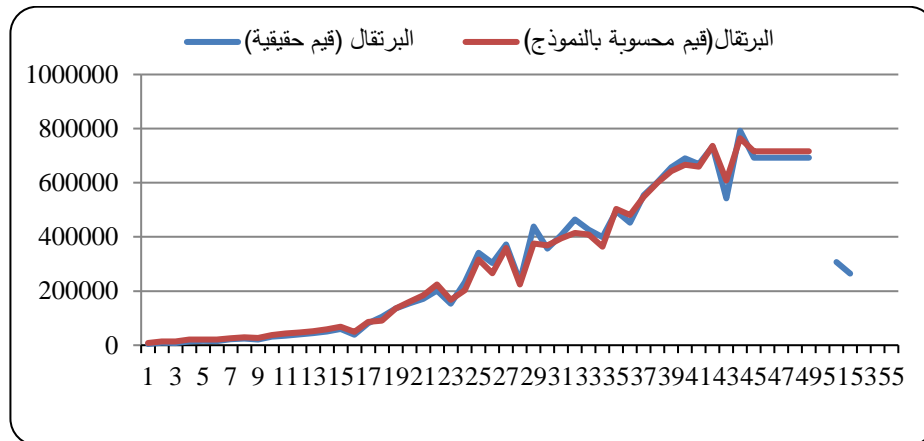
المصدر: مخرجات برنامج SPSS.

علماً أن دقة النموذج هي:

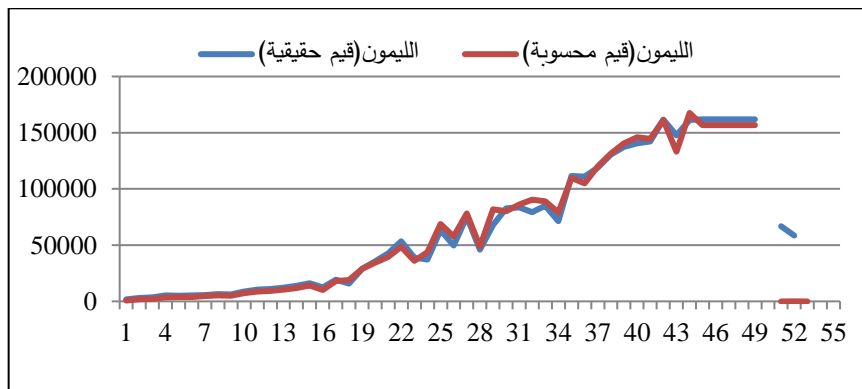
$$\Pr(y1)=1-\frac{\sum_{i=1}^n \left| \frac{y_{11}-y_1}{y_1} \right|}{n} =85.3 \%$$

$$\Pr(y2)=1-\frac{\sum_{i=1}^n \left| \frac{y_{22}-y_2}{y_2} \right|}{n} =88.3 \%$$

ويبين الشكلين (5 ، 6) القيم الحقيقية لإنتاج البرتقال والليمون والقيم المحسوبة بالنموذج الرياضي.



الشكل 5. القيم الحقيقية والمحسوبة لإنتاج البرتقال في سورية



الشكل 6. القيم الحقيقية والمحسوبة لإنتاج الليمون في سورية

الشكلان السابقان يبينان جودة النموذج ونلاحظ كيفية ترافق القيم الحقيقية والمحسوبة وانطباقها شبه التام مع بعضها،

في حين كان أفضل نموذج يمثل سلسلة البرتقال والليمون هو النموذج الخطي الذي لم تتجاوز الدقة لكليهما 40%، ويأخذ الشكل التالي:

$$Y1=-0.0000139 + 17890 t$$

$$R\text{-square}=0.93$$

$$Y2=-0.000316+3938t$$

$$R\text{-square}=0.93$$

ثانياً: مساحة البرتقال والليمون

إذا رمزنا لإنتاج البرتقال بالرمز x_1 وإنتاج الليمون بالرمز x_2 ، بتطبيق الـ PCA على هاتين السلسلتين من خلال مصفوفة الارتباط بين مساحة الليمون ومساحة البرتقال و بإتباع نفس الأسلوب السابق في التحليل كان لدينا مكونان أساسيتان ومواصفاتهما ، جدول (7).

الجدول 7. مواصفات المكونان الأول والثاني لسلاسل المساحة.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.993	99.635	99.635	1.993	99.635	99.635
2	.007	.365	100.000	.007	.365	100.000

المصدر: مخرجات برنامج SPSS.

يتضح أن المكون الأول يفسر 99.635% من تباين مساحة الليمون والبرتقال، بينما يفسر المكون الثاني فقط 0.365% وحسب مقياس قيصر وقياس نسبة التباين نأخذ المكون الأول ونهمل الثاني مع فقدان صغير جداً من صفات السلسلتين.

ويوضح الجدول (8) الأشعة الذاتية التي تربط المكونات بالسلسلتين x_1 و x_2 :

الجدول 8. الأشعة الذاتية للمكونات.

	Component	
	1	2
برتقال	.998	-.06
ليمون	.998	.06

المصدر: مخرجات برنامج SPSS.

حيث:

$$Pc1=0.998 x_{11} + 0.998 x_{22}$$

$$Pc2=-0.082 x_{11} + 0.06 x_{22}$$

x_{11} , x_{22} هي السلاسل المعيارية من السلاسل x_1, x_2

ترتبط السلسلتان x_{11}, x_{22} بالمكون الأول بالعلاقات:

$$x_{11}=0.52*pc1$$

$$x_{22}=0.52*pc1$$

ومنه فإن السلاسل الاصلية:

$$x_1=x_{11} * s(x_1) + m(x_1)$$

$$x_2=x_{22} * s(x_2) + m(x_2)$$

$m(x_1)$, $m(x_2)$ -متوسط السلسلتين x_1 & x_2 على الترتيب.

$s(x_1)$, $s(x_2)$ الانحراف المعياري للسلسلتين x_1 & x_2 على الترتيب.

وبناء نموذج ARIMA على الأول تبين لنا أن أفضل نموذج يصف هذه السلسلة هو النموذج ARIMA

(1,1,0) ، ويوضح الجدول (9) مقدرات النموذج، ومنه يتضح معنوية المعاملات، مع $R\text{-square}=0.99$.

الجدول 9. معاملات نموذج ARIMA (1,1,0)

		Estimate	SE	t	Sig.
No Transformation	Constant	.059	.013	4.620	.000
	AR				
	Lag 1	.509	.129	3.954	.000
	Difference	1			

المصدر: مخرجات برنامج SPSS.

وباستخدام النموذج للتنبؤ لثلاث سنوات قادمة (2019، 2020، 2021) كما هو موضح بالجدول (10).

الجدول 10. نتائج التنبؤ.

العام	Pc1	x1	x2
2019	1.635465	26383.743	7328.134
2020	1.679012	26753.793	7430.389
2021	1.730028	27187.313	7550.183

المصدر: مخرجات برنامج SPSS.

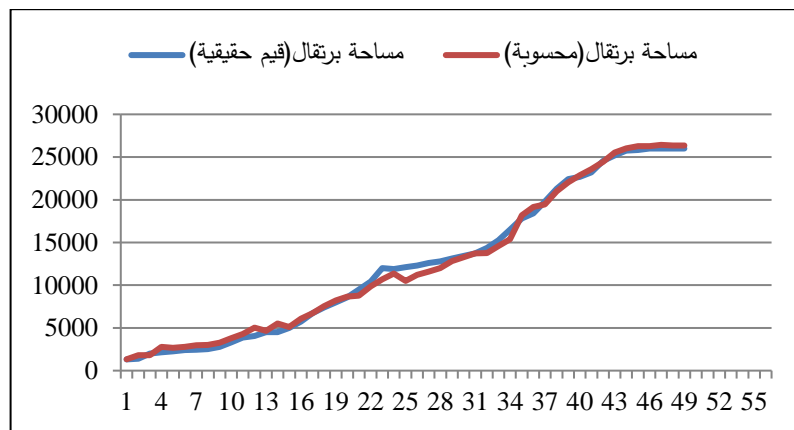
وتم تقدير دقة التنبؤ للنموذج لسلسلتي مساحة البرتقال والليمون كما يلي:

$$\Pr(x_1) = 92.7 \%$$

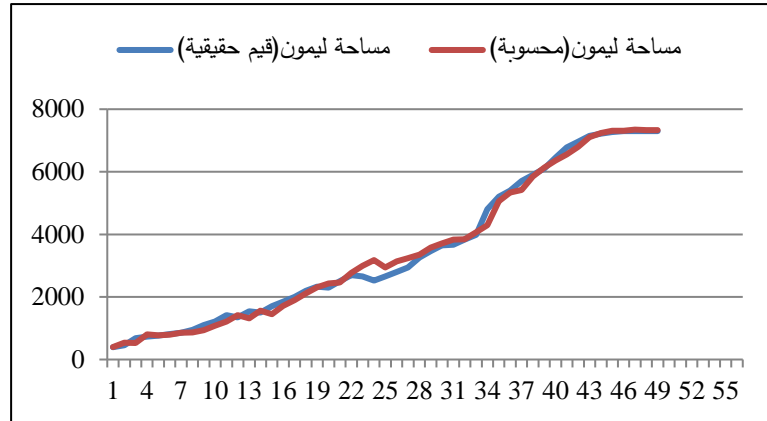
$$\Pr(x_2) = 94.2 \%$$

ويبين كل من الشكلان (7، 8) القيم الحقيقية لمساحة البرتقال والليمون والقيم المحسوبة بالنموذج الرياضي

المقترح.



الشكل 7. القيم الحقيقية والمحسوبة للمساحة المزروعة بالبرتقال.



الشكل 8. القيم الحقيقية والمحسوبة للمساحة المزروعة بالليمون.

أما بالنسبة للنماذج التقليدية فقد وجدنا أن أفضل نموذج يمثل سلسلة مساحة البرتقال هو النموذج التكريري:

$$x1=1552 - 1.086t + 20.809 t^2 - 0.203 t^3$$

$$R\text{-square}=0.98$$

$$pr(x1) = 0.934$$
 وكانت دقة النموذج

أما نموذج سلسلة مساحة الليمون فكان أيضا أفضل نموذج هو النموذج التكريري:

$$X2=685.67 - 9.374 t + 5.086 t^2 - 0.038 t^3$$

$$R\text{-square}=0.983$$

$$pr(x2) = 90.56 \%$$
 وكانت دقة النموذج

وبالمقارنة نجد أن R-square باستخدام سلاسل ARIMA و PCA تعطي نتائج أفضل.

النتائج

- 1- فعالية استخدام التحليل العامل بالأسلوب المكونات الأساسية في تحليل سلاسل إنتاج البرتقال والليمون وكذلك المساحة.
- 2- إمكانية اختصار سلاسل الإنتاج والمساحة إلى سلسلة واحدة.
- 3- إن النموذج الأفضل لتمثيل سلاسل إنتاج الحمضيات هو النموذج ARIMA (5,2,0)
- 4- إن النموذج الأفضل لتمثيل سلاسل مساحة الحمضيات هو النموذج ARIMA (1,1,0)
- 5- فعالية منهجية ARIMA في التنبؤ بإنتاج الحمضيات والاستفادة منها في رسم سياسات الإنتاج والأمن الغذائي.

المراجع:

ابراهيم، نجلاء و هناء عصمان (2013) . تحليل المركبات الأساسية في نظام كشف التطفل الشبكي باستخدام بيانات

NSL-KDD. مجلة الرافدين لعلوم الحاسوب والرياضيات. الناشر: جامعة الموصل. العدد 1. الحجم 10.

عدد الصفحات 251- 261.

بري، عدنان ماجد عبد الرحمن (2002). طرق التنبؤ الإحصائي، جامعة الملك سعود، المملكة العربية السعودية. 344

صفحة.

جماح، ياسر احمد (1985) . تطبيق الاستشعار عن بعد في رصد وتقييم التغيرات بساحل الحديدة- الجمهورية اليمنية "

صنعاء، اليمن.

- عنان، محمد طاهر (2009). استخدام القيم الذاتية والأشعة الذاتية في تنبؤ مؤشرات الانتاج النباتي ، جامعة حلب . حلب. سوريا.
- عنان، محمد طاهر (2014). السلاسل الزمنية، منشورات جامعة حلب، سوريا. 299 صفحة.
- عنبر، رند (2011) . النمذجة الرياضية لتطورات الحالة السكانية في سورية الواقع وآفاق المستقبل جامعة حلب – رسالة ماجستير .
- وزارة الزراعة والإصلاح الزراعي (2018): المجموعة الإحصائية الزراعية. دمشق، سورية.
- وزارة الزراعة والإصلاح الزراعي، المجموعات الإحصائية الزراعية السنوية للأعوام (1970 – 2018).
- وشرن، دين وأخرون (1998) . التحليل الإحصائي للمتغيرات المتعددة من الوجهة التطبيقية. تعريب: عبد المرضي حامد عزام. دار المريخ للنشر. الرياض. المملكة العربية السعودية.
- Mandal, B.N. (2005). Forecasting sugarcane production in India with ARIMA Model. Available from: <http://statjournals.net/ YEAR2005/articles/051002.pdf>.
- Mandala, Q.S. (1977). Econometrics, McGraw- Hill, International book Company, New York 1977. Pp 516.
- Rencher, C. A.(2003). Methods of Multivariate Analysis. Second Edition. ISBN: 978-0-471-46172-2 . Pp738.

Applying ARIMA Models and PCA technique for Forecasting the Production and Area of Citrus Fruits in Syria

Ibtesam Jasem ^{(1)*} and Mohamad Anan ⁽²⁾

(1) General Commission for Scientific Agricultural Research (GCSAR), Damascus, Syria.

(2) Department of Mathematical Statistics, faculty of science, Aleppo University, Aleppo, Syria.

(*Corresponding author: Dr. Ibtesam Jasem E-mail: e_sam_0@hotmail.com).

Received: 16/05/2020

Accepted: 31/05/2020

Abstract

Forecasting economic variables in order to plan and formulate production policies and food security is one of the most important objectives of quantitative economic studies. Multiple methods may be used for the purpose of obtaining economic forecasts. In this study ARIMA models were used that combine the method of Autoregressive and Integrated Moving Averages of time series with PCA technique in order to forecast the production of citrus fruits and its plantation area in Syria. We opted for the ARIMA models as they are known for a high accuracy regarding analysis of time series. In this article, annual data for production and plantation area in Syria for the period) 1970-2018)is being used. Further to statistical tests, ARIMA (5,2,0) turned out to be the best model for forecasting citrus production until 2021. Accuracy values for production were 85.3% in case of oranges and 88.3% in case of lemons. This excelled (substantially) accuracy of conventional models we determined. The best model for forecasting the plantation area was ARIMA (1,1,0). Accuracy was 92.7% in case of orange plantation area and 94.2% in case of lemon plantation area. Here too we determined that ARIMA modeling excelled (substantially) accuracy of conventional models.

Keywords: Orange, Lemon, PCA, Forecasting, Box-Jenkins.