تدقيق السلاسل الزمنية باستخدام برنامج TERROR

محمد علي إسماعيل – يارا عبد المعبود علي– هند سمير طه*

مقدمـــة

تمثل الدقة أحد المكونات الأساسية لجودة البيانات, وتنشأ مشكلة عدم الدقة من مصادر عديدة مثل أخطاء المعاينة وعدم الاستجابة والتغطية و القياس والمعالجة, لذلك هناك حاجة لبحث أسباب عدم الدقة هذه، وإعلان النتائج لمستخدمي البيانات حتى يتخذوا قراراً باستخدام هذه البيانات من عدمه.

يتم تدقيق البيانات عادةً باستخدام طرق يدوية بسيطة مثل مقارنة آخر قيمتين للنمو من فترة لأخرى أو الفروق السنوية والتحقق مما إذا كانت آخر قيمة تمثل قيمة عظمى أو قيمة صغرى بشكل عام،..... الخ، ونجد أن هذه الطرق لها عدة عيوب, أولاً فهي تستخدم جزء من البيانات وليس السلسلة ككل, ثانياً قد تتأثر بشدة بالموسمية وبالأخطاء العشوائية وبالأحداث الخاصة, ثالثاً يحد الفحص اليدوي بشدة من عدد السلاسل التي يمكن فحصها و لذلك يجب أن يتم الفحص على مستوى عال من التجميع, مما يطمس العديد من حقائق السلاسل محل الاعتبار، وعليه تظهر الحاجة إلى أسلوب آلي كفء يأخذ كل المعلومات التي في السلسلة في الاعتبار ويسمح بالفحص على مستوى أكثر تفصيلاً. ويقدم برنامج TERROR طريقة مناسبة لتنفيذ مثل هذا الفحص الآلي، وتهدف هذه الورقة لتطبيق منهجية برنامج TERROR على البيانات الاقتصادية الشهرية المتوفرة على مخزون البيانات القومي بمركز المعلومات ودعم اتخاذ القرار التابع لمجلس الوزراء.

د.محمد على اسماعيل –

د.هند سمير طه

د.يارا عبد المعبود على

رالمتعلقة بجودة البيانات وعلى الدقة كأحد هذه المعايير, ويتم عرض أسباب استخدام برنامج TERROR والمنهجية التي يتبعها ووصف ملخص للبرنامج في المبحث الثالث, ويحتوى المبحث الرابع على وصف للبيانات المستخدمة في تطبيق برنامج TERROR, أيضاً يحتوى على تفسير نتائج تطبيق النتائج على البيانات الاقتصادية الشهرية لمخزون البيانات القومي, وأخيراً يعرض المبحث الخامس ملخص النتائج.

٢. المفاهيم والأبعاد الرئيسية لجودة البيانات

هناك اتفاق عام على أن مفهوم جودة البيانات يصعب قياسه حيث أنه لا يوجد مقياس وحيد مثالي لكل المواقف, وقد قام عدد من المنظمات الدولية باقتراح قائمة من المعايير (الأبعاد) لكي تستخدم في قياس جودة المنتجات الإحصائية، ويوضح جدول رقم (١) الأبعاد المختلفة لجودة البيانات المعترف بها في المنظمات الدولية الآتية: صندوق النقد الدولي ومنظمة التعاون والنمو الاقتصادي واليوروستات ومكتب إحصاءات كندا, وأهم هذه الأبعاد هي:

١. الصلة : درجة تلبية المعلومات الإحصائية للاحتياجات الحقيقية لمستخدميها.

٢. الدقة : تقارب الحسابات أو التقديرات من القيم الفعلية أو الحقيقية التي تهدف الإحصاءات لتقديرها.

٣. الحداثة: سرعة نشر البيانات, أي الفجوة الزمنية بين نهاية الفترة المرجعية (التاريخ المرجعي) وبين نشر البيانات.

٤. إمكانية الوصول إلى المعلومات: السهولة والظروف التي يمكن خلالها الحصول على المعلومات الإحصائية.

٥. الترابط: ملاءمة الإحصاءات بحيث يمكن توفيقها بشكل يعتمد عليه بعدة طرق ولاستخدامات متنوعة.

٦. إمكانية المقارنة/ سلامة المنهجية: مدى إمكانية إرجاع الاختلافات بين الإحصاءات– من أماكن جغرافية مختلفة أو مجالات غير جغرافية أو عبر الزمن– إلى الفروق بين القيم الحقيقية للإحصاءات.

٧. إمكانية التفسير: مدى صحة اللغة والرموز والوحدات ووضوح التعريفات.

جدول رقم (١)

أبعاد جودة البيانات طبقاً لبعض المنظمات الدولية

صندوق النقد الدولي	منظمة التعاون والنمو الاقتصادي	اليوروستات	كندا
مستلزمات الجودة			
	الصلة	الصلة	الصلة
الدقة والاعتمادية	الدقة	الدقة	الدقة
	المصداقية		
النفعية	الحداثة	الحداثة والانتظام	الحداثة
		في نشر البيانات	
إمكانية الوصول إلى المعلومات	إمكانية الوصول إلى المعلومات	إمكانية الوصول إلى المعلومات	إمكانية الوصول إلى المعلومات
		والوضوح	
	تماسك أو ترابط البيانات	تماسك أو ترابط البيانات	تماسك أو ترابط البيانات
سلامة المنهجية		إمكانية المقارنة	
	إمكانية التفسير		إمكانية التفسير
النزاهة والتكامل		الكمال	

تعترف معظم المنظمات الإحصائية بالدقة كأحد معايير جودة البيانات, تركز هذه الورقة على الدقة لأنها تعد بعد ضروري (ولكنه ليس كافي) لجودة البيانات, أيضاً فهي تتناسب عكسياً مع الحداثة, لذلك لابد من التوازن بين الوصول إلى مستوى معين من الدقة وعدم التأثير سلباً على بعد الحداثة بشكل كبير, ومن الجانب الإيجابي فيمكن أن نجزم بالآتي: أولاً تعتبر معرفة الدقة أمر هام جداً للاستخدام الأمثل للبيانات, ثانياً هناك مسئولية على المكاتب الإحصائية لتعليم المستخدمين عن الدقة وتقديم المعلومات اللازمة لتفسير واستخدام قياسات الدقة،ثالثاً لابد أن ينصب اهتمام المكاتب الإحصائية على الفجوات في قدرات قياساتهم خاصةً بالنسبة لخطأ غير المعاينة, تشير الدقة إلى إما تقديم قياسات للدقة مثل متوسط مربع الخطأ أو مؤشرات التقييم الكيفي, لمزيد من التفاصيل, انظر Lee, and Shon (٢٠٠١), إن دقة المعلومات الإحصائية هي درجة توصيفها للظاهرة التي صممت لقياسها بشكل صحيح, وعادة يتم التعبير عنها باستخدام أخطاء التقديرات الإحصائية والتي تنقسم إلى التحيز (الخطأ المنتظم) والتباين

(الخطأ العشوائي), ويمكن أيضاً التعبير عنها من خلال المصادر الرئيسية للخطأ التي تسبب عدم الدقة (مثل التغطية والمعاينة وعدم الاستجابة والاستجابة), لمزيد من التفاصيل, انظر Canadian Institute for Health Information (CIHI) (٢٠٠٥).

٣. برنامج TERROR

يركز البرنامج على قواعد بيانات السلاسل الزمنية, ويعتبر مسألة مراقبة الجودة الآتية: عندما تتوافر بيانات عن الماضي والوقت الحاضر, أي من المشاهدات الجديدة من المرجح أن تحتوي على خطأ؟

ويقوم هذا المبحث ببسط الأسباب وراء استخدام البرنامج ليوضح أهميته, ثم يلقي الضوء على المنهجية التي يتبعها دون الدخول في التفاصيل الفنية, وأخيراً يعطي وصفاً ملخصاً للبرنامج, ويمكن للمهتم بالتفاصيل الفنية سواء عن المنهجية أو عن توصيف البرنامج أن ينظر الكتيب الخاص بالبرنامج, وهو متوفر مجاناً كأحد مطبوعات موقع بنك أسبانيا (www.bde.es).

١٠٣ لماذا برنامج TERROR؟

يمكن أن تحتوي بعض قواعد البيانات على العديد من مئات الآلاف من السلاسل التي تحدث دورياً, وتعتمد مراقبة الجودة للبيانات الجديدة التي تضاف بشدة على الفحص اليدوي باستخدام طرق بسيطة, ولكن هذه الطرق لا تأخذ في اعتبارها المعلومات الكاملة الموجودة في السلسلة ولكن تعتبر فقط عدد صغير من القيم, علاوة على ذلك فيمكن لهذه الطرق أن تتأثر بشدة بالموسمية وبالخطأ العشوائي وبالأحداث الخاصة, كما أن المراقبة اليدوية (الغير آلية) تحد بشدة من عدد السلاسل التي يمكن فحصها و لذلك يجب أن يتم الفحص على مستوى عالٍ من التجميع, وعليه تظهر الحاجة إلى أسلوب آلي كفء يأخذ كل المعلومات التي في السلسلة في الاعتبار ويسمح بالفحص على مستوى أكثر تفصيلاً. طور Gómez and Maravall (١٩٩٢, ١٩٩٣, ١٩٩٤, ١٩٩٦, ٢٠٠١) منهجية (وبرنامج مرتبط بهذه المنهجية) لتحديد نماذج ARIMA آلياً (أو يدوياً), حيث يهدف برنامج TRAMO إلى تقدير والتنبؤ بنماذج الانحدار مع أخطاء الـARIMA المحتملة وبأي تتابع من القيم المفقودة, ويقوم البرنامج بالتعويض عن هذه القيم, ويقوم أيضاً بتحديد وتصحيح عدة أنواع من القيم الشاذة, وأيضاً يقوم بتقدير التأثيرات الخاصة مثل الأعياد

والعطلات الرسمية وأي نوع من متغيرات التدخل بشكل عام, وقد تم اختبار الأداء الآلي لبرنامج TRAMO ("Time Series Regression with ARIMA noise, Missing values, and Outliers") بشكل مكثف, وقد أثبت أنه سريع ويمكن الاعتماد عليه, إن برنامج TERROR هو عبارة عن تطبيق لبرنامج TRAMO ينفذ بأسلوب آلي (ولكن له عدة اختيارات) بهدف مراقبة جودة السلاسل الزمنية, وقد صمم برنامج TERROR ليعالج مجموعات كبيرة من السلاسل الزمنية التي لها تكرارات مشاهدة شهرياً أو ذات تكرارات أقل. ولذلك تستخدم هذه الورقة برنامج TERROR لتدقيق البيانات الاقتصادية الشهرية لمخزون البيانات القومي.

٢٠٣ المنهجية التي يتبعها البرنامج

تقوم المنهجية التي يستخدمها برنامج TERROR للحكم على تقييم المشاهدة الجديدة ما إذا كانت خطأ مشتبه هو أن يتحقق مما إذا كانت بعيدة بشكل كبير عن القيمة المتوقعة لها بالنظر إلى تاريخها الماضي, ويعمل البرنامج بشكل تلقائي من خلال الخطوات الآتية:

١. يقوم بتحديد نموذج ملائم لوصف السلسلة ويقدر معلماته بواسطة دالة الإمكان الأكبر (أو المربعات الصغرى الغير شرطية).

٢. يتحرى عن وجود أنواع متعددة من القيم الشاذة والتأثيرات الخاصة ويقوم بتصحيحها.

٣. ينتج أفضل القيم للتعويض عن المشاهدات المفقودة مع متوسط مربع الخطأ الخاص بها.

٤. يحسب أفضل التنبؤات لكل سلسلة من السلاسل مع متوسط مربع الخطأ الخاص بتلك التنبؤات.

يتم تحديد ما إذا كانت المشاهدة الجديدة خطأ محتمل عندما يكون القيمة المطلقة لخطأ التنبؤ لها أكبر من حد معين معروف مسبقاً, ويتوافر أيضاً ملخص مجمع لنتائج مجموعة السلاسل ككل, وفيما يلي سوف نعرض توصيف للمنهجية بشكل أكثر تفصيلاً.

بفرض أن z_t تصف أحد السلاسل الزمنية، وأن متجه المشاهدات $z = (z_{t1},\ldots,z_{tm})'$ حيث $0 < t_1 < \ldots < t_M$ ، يقوم البرنامج بتقدير نموذج الانحدار الآتي: $z_t = y_t'\beta + x_t$

حيث $\beta=(\beta_1,\ldots,\beta_n)'$ هو متجه معاملات نموذج الانحدار، $y_t=(y_{1t},\ldots,y_{nt})'$ تمثل عدد n من سلسلة المتغيرات، و x_t يتبع نموذج ARIMA كما يلي:

$$\phi(B)\delta(B)x_t=\theta(B)a_t \quad (1)$$

حيث B تمثل معامل الإزاحة للخلف و $\phi(B)$، $\delta(B)$، $\theta(B)$ دوال كثيرات الحدود محدودة في B، و يفترض أن a_t متغيرات مستقلة تتبع نفس التوزيع الطبيعي بمتوسط صفر وتباين σ^2.

بالنسبة للقيم الشاذة التي يتم اكتشافها بواسطة البرنامج تلقائياً يتم معاملتها أيضاً كمتغيرات في معادلة الانحدار، وبالتالي فإذا كان هناك عدد J قيمة شاذة تحدث عند النقاط $T_1,\ldots,T_J$ فإن الحد الذي يعبر عن القيم الشاذة في معادلة الانحدار يمكن التعبير عنه كما يلي:

$$y_t'\omega=\sum_{i=1}^{J}\omega_i\lambda_i(B)I_t^{(T_i)}$$

حيث $I_t^{(T_i)}$ متغير صوري يساوي الواحد الصحيح عندما $t=T_i$، ويساوى الصفر فيما عدا ذلك، ويعبر $\lambda_i(B)$ عن القيم الشاذة.

يتم تقدير معالم نموذج الانحدار الذي يشتمل على متغيرات التقويم والقيم الشاذة والمشاهدات المفقودة من بين القيم الإبتدائية للسلسلة ومعالم نموذج ARIMA بطريقة الإمكان الأكبر (الطريقة الافتراضية للبرنامج)، حيث يتوفر العديد من الخوارزميات لحساب النهاية العظمى لدالة الإمكان (أي إيجاد القيمة الصغرى لمجموع المربعات غير الخطى). ويتم تعظيم دالة الإمكان غير الخطية وحساب الأخطاء المعيارية لقدرات المعالم باستخدام طريقة Marquardt والمشتقة العددية الأولى.

أولاً: تقدير معالم نموذج الانحدار:

يتم تقدير معالم نموذج الانحدار باستخدام تجزئة Cholésky لمقلوب مصفوفة تباينات وتغايرات الأخطاء لتحويل معادلة الانحدار، ويمدنا مرشح Kalman بطريقة كف، لحساب المتغيرات في معادلة الانحدار المحولة. بعد ذلك يتم حل مشكلة المربعات الصغرى الناتجة بواسطة أسلوب خوارزمية تدنية الدالة كثيرة الحدود من الدرجة الثانية.

ينتج عن هذا الإجراء طريقة ذات كفاءة عالية ومستقرة عددياً لحساب مقدرات المربعات الصغرى العامة لمعالم نموذج الانحدار. أما بالنسبة لعملية التنبؤ فيتم تطبيق مرشح Kalman على السلاسل الأصلية.

ثانياً: استكمال المشاهدات المفقودة:

يتم التعامل مع المشاهدات المفقودة بطريقتين متكافئتين: تعتبر الطريقة الأولى هي امتداد للنماذج غير الساكنة لمنهج الاتجاه الوثبي لجونز (١٩٨٠)، وفي هذه الحالة يتم استكمال القيم المفقودة باستخدام طريقة تسليس النقطة الثابتة المبسطة. تتضمن الطريقة الثانية تعيين قيمة اعتباطية وتخصيص قيمة شاذة جمعية لكل مشاهدة مفقودة، على أن تكون القيمة المستكملة هي الفرق بين القيمة التجريبية ومقدر معلمة معادلة الانحدار، و الذي تتطابق قيمته مع القيمة المستكملة بواسطة منهج وثبي.

ثالثاً: اكتشاف ومعالجة القيم الشاذة:

البرنامج مزود بخاصية اكتشاف القيم الشاذة وإزالة أثرها، ويمكن إدخال القيم الشاذة بواسطة المستخدم أو أنها يمكن اكتشافها بواسطة البرنامج أوتوماتيكياً باستخدام منهج Tsay (١٩٨٦) و Chen and Liu (١٩٩٣) حيث يتم اكتشاف القيم الشاذة واحدة تلو الأخرى وفي النهاية يتم استخدام نموذج انحدار متعدد لاكتشاف القيم الشاذة الزائفة. يشبه الإجراء المستخدم لقبول أو رفض القيم الشاذة الإجراء المتبع في نماذج الانحدار التدريجية لاختيار أفضل معادلة انحدار.

ولتلخيص ما سبق، يبدأ البرنامج بتقدير القيم المبدئية لمعلمات كل من الانحدار ونموذج الـARIMA باستخدام انحدارين، ثم يقوم بتقدير المعلمات وبواقي الانحدار الجديدة (النهائية) باستخدام خوارزم تدنية الدالة كثيرة الحدود من الدرجة الثانية مرشح Kalman، ويتم حساب إحصاء (t) لكل مشاهدة لاختبار عدة أنواع من القيم الشاذة، إذا كانت القيم المطلقة لإحصاء (t) لمشاهدة معينة أكبر من قيمة حرجة محددة مسبقاً، يتم تصنيفها على أنها قيمة شاذة، بخلاف ذلك فإن السلسة تخلو من القيم الشاذة.

إذا تم اكتشاف بعض القيم الشاذة في السلسلة، يتم تصحيح هذه القيم، ثم يتم إعادة تقدير معلمات نموذج الـARMA، وفي النهاية يتم تكوين معادلة انحدار متعدد باستخدام مرشح Kalman

وخوارزم تدنية الدالة كثيرة الحدود من الدرجة الثانية. إذا كانت القيمة المطلقة لإحصاء (t̂) أصغر من القيمة الحرجة لبعض المشاهدات, يتم حذف المشاهدة التي لها أصغر قيمة مطلقة من معادلة الانحدار, ثم يتم إعادة تقدير المعادلة, وبعد ذلك يتم حساب قيم إحصاء (t̂) لمختلف أنواع القيم الشاذة ولكل المشاهدات باستخدام البواقي المحسوبة من معادلة الانحدار الأخيرة, وإذا كان مازال هناك قيم شاذة في هذه الخطوة, يقوم البرنامج بتكرار الخطوات السابقة لتصحيحها ثم إعادة تقدير معادلة الانحدار, وهكذا, حتى يتم تقدير معادلة لا تحتوي على قيم شاذة. وتتميز هذه الخوارزمية باعتماد جميع حساباتها على أساليب تحليل الانحدار الخطي، مما يقلل من الوقت المتطلب لإجراء الحسابات.

يتضمن البرنامج أيضاً إمكانية اختبار ما إذا كانت السلسلة تحتاج إلى تحويلة اللوغاريتم, وذلك عن طريق مقارنة معيار المواصفات البايزي لكل من النموذجين (قبل وبعد التحويل), ويتضمن أيضاً إمكانية اختبار وجود بعض آثار التقويم مثل أيام المتاجرة, ويقوم البرنامج كذلك بتحديد نموذج الـARIMA بشكل تلقائي, ويتم ذلك على خطوتين, حيث يتم أولاً تقدير الحدود الغير ساكنة للنموذج (1) ، والتي تأخذ الشكل الضربي في حالة البيانات الموسمية,على أن يتم الحصول على الفروق غير الموسمية بحد أقصى إلى الرتبة الثانية والموسمية بحد أقصى إلى الرتبة الأولى, بينما في الخطوة الثانية يتم تحديد نموذج الـ ARMA للسلسلة الساكنة المصححة بالنسبة للقيم الشاذة.

في النهاية, يقوم البرنامج بتوفيق بعض الإمكانيات التلقائية المتوفرة به لاكتشاف وتصحيح القيم الشاذة, وتحديد نموذج الـARIMA للسلاسل الغير ساكنة, وفي وجود القيم الشاذة والمفقودة, واحتمال إضافة بعض الآثار الأخرى مثل آثار التقويم.

٣٠٣ توصيف البرنامج

أكبر عدد للمشاهدات هو ٦٠٠ بينما أصغر عدد يعتمد على دورية البيانات (بشكل خاص يكون ١٦ للبيانات الربع سنوية بينما يكون ٣٦ للبيانات الشهرية), إن النموذج التلقائي للبرنامج هو نموذج AIRLINE الذي يرجع إلى بوكس وجينكينز (١٩٧٠), وهو كثيراً ما يلائم العديد من السلاسل حيث أن له العديد من الخصائص الملائمة خاصةً أنه يشمل العديد من النماذج الأخرى منها النماذج التي لها اتجاه

عام أو موسمية شبه حتمية, كما أن البرنامج يعتمد على هذا النموذج للسلاسل القصيرة التي لا يمكن الاعتماد على التحديد الآلي للنموذج لها.

وفيما يلي سيتم تسليط الضوء على أهم خصائص البرنامج, حيث أنه فيما يلي سيتم شرح مختصر لمعلمات البرنامج وللملف المدخل والملف المخرج.

أولاً: المعلمات المدخلة

هناك أنواع عديدة من المعلمات المدخلة:

المعلمات الثابتة: تحدد الخطوات التي يقوم البرنامج بتنفيذها تلقائياً ولا يمكن للمستخدم تغييرها, عندما يتم اختيار برنامج TERROR يتم تنفيذ إجراءات خاصة من برنامج TRAMO أغلبها يكون على أساس التحديد التلقائي للنموذج وأيضاً يتم تحديد خطوات التحري عن القيم الشاذة وإجراءات تصحيحها.

المعلمات المدخلة الجديدة: تحدد الخطوات التي ينفذها البرنامج ولكن يتحكم المستخدم فيها من خلال تغيير القيمة التلقائية, تلك المعلمات تؤثر بشكل أساسي على حساسية البرنامج تجاه التحري عن أخطاء التنبؤ المحتملة والمرجحة, أيضاً هناك معلمة بهدف تجنب المغالاة عند التحري عن مثل هذه الأخطاء.

ملخص النتائج للسلاسل الفردية: يمكن التحكم في نتائج السلاسل الفردية من خلال المعلمة NMATRIX, حيث أنها عندما تأخذ القيمة ١ يتم عرض نتائج السلاسل الفردية, لكن عندما تأخذ القيمة صفر لا يتم عرض نتائج السلاسل الفردية.

ملخص النتائج المجمعة لمجموعة السلاسل: يمكن الحصول على نتائج مجمعة من خلال المعلمة MODELSUMM, حيث أنها عندما تأخذ القيمة ١ يتم تقديم ملخص مجمع لمجموعة السلاسل, ولكن عندما تأخذ القيمة صفر لا يتم عرض ملخص مجمع لمجموعة السلاسل.

ثانياً: الملف المدخل

يتوافر ماكرو إكسيل بعنوان (Xlsts.xls) يقوم بتحويل ملف الإكسيل في أي شكل عملي إلى الشكل الملائم للملف المدخل لبرنامج TERROR سواء في شكل ملف إكسيل آخر أو في شكل ملف نصي.

ثالثاً: ملف النتائج الأساسي

توجد نتائج فحص ومراقبة جودة بيانات النتائج بعد تنفيذ TERROR في ملف بعنوان (list.out) في المساحة المحجوزة على القرص المغناطيسي في الحاسوب لتخزين معلومات لوصف ملفات TRAMO\OUTPUT, الملف يحتوي أولاً على المعلمات المستخدمة في التطبيق, ثم على السلاسل التي تم الاشتباه بها, كل سطر يحتوي على اسم السلسلة والتاريخ الذي تم عنده اختبار السلسلة والقيمة الجديدة الفعلية والقيمة المتوقعة لها وخطأ التنبؤ وانحرافه المعياري والقيمة المحسوبة لإحصاء (t) وتصنيف المشاهدة الجديدة على أنها خطأ مرجح أو خطأ محتمل.

٤. نتائج تدقيق البيانات الاقتصادية الشهرية

إن أحد المهام الرئيسية للإدارة العامة لجودة البيانات التابعة لمركز المعلومات ودعم اتخاذ القرار هي تدقيق أهم المؤشرات و الإحصاءات القومية, ويعتبر برنامج TERROR اختياراً مناسباً حيث أنه- كما هو مذكور سابقاً- أنه يسمح بتدقيق مجموعات كبيرة من السلاسل الزمنية بشكل آلي.

يبدأ هذا المبحث بتوصيف البيانات الاقتصادية الشهرية لمخزون البيانات القومي المستخدمة في تطبيق برنامج TERROR، ثم يعرض ملخص لأهم النتائج.

٤-١ توصيف البيانات

تنقسم البيانات المتوفرة على المخزون القومي للبيانات إلى ثلاثة قطاعات رئيسية: البيانات الاقتصادية والبيانات الاجتماعية والبيانات القطاعية, تم أخذ البيانات الاقتصادية الشهرية، وهي تنقسم إلى القطاعات الفرعية الآتية: الأرقام القياسية والاقتصاد العيني والدين العام والقطاع النقدي والمعاملات الخارجية وسوق المال.

يحتوي قطاع الأرقام القياسية على الأرقام القياسية الاقتصادية مثل الرقم القياسي لأسعار المستهلك ومعدل التضخم السنوي، بينما يحتوي قطاع الاقتصاد العيني على بيانات أحكام الإفلاس والاستثمار، كما أنه يحتوي على بعض المؤشرات الاقتصادية، ويحتوي قطاع الدين العام على بيانات تخص كلاً من الدين العام الداخلي والخارجي، بينما يحتوي القطاع النقدي على كل البيانات المتعلقة بالنقود والبنوك، ويحتوي قطاع المعاملات الخارجية على البيانات المتعلقة بالتجارة الخارجية، وأخيراً يحتوي قطاع سوق المال على كل البيانات المتاحة عن البورصة المصرية.

ويوضح جدول رقم (٢) توزيع السلاسل الشهرية للبيانات الاقتصادية المتوفرة على المخزون القومي للبيانات، ويتضح من خلاله أن إجمالي عدد السلاسل للبيانات الاقتصادية الشهرية يبلغ ١٨٥ سلسلة، كما يتبيَّن تركُّز البيانات الاقتصادية في قطاعين فرعيين هما القطاع النقدي (٧٠ سلسلة) وقطاع سوق المال (٦٢ سلسلة). كذلك يوضح الجدول أيضاً عدد السلاسل التي تم تطبيق البرنامج عليها ونسبة السلاسل التي تم استبعادها من البرنامج سواء بسبب قصرها، أو لاحتوائها على عدد كبير من القيم المفقودة، أو نتيجة للثبات النسبي لقيم السلسلة، ويمكن ملاحظة أن نسبة السلاسل التي تم استبعادها تبلغ حوالي ٩٪ من إجمالي البيانات الاقتصادية الشهرية. ووفقاً للقطاعات الفرعية، نجد أن هذه النسبة تكون أعلى فيما يمكن لكل من قطاعي الدين العام والمعاملات الخارجية، حيث تم استبعاد ٢٩٪ من إجمالي السلاسل المتاحة في كل من القطاعين، بينما تكون أقل ما يمكن في قطاع الأرقام القياسية والذي لم يتم استبعاد أي من السلاسل التي تندرج تحت هذا القطاع.

جدول رقم(٢)

السلاسل الشهرية للبيانات الاقتصادية

القطاع	الأرقام القياسية	الاقتصاد العيني	الدين العام	القطاع النقدي	المعاملات الخارجية	سوق المال	إجمالي البيانات الاقتصادية
العدد الكلي للسلاسل	٢٠	٢١	٧	٧٠	٤٣	٦٢	١٨٥

القطاع	الأرقام القياسية	الاقتصاد العيني	الدين العام	القطاع النقدي	المعاملات الخارجية	سوق المال	إجمالي البيانات الاقتصادية
الشهرية							
عدد السلاسل التي طبق عليها البرنامج	٥	٢٧	٥	٦٧	١٠	٥٨	١٦٨
نسبة السلاسل الشهرية المستبعدة	٪٠	٪١٣	٪٢٩	٪٤	٪٢٩	٪٦	٪٩

٢٠٤ أهم النتائج

فيما يلي سيتم عرض أهم نتائج البرنامج من خلال عرض جدولي تحليلي, يعطي الجدول الأول بعض الإحصاءات التلخيصية عن القيم الشاذة والمفقودة, ويتم توضيح بعض خصائص النماذج الموفقة من خلال الجدول الثاني, بينما يتم إعطاء بعض الإحصاءات التلخيصية عن طول السلسلة الزمنية وعدد معلمات نموذج الـARMA من خلال الجدول الثالث, ويوضح الجدول الرابع نتائج اختبارات جودة النماذج الموفقة, وأخيراً يعطي الجدول الخامس نتائج اختبارات التنبؤ لما بعد العينة.

يوضح جدول رقم (٣) نسبة السلاسل الزمنية التي تمّ تطبيق البرنامج عليها ووُجد أنها تحتوي على قيم شاذة أو مفقودة، كما يوضح المتوسط والقيمة العظمى والقيمة الصغرى لعدد القيم الشاذة أو المفقودة للسلسلة الواحدة، مع تقسيم القيم الشاذة حسب نوعها ووفقاً القطاعات الفرعية للبيانات الاقتصادية. ويتضح من النتائج أن نسبة السلاسل الزمنية التي تحتوي على قيم مفقودة للبيانات الاقتصادية الشهرية هي ١٠,١٪, حيث لا توجد أي سلاسل تحتوي على قيم مفقودة في كلٍ من قطاع الأرقام القياسية وقطاع الدين العام، بينما يحتوي حوالي ٢٢,٢٪ من السلاسل التي تنتمي إلى الاقتصاد العيني على قيم مفقودة. أما بالنسبة لمتوسط عدد القيم المفقودة في السلسلة الواحدة للبيانات الاقتصادية الشهرية ككل فيبلغ حوالي

٣,٥، وبالنظر إلى القطاعات الفرعية يمكن ملاحظة أن أعلى قيمة لمتوسط عدد القيم المفقودة في السلسلة الواحدة توجد في قطاع الاقتصاد العيني (٦,٣ قيمة مفقودة لكل سلسلة). بالنسبة للقيم الشاذة, فإنه يتضح أن نسبة السلاسل التي تحتوي على قيم شاذة بالبيانات الاقتصادية الشهرية ككل هي ٨٣,٩٪ بمتوسط ٤,٥ قيمة شاذة لكل سلسلة، وبالنظر إلى توزيعها وفقاً للقطاعات الفرعية يمكن ملاحظة أن جميع السلاسل التي تنتمي لقطاع الأرقام القياسية أو لقطاع المعاملات الخارجية وحوالي ٩٥,٥٪ من سلاسل القطاع النقدي، و٨٠٪ من سلاسل الدين العام، و٧٤,١٪ من سلاسل الاقتصاد العيني، و٧٢,٤٪ من سلاسل سوق المال تحتوي على الأقل على نوع واحد من القيم الشاذة، مما يعكس ارتفاع نسبة القيم الشاذة في كل القطاعات. وبالنظر إلى توزيع القيم الشاذة وفقاً لنوعها، نجد أن ٦٣,٧٪ من السلاسل تحتوي على قيم شاذة جمعية بمتوسط ١,٨ قيمة شاذة جمعية لكل سلسلة, بينما تحتوي ٦١,٩٪ من السلاسل على قيم شاذة تعبِّر عن تحوُّلات في مستوى السلسلة بمتوسط ١,٦ قيمة لكل سلسلة،وأخيراً تحتوي ٥٣٪ من السلاسل تحتوي على قيم شاذة تعبِّر عن تغيُّرات عابرة بمتوسط يبلغ قيمة واحدة لكل سلسلة.

جدول رقم (٣)

القيم المفقودة والشاذة بالسلاسل الشهرية بالبيانات الاقتصادية

القطاع	الأرقام القياسية	الاقتصاد العيني	الدين العام	القطاع النقدي	المعاملات الخارجية	سوق المال	إجمالي البيانات الاقتصادية
أولاً- القيم المفقودة							
نسبة السلاسل التي تحتوي عليها	٪٠	٪٢٢.٢	٪٠	٪٧.٥	٪١٠	٪١٢.١	٪١٠.١
متوسط عدد القيم في السلسلة	٠	٦.٣	٠	١	٧	٣.٤	٣.٥
القيمة العظمى لعدد القيم في السلسلة	٠	١٢	٠	١	٧	١٢	١٢
القيمة الصغرى لعدد القيم في	٠	٠	٠	٠	٠	٠	٠

القطاع	الأرقام القياسية	الاقتصاد العيني	الدين العام	القطاع النقدي	المعاملات الخارجية	سوق المال	إجمالي البيانات الاقتصادية
السلسلة							
ثانياً- القيم الشاذة الجمعية							
نسبة السلاسل التي تحتوي عليها	٪١٠٠	٪٥٩.٣	٪٦٠	٪٦٥.٧	٪٩٠	٥٨.٦	٪٦٣.٧
متوسط عدد القيم في السلسلة	٤	١.٧	١.٨	١.٦	٢.٧	١.٩	١.٨
القيمة العظمى لعدد القيم في السلسلة	١١	٩	٤	٦	١٠	١٢	١٢
القيمة الصغرى لعدد القيم في السلسلة	١	٠	٠	٠	٠	٠	٠
ثالثاً- القيم الشاذة التي تعبّر عن التغيُرات العابرة							
نسبة السلاسل التي تحتوي عليها	٪٦٠	٪٣٣.٣	٪٦٠	٪٦٤.٢	٪٦٠	٪٤٦.٦	٪٥٣.٠
متوسط عدد القيم في السلسلة	١.٢	٠.٨	١	١.٢	١.٤	٠.٩	١.٠
القيمة العظمى لعدد القيم في السلسلة	٢	٥	٢	٥	٥	٤	٥
القيمة الصغرى لعدد القيم في السلسلة	٠	٠	٠	٠	٠	٠	٠
رابعاً- القيم الشاذة التي تعبّر عن تحوُل المستوى							
نسبة السلاسل التي تحتوي	٪٨٠	٪٤٨.١	٪٦٠	٪٨٢.١	٪٤٠	٪٤٨.٣	٪٦١.٩

القطاع	الأرقام القياسية	الاقتصاد العيني	الدين العام	القطاع النقدي	المعاملات الخارجية	سوق المال	إجمالي البيانات الاقتصادية
عليها							
متوسط عدد القيم في السلسلة	٢.٢	١.١	١.٨	٢.٦	٠.٨	٠.٩	١.٦
القيمة العظمى لعدد القيم في السلسلة	٥	٥	٤	٨	٤	٧	٨
القيمة الصغرى لعدد القيم في السلسلة	٠	٠	٠	٠	٠	٠	٠
خامساً- إجمالي القيم الشاذة							
نسبة السلاسل التي تحتوي عليها	٪١٠٠	٪٧٤.١	٪٨٠	٪٩٥.٥	٪١٠٠	٪٧٢.٤	٪٨٣.٩
متوسط عدد القيم في السلسلة	٧.٤	٣.٦	٤.٦	٥.٤	٤.٩	٣.٨	٤.٥
القيمة العظمى لعدد القيم في السلسلة	١٣	١٧	١٠	١٥	١٦	١٧	١٧
القيمة الصغرى لعدد القيم في السلسلة	٥	٠	٠	٠	١	٠	٠

وكما اتضح في السابق من خلال جدول رقم (٣), فإن البيانات الاقتصادية الشهرية تتسم بارتفاع نسبة

القيم الشاذة بها, ويرجع ذلك إلى عدة عوامل أهمها: بالنسبة لقطاع الأرقام القياسية, يرجع السبب في ذلك إلى تغيير سنة الأساس في حساب السلسلة الشهرية الخاصة بالرقم القياسي لأسعار الجملة إلى سنة ١٩٩٩/ ٢٠٠٠ بدلاً من سنة ١٩٨٦/ ١٩٨٧ ابتداءً من شهر سبتمبر لسنة ٢٠٠٤.

١. بالنسبة لقطاع الاقتصاد العيني, يمكن ملاحظة ارتفاع نسبة القيم المفقودة في هذا القطاع مما قد يؤثر سلباً على السلاسل التي تنتمي إليه, كذلك فإنه قد تمّ تغيير سنة الأساس لبعض المتغيرات أكثر من مرة.

٢. بالنظر إلى القطاع النقدي فإنه اعتباراً من سبتمبر ٢٠٠٥ تمّ تعديل حساب متوسطات العوائد باستخدام المتوسط المرجح لعينة من البنوك تمثّل نحو ٨٠٪ من النظام المصرفي بدلاً من المتوسط البسيط لغالبية البنوك، وذلك على أساس شهري بدلاً من أسبوعي، علاوة على ذلك فإنه قد حدث تراجُع في بعض المتغيرات، ويعزى هذا التراجع أساساً إلى سداد وزارة المالية لجانب من مستحقات بنوك القطاع العام التجارية تجاه شركات تجارة وتصدير الأقطان قدرها ٣٠٦٦ مليون جنيه بموجب أذون على الخزانة العامة، أيضاً بسبب سداد مبلغ ٤١١ مليون دولار في كلٍ من يناير ويوليو ٢٠٠٤ من ديون مصر الخارجية والمستحق الجانب الأكبر منها لدول نادي باريس.

يوضح جدول رقم (٤) خصائص النماذج المطلوبة لمعالجة السلاسل الزمنية الشهرية بالبيانات الاقتصادية، ومنه يلاحظ أن ٦٦,٠٧٪ من السلاسل الشهرية للبيانات الاقتصادية تحتاج إلى تحويلة اللوغاريتم لتسكين السلسلة في التباين، وتتركز هذه النسبة في بيانات القطاع النقدي حيث أن ٨٣,٥٨٪ من السلاسل التي تنتمي إليه تحتاج إلى تحويلة اللوغاريتم، كذلك فإن حوالي ٧,١٤٪ فقط من السلاسل تعتبر ساكنة. ووفقاً للقطاعات الفرعية فإن كل السلاسل الزمنية التي تنتمي لكلٍ من قطاع الأرقام القياسية والدين العام والقطاع النقدي تعتبر غير ساكنة, وبالنظر إلى أقل فرق مطلوب لتسكين السلسلة يتضح أن ٥٠٪ من السلاسل احتاجت أخذ كلاً من فرق من الدرجة الأولى والفرق الموسمي $\Delta\Delta_{12}$، وتتركز هذه السلاسل في القطاعات التالية: سوق المال (٧٤٪)، والاقتصاد العيني (٦٢,٩٦٪)، والأرقام القياسية (٦٠٪). يتضح كذلك من النتائج أن حوالي ٣٥,٧١٪ من السلاسل احتاجت إلى أخذ فرق من الدرجة الأولى Δ, وتتركز هذه النسبة في القطاعات التالية: المعاملات الخارجية (٧٠٪)، والدين العام (٦٠٪)، والقطاع النقدي (٥٥,٢٢٪)، والأرقام القياسية (٤٠٪). ومن خلال النتائج يمكن ملاحظة أن أخذ كلاً من الفرق من الدرجة الثانية مع الفرق الموسمي $\Delta^2\Delta_{12}$ وأخذ الفرق من الدرجة الثانية فقط Δ^2 لم ينتج عنها نتائج مرضية في أغلب السلاسل حيث لم يحققا أي نتائج مرضية سوى في القطاع النقدي فقط بنسب ١١,٩٤٪ و٥,٩٧٪

للنوعين على الترتيب، علاوة على ذلك، نجد أن الفرق الموسمي Δ_{12} لم يؤد إلى نتائج مرضية في أي من السلاسل، وأخيراً فإن النموذج التلقائي في البرنامج (AIRLINE) حقق نتائج مرضية في ٤٧,٦٢٪ من السلاسل، وتتركز هذه النسبة في القطاعات التالية: سوق المال (٧٢,٤١٪)، والاقتصاد العيني (٦٢,٩٦٪)، والأرقام القياسية (٦٠٪).

جدول رقم (٤)

خصائص النماذج المطلوبة لمعالجة السلاسل الشهرية بالبيانات الاقتصادية

القطاع	الأرقام القياسية	الاقتصاد العيني	الدين العام	القطاع النقدي	المعاملات الخارجية	سوق المال	إجمالي البيانات الاقتصادية
أولاً- سلوك السكون والفروق							
اللوغاريتم	٢٠٪	٣٧.٠٤٪	٦٠٪	٨٣.٥٨٪	٦٠٪	٦٢.٠٧٪	٦٦.٠٧٪
السلاسل الساكنة	٠٪	١٨.٥٢٪	٠٪	٠٪	١٠٪	١٠.٣٤٪	٧.١٤٪
ثانياً- أقل فرق مطلوب							
Δ	٤٠٪	١٨.٥٢٪	٦٠٪	٥٥.٢٢٪	٧٠٪	١٦٪	٣٥.٧١٪
Δ^2	٠٪	٠٪	٠٪	٥.٩٧٪	٠٪	٠٪	٢.٣٨٪
Δ_{12}	٠٪	٠٪	٠٪	٠٪	٠٪	٠٪	٠٪
$\Delta\Delta_{12}$	٦٠٪	٦٢.٩٦٪	٤٠٪	٢٦.٨٧٪	٢٠٪	٧٤٪	٥٠.٠٪
$\Delta^2\Delta_{12}$	٠٪	٠٪	٠٪	١١.٩٤٪	٠٪	٠٪	٤.٧٦٪

القطاع	الأرقام القياسية	الاقتصاد العيني	الدين العام	القطاع النقدي	المعاملات الخارجية	سوق المال	إجمالي البيانات الاقتصادية
ثالثاً- تحديد النموذج							
نموذج AIRLINE	٪٦٠	٪٦٢.٩٦	٪٤٠	٪٢٢.٣٩	٪٢٠	٪٧٢.٤١	٪٤٧.٦٢

يوضح جدول رقم (٥) بعض الإحصاءات التلخيصية عن السلاسل والنماذج المستخدمة، حيث يتضح أن متوسط طول السلسلة الزمنية للبيانات الاقتصادية الشهرية ١٠١,٢ مشاهدة بانحراف معياري ٣٧,٦, مما يعكس التفاوت الكبير في طول السلسلة الزمنية حيث أن أقصر سلسلة طولها ٣٦ مشاهدة – وهي تنتمي لقطاع الدين العام – بينما أطول سلسلة تتكون من ٢١٤ مشاهدة وتنتمي لقطاع الأرقام القياسية،س بالنسبة لعدد معلمات نموذج ARMA، توضح النتائج أنه في المتوسط يتكون نموذج ARMA من ١,٨ معلمة بانحراف معياري ٠,٨, كذلك فإن أصغر نموذج يحتوي على معلمة واحدة فقط بينما يحتوي أكبر نموذج على خمس معلمات وهو ينتمي للقطاع النقدي.

جدول رقم (٥)

بعض الإحصاءات التلخيصية للسلاسل الشهرية بالبيانات الاقتصادية

القطاع	الأرقام القياسية	الاقتصاد العيني	الدين العام	القطاع النقدي	المعاملات الخارجية	سوق المال	إجمالي البيانات الاقتصادية
أولاً- بعض الإحصاءات الخاصة بطول السلسلة							
المتوسط	١٦٦	٨٥.٢	٩١.٢	١٢٩	١١٨.٣	٩٦.٩	١٠١.٢
الانحراف المعياري	٥٥.٧	٢٩.٥	٤٥.١	٢٠.٩	١٩.٧	١٦.٧	٣٧.٦
القيمة العظمى	٢١٤	١١٩	١٢٨	١٨٩	١٥٣	١١٩	٢١٤

القطاع	الأرقام القياسية	الاقتصاد العيني	الدين العام	القطاع النقدي	المعاملات الخارجية	سوق المال	إجمالي البيانات الاقتصادية
القيمة الصغرى	٨١	٤٥	٣٦	٨٠	٩٣	٥٩	٣٦
في السلسلةARMAثانياً- بعض الإحصاءات الخاصة بعدد معلمات نموذج							
المتوسط	١.٨	١.٩	٢	١.٨	١.٩	١.٨	١.٨
الانحراف المعياري	٠.٤	٠.٨	١.٣	١	١	٠.٥	٠.٨
القيمة العظمى	٢	٤	٤	٥	٤	٣	٥
القيمة الصغرى	١	١	١	١	١	١	١

ويوضح جدول رقم (٦) نتائج بعض الاختبارات لجودة النماذج المختلفة بدرجة ثقة ٩٩٪، ويتضح منها اختبار Ljung-Box ويشير إلى أن ٩٦,٤٪ من نماذج السلاسل الزمنية لها بواقي غير مرتبطة. وبالنظر إلى نتائج اختبار Bowman Shenton للتحقق من إتباع البواقي للتوزيع المعتاد أو الطبيعي يتضح أن ٨٩,٩٪ من مجموع نماذج السلاسل الزمنية الشهرية للبيانات الاقتصادية لها بواقي تتبع التوزيع المعتاد، كما أن ٩٤,٦٪ منها اجتاز اختبار الالتواء لتوزيع البواقي أيضاً بنجاح, علاوة على ذلك فقد اجتاز حوالي ٩٠,٥٪ من السلاسل اختبار التفرطح لتوزيع البواقي بنجاح. كذلك تشير نتائج اختبار Pierce إلى عدم ارتباط البواقي الخاصة بكل السلاسل الزمنية الشهرية للبيانات الاقتصادية موسمياً، كما يوضح اختبار Ljung-Box لمربعات البواقي أن ٩٢,٩٪ من السلاسل الزمنية لها بواقي مستقلة, وأخيراً توضح نتائج اختبار Runs أن ٩٨,٨٪ من السلاسل الزمنية الشهرية للبيانات الاقتصادية تمتاز البواقي للنماذج الخاصة بها بالعشوائية.

جدول رقم (٦)

نتائج اختبارات جودة النماذج الخاصة بالسلاسل الشهرية بالبيانات الاقتصادية

القطاع	الأرقام القياسية	الاقتصاد العيني	الدين العام	القطاع النقدي	العاملات الخارجية	سوق المال	إجمالي البيانات الاقتصادية
البواقي غير المرتبطة	٪١٠٠	٪٩٦.٣	٪١٠٠	٪١٠٠	٪١٠٠	٪٩٦.٦	٪٩٦.٤
البواقي الطبيعية	٪٤٠	٪٨٥.٢	٪٨٠	٪٩٤	٪١٠٠	٪٩١.٤	٪٨٩.٩
الالتواء	٪٦٠	٪٨٨.٩	٪١٠٠	٪٩٧	٪١٠٠	٪٩٦.٦	٪٩٤.٦
التفرطح	٪٤٠	٪٨٨.٩	٪٨٠	٪٩٤	٪١٠٠	٪٩١.٤	٪٩٠.٥
البواقي غير المرتبطة موسمياً	٪١٠٠	٪١٠٠	٪١٠٠	٪١٠٠	٪١٠٠	٪١٠٠	٪١٠٠
البواقي المستقلة	٪٦٠	٪١٠٠	٪٨٠	٪٩١	٪١٠٠	٪٩٦.٦	٪٩٢.٩
عشوائية علامات البواقي	٪١٠٠	٪٩٦.٣	٪١٠٠	٪٩٨.٥	٪١٠٠	٪١٠٠	٪٩٨.٨

وأخيراً يوضح جدول رقم (٧) نتائج اختبارات التنبؤ خارج إطار العينة، حيث يوضح التوزيع النسبي للسلاسل الزمنية الشهرية الخاصة بالبيانات الاقتصادية التي قام البرنامج بمعالجتها وفقاً لاحتمال وجود خطأ بالقيمة الأخيرة بالسلسلة من خلال التنبؤ بها, وتوضح النتائج أن حوالي ٩١,٦٧٪ من السلاسل الزمنية أعطت نتائج متسقة مع النموذج المستخدم للتنبؤ، وبالتالي فإنها خالية من الأخطاء المحتملة بالنظر إلى القيمة الأخيرة للسلسلة، وتتركز هذه النسبة في القطاع النقدي حيث تم قبول جميع السلاسل المتاحة فيه بدون أخطاء. كذلك توضح النتائج أن حوالي ٢,٣٨٪ من السلاسل الزمنية الشهرية بالبيانات الاقتصادية يشتبه أنها تحتوي على خطأ, وتتركز هذه النسبة في قطاعين هما الأرقام القياسية (٢٠٪) والدين العام (٢٠٪)، أما عن نسبة السلاسل الزمنية الأكثر اشتباهاً بوجود خطأ بالقيمة الأخيرة

لها فإنها تبلغ حوالي ٥,٩٥٪، وتتوزع هذه النسبة ما بين قطاعين اثنين فقط هما سوق المال (١٥,٥٢٪) والاقتصاد العيني (٣,٧٪).

جدول رقم (٧)

التوزيع النسبي للسلاسل الشهرية بالبيانات الاقتصادية وفقاً لاحتمال وجود خطأ بالقيمة الأخيرة لها

القطاع	الأرقام القياسية	الاقتصاد العيني	الدين العام	القطاع النقدي	المعاملات الخارجية	سوق المال	إجمالي البيانات الاقتصادية
السلاسل المقبولة بدون خطأ	٨٠٪	٨٨.٨٩٪	٨٠٪	١٠٠٪	٩٠٪	٨٤.٤٨٪	٩١.٦٧٪
المشتبه بها	٢٠٪	٧.٤١٪	٢٠٪	٠٪	١٠٪	٠٪	٢.٣٨٪
الأكثر اشتباهاً	٠٪	٣.٧٠٪	٠٪	٠٪	٠٪	١٥.٥٢٪	٥.٩٥٪

ويتضح من جدول رقم (٧) أن القطاع الوحيد الذي أعطى نتائج مرضية في التنبؤ بنسبة ١٠٠٪ هو القطاع النقدي, وبالبحث وراء أسباب وجود بعض السلاسل التي يُحتمل وجود خطأ بها بالنسبة للقيمة الأخيرة لها في القطاعات الأخرى، يتبين أن أهم تلك الأسباب تكمن في أن نهاية بعض السلاسل تكون قيم تقديرية وليست حقيقية، ويمكن أن تكون مقدرة بنموذج مختلف عن النموذج الذي قام البرنامج بتوفيقه لهذه السلاسل، مما قد يؤدي إلى فروق واضحة بين القيمة الأصلية الموجودة في السلسلة وبين القيمة المتوقعة.

٥. النتائج والتوصيات

من خلال التطبيق على السلاسل الاقتصادية الشهرية الموجودة بالمخزون باختلاف تصنيفاتها الفرعية اتضح أن أهم أسباب وجود القيم الشاذة و/ أو الحصول على نتائج غير مرضية للتنبؤ بهذه السلاسل هي:

١. تغيير سنة الأساس لبعض المتغيِّرات داخل نفس السلسلة الزمنية: ويمكن تفادي ذلك من خلال توحيد سنة الأساس للسلسلة ككل، أو قسمة السلسلة إلى سلسلتين أو أكثر بحيث يكون لكل واحدة من هذه السلاسل سنة أساس موحدة، أو استبعاد القيم المحسوبة عند سنة الأساس القديمة من السلسلة.

٢. ارتفاع نسبة القيم المفقودة لبعض السلاسل: ويمكن التغلُّب على ذلك من خلال استكمال تلك السلاسل إن أمكن، أو ذكر أسباب الفجوات في تلك السلاسل في الملف الخاص بالملاحظات الملحق مع البيانات، أو حذفها من بيانات المخزون تماماً.

٣. تغيير تعريف أو منهجية حساب بعض المؤشرات: ويمكن التغلُّب على ذلك من خلال قسمة تلك المؤشرات إلى عدة سلاسل تحتوي كل واحدة منها على القيم المحسوبة للمؤشر بنفس المنهجية ووضع تعريف واضح لكل مؤشر من هذه المؤشرات يوضح الفرق بينها.

٤. وجود بيانات تقديرية وليست فعلية للقيم الموجودة في نهايات بعض السلاسل: ويمكن تفادي ذلك باستبدال البيانات التقديرية بالبيانات الفعلية إن أمكن، أو ذكر سبب وضع القيم التقديرية مع توضيح الطريقة المستخدمة في التقدير في الملف الخاص بالملاحظات.

وفي النهاية يمكن القول أنه يمكن بتطبيق البرنامج على بيانات المخزون القومي لكل سلسلة على حدا, ومراجعة القيم الشاذة مع مصادر البيانات, بحيث يتمُّ تصحيحها في حالة كونها تمثِّل أخطاء أو توضيح الأسباب وراء بعدها عن باقي قيم السلسلة في حالة صحتها وذكره في ملف الملاحظات الموجود مع البيانات مساعدة المستخدم على فهم البيانات بشكل أفضل وبصورة أدق. كذلك فإنه من الهام استكمال السلاسل القصيرة إن أمكن أو ذكر سبب قصرها أو حذفها من بيانات المخزون القومي، واستبدال القيم التقديرية بالقيم الفعلية إن أمكن أو حذف القيم التقديرية وفصلها في سلاسل خاصة بها.

المراجع

1. Anderson, B. and J. Moore, (1979), "Optimal Filtering", **New Jersey: Prentice Hall.**
2. BOX, G.E.P. and JENKINS, G.M. (1970), "Time Series Analysis: Forecasting and Control", **San Francisco: Holden-Day.**
3. Brackstone, G. (2001), "How Important Is Accuracy?", **Proceedings of Statistics Canada Symposium.**
4. Brackstone, G. (1999), "Managing Data Quality in a Statistical Agency", **Statistics Canada.**
5. Canadian Institute for Health Information (2005), "The CIHI Data Quality Framework", **Canada.**
6. Caporello, G. and Marvall, A (2004), "A Tool for Quality Control of Time Series Data Program TERROR", **Bank of Spain, Service of Studies.**
7. Census Bureau Methodology & Standards Council (2006), "Definition of Data Quality", **US Census Bureau.**
8. Chen, C. and L.M. Liu (1993), "Joint Estimation of Model Parameters and Outlier Effects in Time Series", **Journal of the American Statistical Association 88, 284-297.**
9. European Central Bank, Annual Quality Report (2006), "Euro Area Balance of Payments and International Investment Position Statistics", **Germany.**
10. Fischer B. and Planas C. (1998), "Large Scale Fitting of ARIMA Models and Stylised Facts of Economic Time Series", **Planistat Luxembourg, Luxembourg and Joint Research Centre, Italy.**
11. Froeschl K. A. and Grossmann W. (2002), "Deciding Statistical Data Quality", **Department of Statistics and Decision Support Systems University of Vienna.**
12. Full, S., Haworth, M. and Stephens A. (2001), "Quality Measurement and Reporting in the UK Office for National Statistics", **Proceedings of Statistics Canada Symposium.**

13. GÓMEZ, V. and MARAVALL, A. (1992), "Time Series Regression with ARIMA Noise and Missing Observations - Program TRAM", EUI Working Paper ECO No. 92/81, Department of Economics, European University Institute.

14. GÓMEZ, V. and MARAVALL, A. (1993), "Initializing the Kalman Filter with Incompletely Specified Initial Conditions", in Chen, G.R. (ed.), Approximate Kalman Filtering (Series on Approximation and Decomposition), London: World Scientific Publ. Co.

15. GÓMEZ, V. and MARAVALL, A. (1994), "Estimation, Prediction and Interpolation for Non-stationary Series with the Kalman Filter", Journal of the American Statistical Association 89, 611-624.

16. GÓMEZ, V. and MARAVALL, A. (1996), "Programs TRAMO (Time series Regression with ARIMA noise, Missing observations, and Outliers) and SEATS (Signal Extraction in ARIMA Time Series), Instructions for the User", Working Paper 9628, Servicio de Estudios, Banco de España.

17. GÓMEZ, V. and MARAVALL, A. (2001), "Automatic Modeling Methods for Univariate Series", Ch.7 in Peña D., Tiao G.C. and Tsay, R.S. (eds.) **A Course in Time Series Analysis, New York: J. Wiley and Sons.**

18. Jones, R. (1980) "Maximum Likelihood Fitting of ARMA Models to Time Series with Missing Observations", **Technometrics 22, 389-395.**

19. Kalton, G. (2001), "How Important Is Accuracy?", **Proceedings of Statistics Canada Symposium.**

20. Kuchler, C. and Martin, S. (2006), "The Data Quality Concept of Accuracy in the Context of Public Use Data Sets", **German Institute for Economic Research.**

21. Lee, D. and Shon A. (2001), "Korea's Experiences in Statistical Quality Assessment", **Proceedings of Statistics Canada Symposium.**

22. **Organization for Economic Co-operation and Development (OECD),** OECD Glossary, http://stats.oecd.org/glossary/detail.asp?ID=21> Retrieved on April 21, 2007, last updated on March 08, 2006, Web page.

23. Pipino L. L, Lee Y. W, and Wang R. Y. (2001), "Data Quality Assessment", Information and Methods of Calculation, **Communications of the ACM, US.**

24. Scheuren, F. (2001), "How Important Is Accuracy ", **Proceedings of Statistics Canada Symposium.**

25. Software and Standards Management Branch, Systems Support Division (1998), "IT Standard 16.0.0: Survey Design and Statistical Methodology Metadata", **US Bureau of the Census.**

26. Tayi, K.G. and Ballou D. P (1998), "Examining Data Quality", **Communications of the ACM, US.**

27. Tsay, R.S. (1986), "Time Series Model Specification in the Presence of Outliers", **Journal of the American Statistical Association 81, 132-141.**

Wald, W. G. and Linden H. (2001), "Quality Measurement-Eurostat Experiences", **Proceedings of Statistics Canada Symposium**