نموذج ARIMA

فصل 10. نموذج ARIMA

مكونات النموذج – مراحل صياغة النموذج واختباره –دراسة حالة – المراجع

توطئة.

نموذج ARIMA هو طريقة مشهورة للتنبؤ من بين مجموعة طرق الموائمة للمنحنى (Curve-fit methods). تصلح هذه الطريقة الموثوقة لقطاع واسع من الحالات العملية؛ وذلك بفضل مرونتها التي تتأتى من تغيير المعالم ومن تعدد النماذج واستيعاب السلاسل غير المستقرة عن طريق التحويل. مثلا، تستوعب الطريقة السلاسل ذات التوجه الخطي وغير الخطي كما تستوعب السلاسل الخالية من الموسمية (ARIMA) وتلك التي تتضمن موسمية باستخدام طريقة مجاورة هي SARIMA. حجم البيانات التي تتطلبها الطريقة يتوقف على طبيعة السلسلة لكن الطريقة ليست من النوع الذي يتطلب بيانات كبيرة، فالحد الأدنى هو من خمسين إلى ستين مشاهدة.

في نهاية هذا الفصل يفترض أن يصبح الطالب قادرا على استخدام الطريقة على برنامج R، مع التحقق من اشتراطاتها، وتقييم النماذج المستخرجة، والمفاضلة بينها، ومن ثم التنبؤ.

1. مكونات النموذج

التفريق

الانحدار الذاتي

المتوسطات المتحركة

نماذج ARIMA(p, d, q) هو اختصار ل AutoRegressive Integrated Moving Average، وهي صنف من النماذج تستخدم ماضي السلسلة والخطأ للتنبؤ. يجمع هذا الصنف من النماذج بين ثلاث مكونات: التفريق؛ والانحدار الذاتي؛ والمتوسطات المتحركة (في الواقع كلمة Integrated هنا تعني الدمج أي عكس عملية التفريق). فيما يلي نستعرض باختصار هذه المكونات.

1-1. التفريق

تفترض النمذجة بطريقة ARIMA أن السلسلة مستقرة، لذلك نحتاج قبل البحث عن نموذج المناسب أن نتحقق من أن السلسلة مستقرة، فبدونها لا تكون النتائج موثوقة لأن النموذج لا يتضمن ما تحتويه السلسلة من أنماط. استقرارية السلسلة تعني أن تكون خصائصها الاحصائية مستقرة خلال الزمن، أو بتعبير أدق: **أن يكون كل من المتوسط، والتباين، وهيكل الارتباطات الذاتية غير مرتبطة بالزمن**[1]. السلسلة الزمنية غير المستقرة هي التي تتضمن توجها أو موسمية أو أي تغير في نمطها. لتحقيق الاستقرارية غالبا ما يكفي تفريقها من الدرجة الأولى.

$$D(y_t) = y_t - y_{t-1}$$

عندما لا يكفي ذلك قد نلجأ إلى التفريق من الدرجة الثانية، أي تفريق التفريق. نسمي النموذج الذي تم فيه اللجوء إلى التفريق ب ARIMA(p, d, q) حيث d درجة التفريق. يمكن أيضا الحصول على الاستقرارية بطرق أخرى مثل تفكيك السلسلة لنزع التوجه والموسمية أو التحويل اللغرتمي أو تحويل بوكس –كوكس، سواء مع أو بدون تفريق.

1-2. نموذج الانحدار الذاتي *AR*

يربط y بماضيها إلى عدد محدد p من الفترات (يسمى درجة AR).

$$AR(p) model: y_t = c + \emptyset_1 y_{t-1} + \cdots + \emptyset_p y_{t-p} + \varepsilon_t$$

حيث P هو عدد الفترات (lags) التي تمتد عبرها العلاقة بين y_t وماضيها، أي عدد الحدود في نموذج الانحدار الذاتي الخطي التي يتعين تقديرها: إذا كانت p=0 فهذا يعني أن y_t غير مرتبطة خطيا بماضيها، إذا كان p=1 فهذا يعني أن y_t مرتبطة ب y_{t-1} فقط، والذي تعبر عنه $\emptyset_1$، إذا كانت p = 2 فهذا يعني أن y_t مرتبطة ب y_{t-1} و y_{t-2}، وهكذا. ε_t هو ضجيج أبيض.

مثلا إذا كان p=2:

$$y_t = c + \emptyset_1 y_{t-1} + \emptyset_p y_{t-2} + \varepsilon_t$$

---

[1] نقول عن**(X_t)** أنها مستقر بالمفهوم الدقيق فقط إذا كان التوزيع الاحتمالي ل $(X_{t+1}, \ldots, X_{t+h})$ يتوقف فقط على h. هذه الصيغة تعد أحيانا متشددة وتفضل عليها صيغة الاستقرارية من الدرجة الثانية وهي: نقول عن سلسلة أنها مستقرة فقط إذا كان، من أجل أي عدد طبيعي h، فإن $E(X_{t+h})$ و $cov(X_t, X_{t+h})$ مرتبطة فقط ب h.

3-1. نموذج المتوسطات المتحركة *MA*

يربط y بماضي خطأ التنبؤ إلى عدد محدد q من الفترات (يسمى درجة MA).

$$MA(q)model: y_t = c + \theta_1\varepsilon_{t-1} + \cdots + \theta_{t-q}\varepsilon_{t-q}$$

q هو عدد مكونات المتوسط المتحرك في نموذج. في حالة q=0 هذا يعني عدم وجود هذا المكون، في حالة q= 1 هذا يعني ارتباط y_t بمكون الخطأ السابق فقط، في حالة q=2 هذا يعني أن y_t مرتبط بحدين للخطأ: السابق والذي قبله، وهكذا.

مثلا إذا كان q=2:

$$y_t = c + \theta_1\varepsilon_{t-1} + \theta_{t-2}\varepsilon_{t-2}$$

القيمة الأولى التقديرية y^ هي المتوسط، والخطأ عندها هو الفرق بين المتوسط والقيمة الحقيقية، وتستمر هذه العملية لإنتاج $(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots)$.

أحيانا تكون y_t تتضمن كلا من مكوني الانحدار الذاتي والمتوسطات المتحركة، في هذه الحالة نستخدم نموذج ARMA الذي يدمج بين النموذجين ويكتب كما يلي:

$$ARMA(p,q)\ model: y_t = c + \emptyset_1 y_{t-1} + \cdots + \emptyset_p y_{t-p} + \theta_1\varepsilon_{t-1} + \cdots \theta_{t-q}\varepsilon_{t-q} + \varepsilon_t$$

مثلا إذا كان p = 2 و q = 1 ، نكتب النموذج ARMA(2,1) كما يلي:

$$y_t = c + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \theta_1\varepsilon_{t-1} + \varepsilon_t$$

إذا كان p = 1 و q = 1 نكتب النموذج ARMA(1,1) كما يلي:

$$y_t = c + \emptyset_1 y_{t-1} + \theta_1\varepsilon_{t-1} + \varepsilon_t$$

2. مراحل صياغة النموذج واختباره

يمكن القول أن صياغة النموذج والتحقق منه تتم عبر ست مراحل.

2-1. الاستقرارية

النمذجة ب ARIMA كما العديد من التقنيات الأخرى تتطلب استقرارية السلسلة. لذلك قبل البحث عن النموذج المناسب لبيانات السلسلة يجب التحقق من أن هذه الأخيرة مستقرة. إذا لم تكن كذلك نحولها إلى مستقرة، وهناك عدة طرق لذلك، أهمها التفريق.

تعريف الاستقرارية

بتعبير بسيط، يمكن أن نقول السلسلة المستقرة هل التي يكون فيها التوقع والتباين للسلسلة، أي $E(X_t)$ et $var(X_t)$ ثابتين خلال الزمن[1]. هذا التعريف يسهل تفحص الاستقرارية من خلال الرسم ابتداء.

تفحص الاستقرارية بالتمثيل البياني

التحقق من الاستقرارية يكون أولا بالرسم البياني للسلسلة. تأتي السلسلة المستقرة في شكل يدل على عدم الارتباط بالزمن لكل من المتوسط والتباين؛ بمعنى شكل تغاير حول خط أفقي نوعا ما بدون انتقال في المتوسط ومع تباين ثابت؛ أي بدون توجه ولا موسمية. الدورة في المقابل لا تعني وجود عدم استقرارية لأن الدورة غير منتظمة، فلا نعلم مثلا متى نصل إلى القمة أو القاع في الدورة (Hyndman & Athanasopoulos, 2018). لتفحص وجود توجه أو موسمية يمكن تفكيك السلسلة إلى مكوناتها.

لشرح أكثر للاستقرارية لا بأس أن نستعير عن الكتاب الالكتروني الممتاز لهيندمان وأثاناسوبولوس (Hyndman & Athanasopoulos, 2018) الأمثلة التالية لسلاسل زمنية مختلفة ونستخرج أيها يمكن القول عنه أن سلسلة مستقرة. التدقيق يظهر أن السلسلتين (b) و (g) هما الوحيدتان اللتان ينطبق عليهما وصف الاستقرار. بقية السلاسل تتضمن إما توجها (e- i) أو موسمية (d-h-i) أو انتقالا للمتوسط (a - c - f) أو تغيرا في التباين (i). نمط الدورة في (g) لا يخرجها عن الاستقرار لأن الدورة غير منتظمة، وهي السلسلة الوحيدة المستقرة بدون تفريق، وأما (b) فهي مثال على نجاح التفريق الأول في الحصول على الاستقرارية.

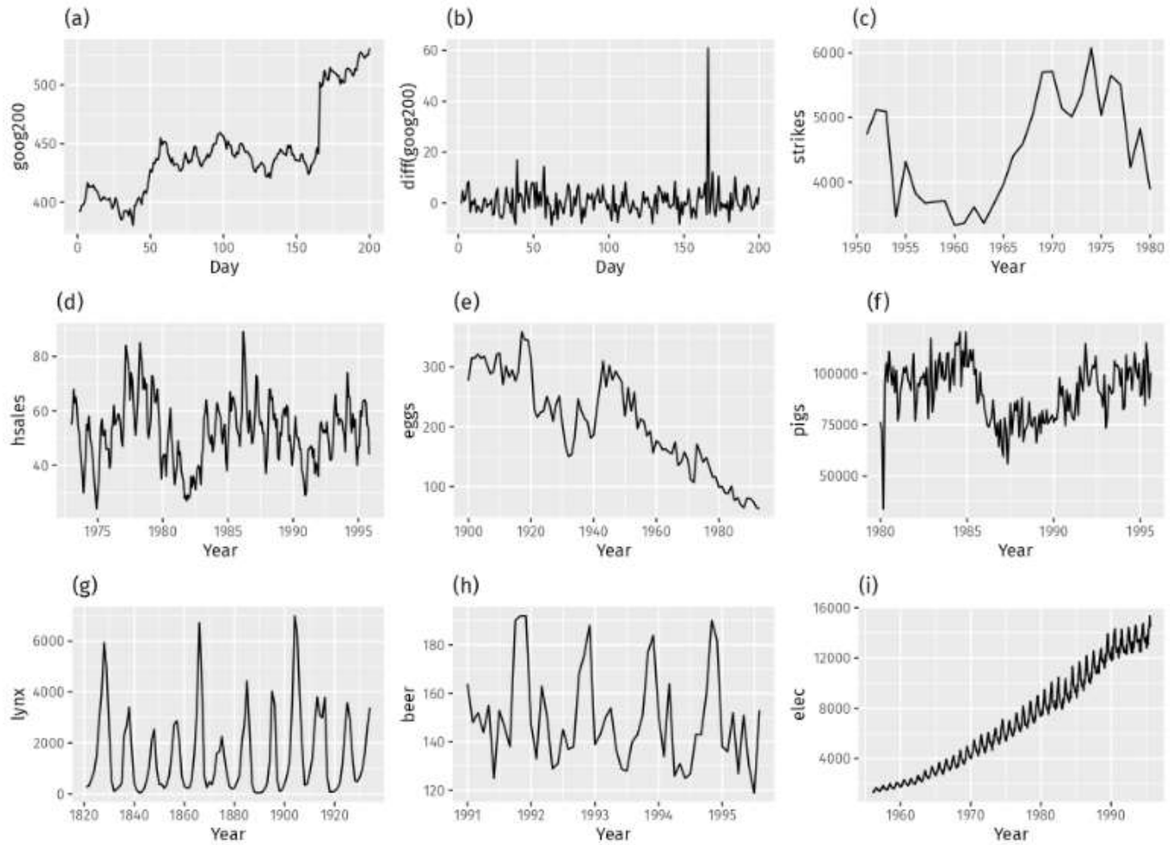
---

[1] تعريف مسار عشوائي ما يكافئ معرفة القانون الاحتمالي لكل شعاع $(X_{t+1}, ..., X_{t+h})$، حيث h عدد صحيح. التعريف الدقيق للاستقرارية (sens strict) هو أن القانون الاحتمالي ل $(X_{t+1}, ..., X_{t+h})$ يتوقف فقط على h. هذا التعريف قد يكون أحيانا شديد التقييد، لذلك يفضل عليه تعريف الاستقرارية من الدرجة الثانية، وهو أن يكن من أجل أي قيمة ل h فإن: $E(X_{t+h})$ و $cov(X_t, X_{t+h})$ لا تتوقف إلا على h.

هذين التعريفين يتكافئان في حالة المسار الطبيعي (عندما يكون القانون الاحتمالي لكل شعاع $(Xt+1, ..., Xt+h)$ يتبع التوزيع الطبيعي).

من المهم التمييز بين الاستقرارية والضجيج الأبيض (bruit blanc). هذا الأخير يعرف (التعريف الدقيق) بأنه حالة الغياب الكلي لكل ارتباط ذاتي في السلسلة، أي أن المتغيرات X_t مستقلة ولها ذات التوزيع الاحتمالي، ونكتب: (Xt) sont IID

التعريف الأقل تشددا للضجيج الأبيض هو أن تكون المتغيرات (X_t) لها نفس التوزيع الاحتمالي وغير مرتبطة خطيا، ونكتب $(X_t) \rightarrow WN(0, \sigma^2)$. في حالة الضجيج الأبيض الطبيعي فإن التعريفين يتطابقان.



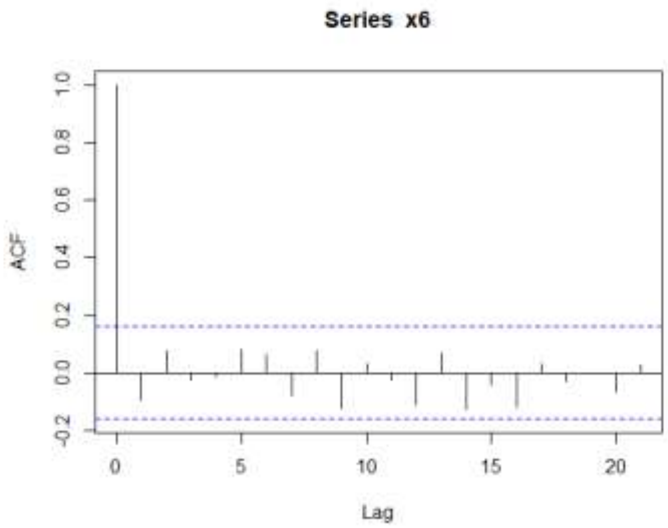
(a) Google stock price for 200 consecutive days; (b) Daily change in the Google stock price for 200 consecutive days; (c) Annual number of strikes in the US; (d) Monthly sales of new one-family houses sold in the US; (e) Annual price of a dozen eggs in the US (constant dollars); (f) Monthly total of pigs slaughtered in Victoria, Australia; (g) Annual total of lynx trapped in the McKenzie River district of north-west Canada; (h) Monthly Australian beer production; (i) Monthly Australian electricity production.

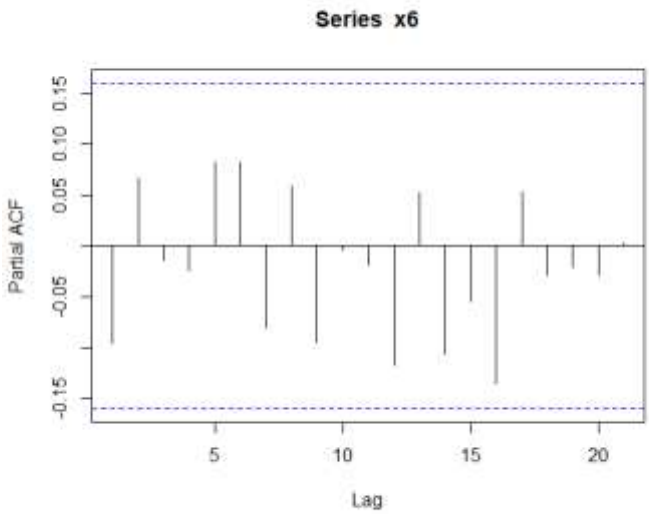
تفحص الاستقرارية بدالة الارتباط الذاتي

يمكن كذلك التمثيل البياني لدالتي الارتباط الذاتي ACF (أو أيضا الارتباط الذاتي الجزئيPACF). في حالة السلسلة غير المستقرة غالبا يكون الارتباط الذاتي الأول r1 عاليا وموجبا، وتنحدر الارتباطات الذاتية إلى الصفر ببطء نسبيا؛ بينما في السلسلة المستقرة تنحدر الارتباطات الذاتية ACF بسرعة نسبيا نحو الصفر، وتأتي غير دالة من البداية.

مثال: أنشىء سلسلة مستقرة باستخدام (rnorm) ثم مثل لها دالتي الارتباط الذاتي والارتباط الذاتي الجزئي.

```
 #ACF and PACF to check for stationarity
x6 <- rnorm (150)
plot.ts(x6)
acf(x6)
pacf(x6)
```



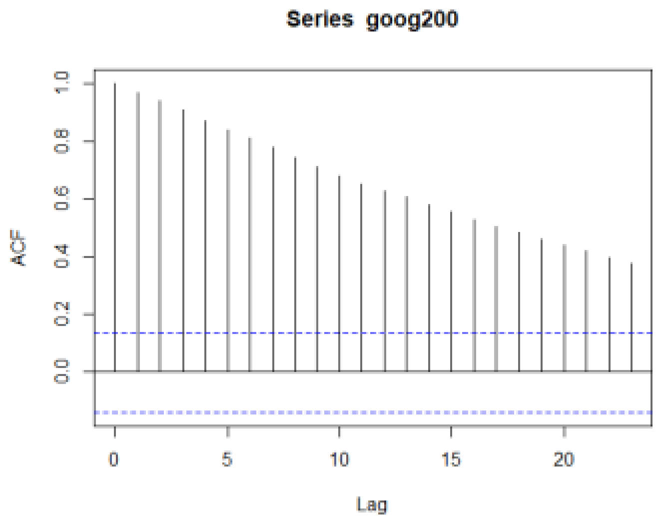


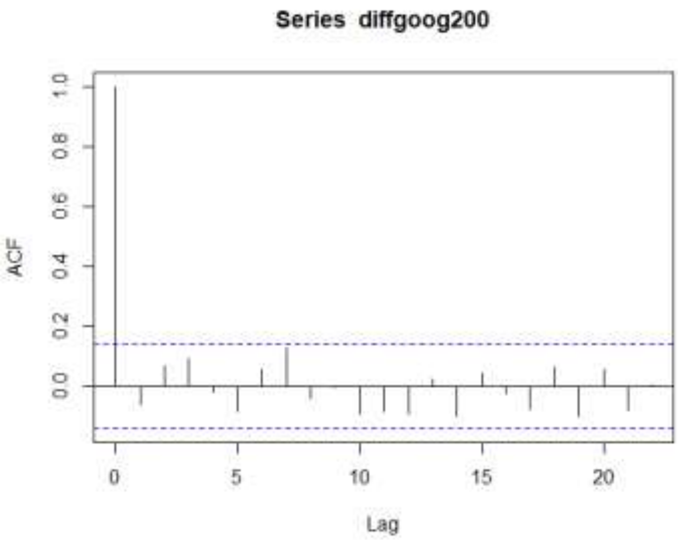
لاحظ أن الرسمين يبينان أن دالة الارتباط الذاتي ACF أتت بارتباطات ذاتية غير دالة من البداية وكذلك دالة الارتباطي الذاتي الجزئيPACF. هذا لأن المتغيرة هي متغيرة عشوائية طبيعية.

مثال 2. استخرج ACF لكل سعر سهم قوقل ل 200 يوم – السلسلة موجودة ضمن قواعد بيانات R باسم goog200– وقم بتفريق السلسلة ذاتها ، وقارن بين ACF لكل من السلسلتين قبل وبعد التفريق.

```
diffgoog200<-diff(goog200)
acf(goog200)
acf(diffgoog200)
```

Output :





الفرق واضح بين الرسمين، فالتفريق سمح بالتخلص من الارتباط الذاتي الموجود في السلسلة والحصول على سلسلة مستقرة.

تفحص الاستقرارية بالاختبار الاحصائي ديكي– فولر

طريقة أخرى لتفحص الاستقرارية هي إجراء اختبار إحصائي، وغالبا يستخدم الاختبار المضخم لديكي – فولر [1] Augmanted Dickey-Fuller test (ADF)،.

[1] الاختبار ADF يختبر نوعا من عدم الاستقرارية هو جذر الوحدة؛ وهو عندما تكون القيم المستقبلية للسلسلة متأثرة بالماضي. لنبدأ بحالة AR(1). يقوم ADF باختبار معامل انحدار y على تأخر y: $y_t = c + Ø_1 y_{t-1} + \varepsilon_t$. الفرضية الصفرية هي أن $Ø_1 = 1$ والفرضية البديلة هي $Ø_1 < 1$. هذه المعادلة نحولها إلى التفريق بطرح y_{t-1} من الطرفين فنحصل على :

$\Delta y_t = c + (Ø_1 - 1) y_{t-1} + \varepsilon_t = c + \lambda y_{t-1} + \varepsilon_t$. الفرضية الصفرية تصبح أن معامل الانحدار λ يساوي 0 (أي أن $Ø_1$ يساوي 1، وهو مصدر كلمة جذر الوحدة) وفي هذه الحالة تكون السلسلة غير مستقرة. الفرضية البديلة هي λ أقل من صفر (أي $Ø_1$ أقل من 1). في حال رفض الفرضية الصفرية نستدل على أن السلسلة مستقرة. إحصائية الاختبار تحسب بالطرقة العادية كما في اختبار تحليل الانحدار البسيط، أي قيمة

الفرضية البديلة لاختبار ADF هي الاستقرارية، أي أن الاختبار يختبر في الواقع عدم الاستقرارية. الفرضية الصفرية: معامل تأثير تأخير y على تفريقها يساوي 1 (أو أكبر)، والفرضية البديلة هي أن المعامل **أقل من 1**. إذا رفضت الفرضية الصفرية نستدل على أن السلسلة مستقرة. على العكس الفرضية البديلة لاختبار KPSS هي عدم الاستقرارية، فإذا رفضت الفرضية الصفرية نواصل التفريق.

في بعض البرمجيات (مثل STATA) يعطي الاختبار ADF إحصائية الاختبار والقيم الحرجة الجدولية من أجل مستويات دلالة 0.01، 0.05، 0.10. غالبا نختار مستوى الدلالة 0.05. نرفض الفرضية الصفرية (ونستدل على الاستقرارية) فقط إذا جاءت إحصائية الاختبار (بالقيمة المطلقة) أكبر من القيمة المطلقة للقيمة الحرجة. في برمجيات أخرى (مثل R وSPSS) لدينا مباشرة مستوى المعنوية الفعلي (p-value) وهذا يعني أننا نرفض الفرضية الصفرية إذا جاء مستوى المعنوية الفعلي أقل من 0.05. بعض البرمجيات (مثل Python) تعطي الاثنين: القيم الجدولية ومستوى المعنوية.

تفحص الاستقرارية باختبار KPSS

اختبار KPSS أو Kwatkowski-Phillips-Schmidt-Shin testهو اختبار مكمل لاختبار ADF، أي أن أحدهما لا يغني ببساطة عن الآخر. يختلف KPSS عن DF وADF في أكثر من ناحية، ومن أبرزها أن KPSS ينطلق من فرضية صفرية هي الاستقرارية، والفرضية البديلة هي وجود جذر الوحدة عكس الاختبار ديكي – فولر. الاختلاف الآخر المهم هو أن KPSS صنفين: صنف أول يستكشف عدم

المعامل مقسوما على خطئه المعياري، لكن الاختبار ستيودنت لا يصلح في هذه الحالة (لأننا ننطلق من فرضية عدم الاستقرارية وفي هذه الحالة لا تنطبق نظرية النهاية المركزية) وتستخدم القيم الجدولية لتوزيع ديكي–فولر.

لكن ليست كل السلاسل لها تأخر واحد فقط، وهناك حالة افتراض انعدام الثابت وحالة إدراج الزمن نفسه في النموذج. هناك في الواقع 4 حالات:

أ- صيغة اختبار DF لحالة افتراض عدم وجود ثابت ولا توجه (no trend no drift)، أي نموذج إنحدار التفريق على تأخر y أي على y_{t-1} فقط، وهو السير العشوائي (الزمن t لا يدخل في دالة النموذج)، $\Delta y_t = \lambda y_{t-1} + \varepsilon_t$

ب- صيغة DF لحالة عدم افتراض انعدام الثابت، أي نفس المعادلة لكن تتضمن ثابت (لكن لا نختبره)؛ $\Delta y_t = c + \lambda y_{t-1} + \varepsilon_t$

ت- صيغة DF لحالة افتراض وجود كل من الثابت والتوجه، وهنا يدخل الثابت وندرج الزمن t كمتغير مستقل في المعادلة، ولكن لا نختبر إلا الميل. $\Delta y_t = c + c_1 t + \lambda y_{t-1} + \varepsilon_t$

ث- الصيغة ADF هي لحالة إدراج عدة تأخرات وليس تأخر واحد: $y_t = c + \varnothing_1 y_{t-1} + \varnothing_2 y_{t-2} + \varnothing_3 y_{t-3} + \ldots + \varnothing_p y_{t-p} + \varepsilon_t$. هنا أيضا نحول الدالة إلى التفريق فنحصل على:

$$\Delta y_t = c + (\varnothing_1 - 1) y_{t-1} + \sum_{i=1,p} \beta_i \Delta y_{t-i} + \varepsilon_t \ .$$

هنا أيضا الاختبار هو فقط لمعامل تأثير تأخير y على التفريق. البرمجيات الإحصائية مثل Eviews تعطي إمكانية استخدام أي من الخيارات السابقة مع إمكانية تحديد عدد التأخرات آليا. المفاضلة بين هذه الخيارات ليست بالأمر السهل أو الهين، لأنها تأثر مثلا على قوة الاختبار. بعض الباحثين ببساطة يلجأ إلى إجراء الاختبار في الحالات المختلفة للتحقق من أن الاستقرارية محققة فيها جميعا، لكن ماذا إن تحققت الاستقرارية في حالات ولم تتحقق في أخرى. الأبحاث مستمرة لحل مثل هذه الإشكالات، وقدمت حلول مختلفة.

الاستقرار حول المتوسط (وهي الصيغة العادية) وصنف ثاني يستكشف ما يسمى عدم الاستقرارية حول توجه (trend stationary)، وهذا الصنف لاختبار KPSS فرضيته الصفرية هي أن السلسلة مستقرة حول توجه ما، والفرضية البديلة هي أنها ليست مستقرة حول توجه، لذلك إذا جاءت السلسلة مستقرة حسب KPSS من الصنف الثاني وغير مستقرة حسب اختبار ADF فهذا يعني أن السلسلة مستقرة حول توجه ما وهذا النوع من السلاسل يمكن نمذجته. المقصود بالتوجه هنا هو توجه دائم (deterministic trend) أو نمط يبقى في السلسلة إلى نهايتها وإن انقطع مؤقتا بسبب صدمة مثلا يعود مجددا.

إذا وجد أن السلسلة غير مستقرة يتعين تحويل السلسلة إما بالتفريق أو التحويل أو كلاهما. التفريق يستخدم للتغلب على حركة المتوسط، وعلى التوجه والموسمية؛ بينما التحويل يستخدم في معالجة مشكلة تغير التباين. يمكن أن يستخدم التحويل اللغرتمي أيضا لتقليص الانحدار الشديد للسلسلة. غالبا ما يكفي تفريق من الدرجة الأولى (توجه خطي) أو الثانية (توجه تربيعي: إرجع إلى فقرة التفريق في درس تحليل التوجه).

$$d_t = y_t - y_{t-1} \; ; \; d_2t = d_t - d_{t-1}$$

لاحظ:

- بما أن التفريق غالبا سيتضمن قيما سالبة فإن البرمجيات تتخلص من هذه القيم السالبة أتوماتيكيا عند التحويل اللغرتمي بإضافة قيمة إلى التفريق، ستنشأ أيضا صعوبة في تفسير المتغيرة المحولة[1].

إختبار ADF في R

لإجراء اختبار الاستقرارية في R يمكن استخدام الدالة ()adf.test لاختبار ADF أو ()kpss.test لاختبار KPSS وكلاهما يوجد في 'tseries' package.

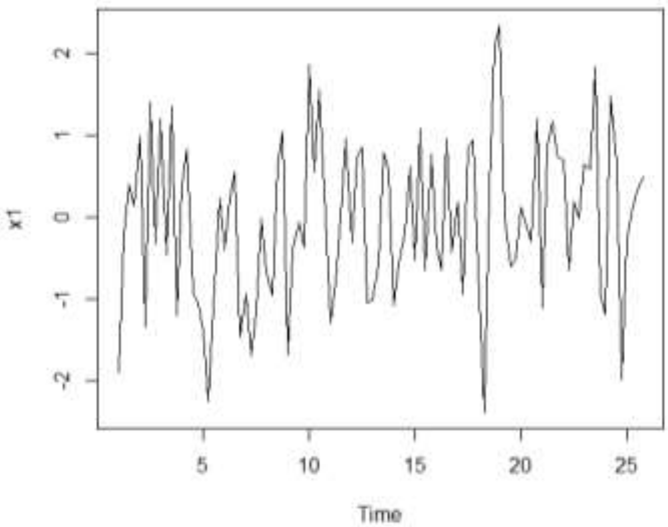
مثال 1: لننشئ سلسلة باستخدام التوزيع الطبيعي المعياري، حجمها 100، وذات موسمية درجتها 4 (بيانات فصلية). نسمي السلسلة x1. للتمكن من إجراء الاختبار ADF نثبت ثم نحمل الحزمة tseries . قبل الاختبار، نقوم أولا بإظهار بيانات السلسلة ثم نمثلها بيانيا.

```
x1<-ts(rnorm(100),frequency=4)
#install.packages("tseries") #if needed
library(tseries)
x1
plot(x1)
adf.test(x1)
```

[1] متوسط الفروق من الدرجة الأولى (m_{d1}) هو ميل y_t على y_{t-1} ، أي معامل AR(1)، أي ميل التوجه الخطي في السلسلة. عدد المشاهدات يتناقص بدرجة عند التفريق الأول، وبدرجتين عند التفريق الثنائي وهكذا.

Output:

```
          Qtr1        Qtr2        Qtr3        Qtr4
1  -1.89646436 -0.22703569  0.39193281  0.13158407
2   0.98150434 -1.34322207  1.39253464 -0.33225581
3   1.20189484 -0.46853576  1.34758900 -1.20179542
4   0.38255761  0.82969894 -0.92809172 -1.06445340
5  -1.39361456 -2.24678844 -1.09213499  0.22083153
6  -0.40662404  0.23379483  0.55801070 -1.46189484
7  -0.94066387 -1.70020728 -1.16089914 -0.01755214
8  -0.68671869 -0.96356948  0.63964171  1.04898393
9  -1.68064016 -0.38846565 -0.06130134 -0.37102669
10  1.86555994  0.54706201  1.54663683  0.01620678
11 -1.30334673 -0.80025589  0.01281325  0.95214569
12 -0.30880604  0.73410984  0.85153425 -1.04570910
13 -1.02324633 -0.58497467  0.77720038  0.58607254
14 -1.07259676 -0.57013846 -0.18970287  0.61743713
15 -0.52037339  1.07567049 -0.64744161  0.77326830
16 -0.35856023 -0.65641459  0.95765451 -0.44086032
17  0.17993828 -0.94642729  0.84377336  0.94216158
18 -0.63087236 -2.39317960  0.33757962  2.06976812
19  2.34509716  0.01710032 -0.60514938 -0.50823865
20  0.12635343 -0.08838281 -0.28969659  1.20343972
21 -1.10935256  0.87557435  1.16920948  0.73288681
22  0.69237493 -0.64499416  0.18137784 -0.01697952
23  0.63813212  0.58725441  1.82670072 -0.94603194
24 -1.19222236  1.48094764  0.66378229 -1.99006734
25 -0.27442396  0.08043719  0.29849570  0.48351084
```



Augmented Dickey-Fuller Test

data: x1
Dickey-Fuller = -4.743, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary

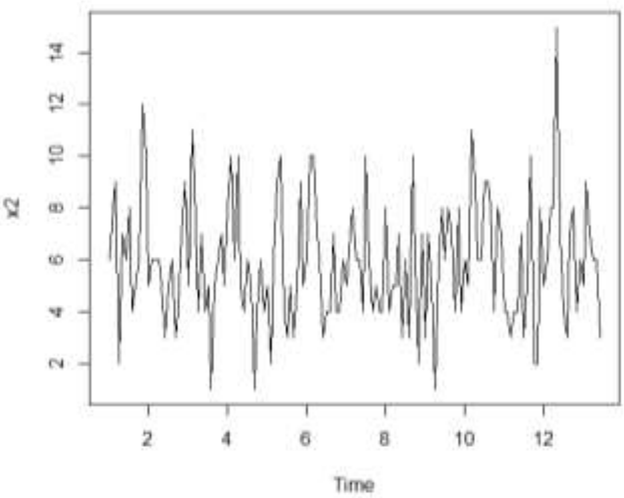
الرسم يظهر أن السلسلة تتباين حول متوسط ثابت (الصفر) وبدون توجه أو موسمية والتباين ثابت، فهي تبدو مستقرة. الاختبار الإحصائي ADF أكد ذلك: مستوى المعنوية أقل من 0.05 وبالتالي يمكن قبول الفرضية البديلة وهي أن السلسلة مستقرة.

مثال 2: لننشئ سلسلة باستخدام التوزيع بواسون بمعلمة 6، حجمها 150، ذات موسمية درجتها 12 (بيانات شهرية)، ثم نجري الاختبار ذاته (لا حاجة هنا إلى إعادة تحميل الحزمة)، لكن قبل الاختبار نظهر قيم السلسلة ونمثلها بيانيا.

```
x2 <- ts(rpois(150,6),frequency=12)
x2
plot(x2)
adf.test(x2)
```

Output:

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	6	8	9	2	7	6	8	4	5	6	12	10
2	5	6	6	6	5	3	5	6	3	4	7	9
3	5	11	9	4	7	4	5	1	5	6	7	5
4	8	10	6	10	5	4	6	5	1	4	6	4
5	5	2	6	9	10	4	3	5	3	5	9	5
6	6	10	10	7	6	3	4	4	7	4	4	6
7	5	7	8	6	6	4	10	6	4	5	4	4
8	8	4	5	5	7	3	6	3	10	6	2	7
9	3	7	5	1	6	8	6	8	7	4	8	4
10	6	5	11	9	6	6	9	9	8	4	8	7
11	4	4	3	4	4	7	3	5	10	2	2	8
12	5	6	8	8	15	7	4	3	7	8	4	6
13	5	9	7	6	6	3						




```
Augmented Dickey-Fuller Test

data:  x2
Dickey-Fuller = -5.5922, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
Warning message:
In adf.test(x2) : p-value smaller than printed p-value
```

تظهر النتيجة أن مستوى المعنوية أقل من 0.05 وبالتالي يمكن قبول الفرضية البديلة وهي أن السلسلة مستقرة.

في حال جاءت p-value أكبر من 0.05، لا يمكن الاستدلال على الاستقرارية وفي هذه الحالة يتعين كما ذكرنا اللجوء إلى التفريق (الدالة ()diff) أو التحويل أو كلاهما. هذا ما يظهر من المثال التالي:

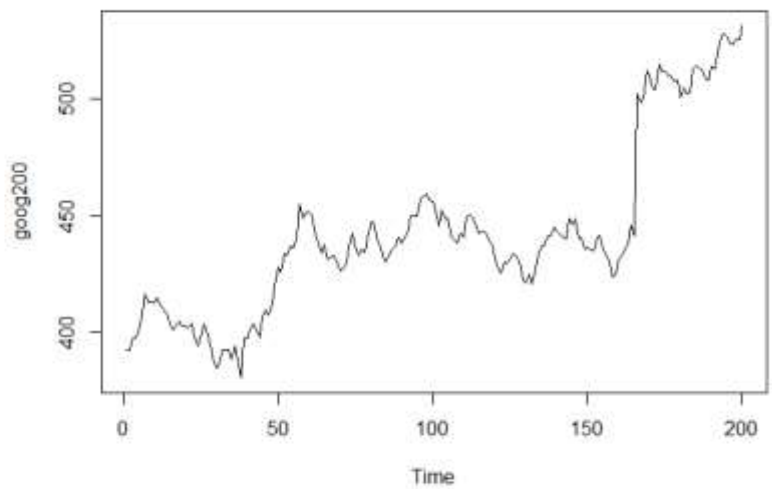
مثال3. نختبر السلسلة goog200 (سعر سهم 'قوقل' في 200 يوم)، ولفعل ذلك نحمل أولا الحزمة fpp2 التي تنتمي إليها:

```
library(fpp2)
adf.test(goog200)
```

Output:

```
Augmented Dickey-Fuller Test

data:  goog200
Dickey-Fuller = -1.7775, Lag order = 5, p-value = 0.6693
alternative hypothesis: stationary
```

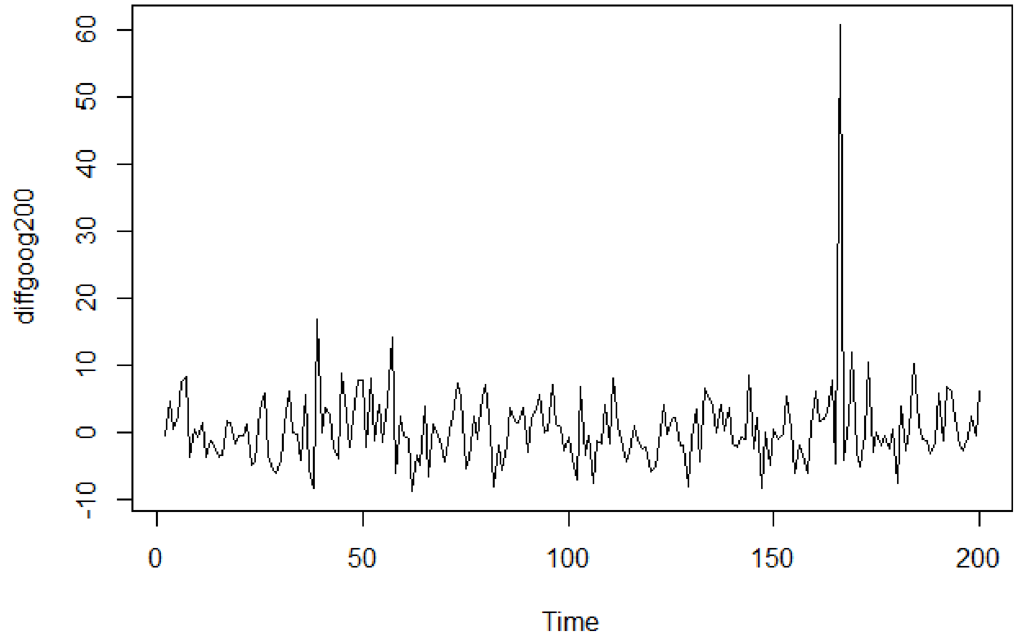
بما أن مستوى المعنوية أكبر بكثير من 0.05 نقبل الفرضية الصفرية بأن السلسلة غير مستقرة. لمعرفة السبب نظهر الرسم البياني للسلسلة.



الرسم يبين أن المتوسط غير ثابت وهذا سبب عدم استقرارية السلسلة. في مثل هذا الحالة يتعين إذا أردنا التنبؤ باستخدام طريقة ARIMA تفريق السلسلة؛ كما يلي:

```
diffgoog200<-diff(goog200)
plot(diffgoog200)
```

Output:



يبدو من الرسم أن التفريق من الدرجة الأولى هو كاف لمعالجة تغير المتوسط. للتحقق نختبر السلسلة الجديدة:

```
adf.test(diffgoog200)
```

output :

```
                    Augmented Dickey-Fuller Test

data:  diffgoog200
Dickey-Fuller = -5.7435, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(diffgoog200) : p-value smaller than printed p-value
```

النتيجة تأكد ما أظهره الرسم من أن السلسلة الجديدة (بعد التفريق) مستقرة. يمكن إذن استخدام السلسلة الجديدة للحصول على نموذج ARIMA، ب d قيمتها 1. لو وجدنا التفريق لم يكف للحصول على الاستقرارية نقوم بتفريق من الدرجة الثانية (تفريق التفريق).

اختبار KPSS في R

الدالة المستخدمة في R هي kpss.test().

مثال 4 : قم بإنشاء سلسلة مستقرة ثم اختبر عدم الاستقرارية باختبار KPSS لاستقرار المستوى(level- stationary) ولاستقرار التوجه (trend- stationary).

من أجل **اختبار KPSS لاستقرار المستوى:** حالة استقرار المستوى:

```
set.seed（100）
x3 <- rnorm（100）
kpss.test(x3)
```

Output

```
KPSS Test for Level Stationarity

data:  x3
KPSS Level = 0.13609, Truncation lag parameter = 4, p-value = 0.1

Warning message:
In kpss.test(x3) : p-value greater than printed p-value
```

بما أن p-value>0.05 لا نرفض الفرضية الصفرية، أي أن السلسلة مستقرة.

من أجل **اختبار KPSS لاستقرار التوجه:** حالة استقرار التوجه:

```
kpss.test(x3,null="Trend")
```

Output:

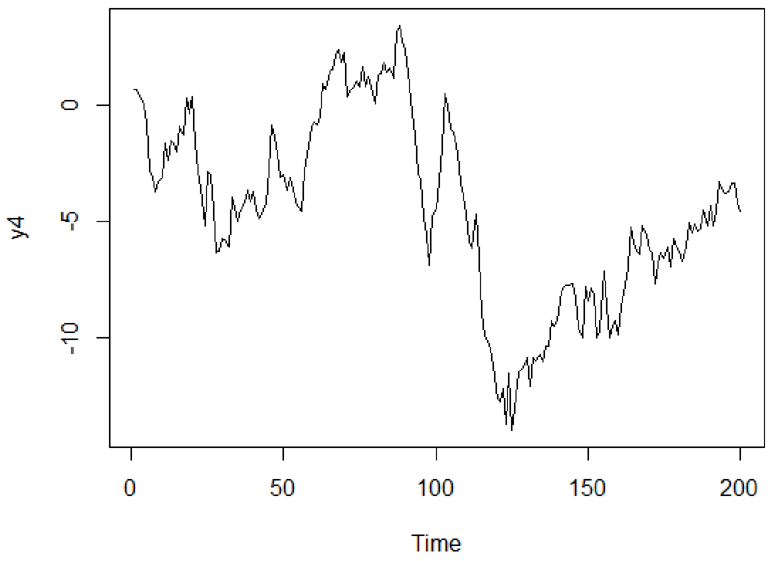
```
KPSS Test for Trend Stationarity
data:  x3
KPSS Trend = 0.034563, Truncation lag parameter = 4, p-value = 0.1

Warning message:
In kpss.test(x3, null = "Trend") : p-value greater than printed p-value
```

هنا أيضا لا نرفض الفرضية الصفرية (استقرار التوجه). يمكن أيضا طلب الاختبارين معا: بحيث تكون الفرضية الصفرية هي اختبار استقرار المستوى أو اختبار استقرار التوجه وتكون الفرضية البديلة هي وجود جذر الوحدة.

مثال 5. حالة عدم استقرار المستوى:

```
x4<-rnorm（200）
y4 <- cumsum(x4)  # has unit root
plot.ts(y4)
kpss.test(y4)
```

output :

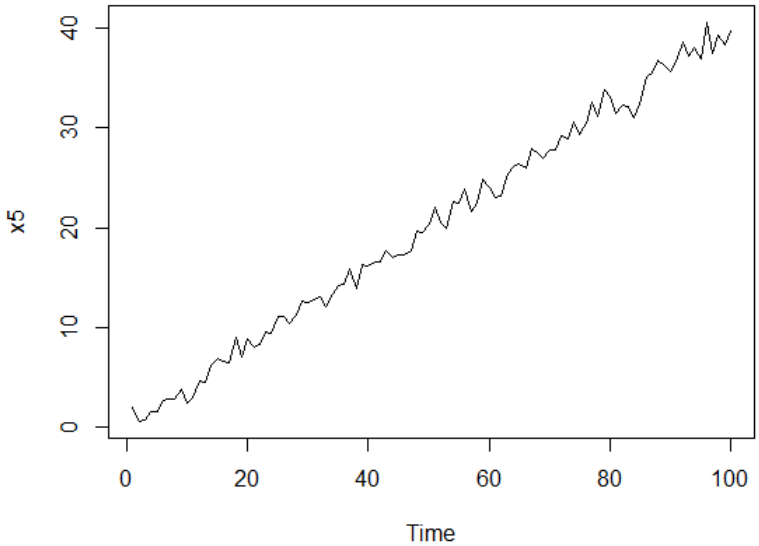



```
KPSS Test for Level Stationarity

data:  y4
KPSS Level = 1.5264, Truncation lag parameter = 4, p-value = 0.01

Warning message:
In kpss.test(y4) : p-value smaller than printed p-value
```

حالة استقرار التوجه (مع عدم استقرار المستوى)

```
x5 <- 0.4*(1:100)+rnorm(100)  # is trend stationary
plot.ts(x5)
kpss.test(x5, null = "Trend")
```

Output :




```
KPSS Test for Trend Stationarity

data:  x5
KPSS Trend = 0.026369, Truncation lag parameter = 4, p-value = 0.1
```

```
Warning message:
In kpss.test(x5, null = "Trend") : p-value greater than printed p-value
```

يتم إذن تحديد قيمة d لتحقيق استقرارية السلسلة قبل المعلمتين p وq. غالبا ما تكون d صغيرة: 0 أو 1 أو أحيانا 2؛ (2: إذا كان لها توجه تربيعي لا خطي). قيمة d تكون 0 إذا كانت السلسلة لا تحتاج إلى تفريق، وفي هذه الحالة يقصى d من النموذج فنستخدم AR(p) أو MA(q) أو ARMA(p,q).

2-2. تحديد النموذج (أفضل قيم لمعالم النموذج p, q)

البرمجيات الإحصائية اليوم (مثل SPSS وR و Eviews) مزودة بخوارزميات تمكن من تجربة عدد كبير من النماذج ومن ثم انتقاء أفضلها بناء على عدد من المعايير، وبهذا تختصر هذه البرمجيات الجهد على المحلل في بحثه عن النموذج الأفضل.

إذا شاء المحلل تحديد النموذج بدون الاستعانة بهذه الطرق الاتوماتيكية، فإن تحديد النموذج يكون بالاستعانة ب ACF وPACF (عدد الارتباطات الدالة) للسلسلة (الأصلية أو المحولة إن حولنا). ليس هناك طريقة محددة لتحديد النموذج صالحة في كل الحالات، ولكن هناك قواعد عامة وخطوط عريضة يستخدمها المحلل أو الباحث، منها (أنظر الملحق):

- نختار نموذج AR(p): إذا كان ACF ينحدر بسرعة نحو الصفر وPACF له اختراقات دالة في التأخرات من 1 إلى p.
- تختار نموذج MA(q): إذا كان PACF ينحدر بسرعة نحو الصفر وACF له اختراقات دالة من 1 إلى q
- قد يتحقق الاثنان وفي هذه الحالة ندمج بين النموذجين ونستخدم نموذج ARMA.

2-3. تقدير النموذج

عندما يتم تحديد معالم النموذج الثلاث (p, d, q) يمكن تقدير معاملاته (معاملات الانحدار والثابت) واختبارها باستخدام البرمجيات وكتابة دالة النموذج.

2-4. تفحص النموذج (المتبقي)

ما هي الاشتراطات؟

المهم في النموذج أن يلتقط (يستغل) كل المعلومات الموجودة في السلسلة ولا يبقي إلا متغيرات عشوائية مستقلة (ضجيج أبيض). لذلك يتعين التحقق من المتبقي من خلال مجموعة من الاختبارات البيانية والإحصائية. إن لم يظهر أن المتبقي عشوائيا عندها يتعين تحسين النموذج.

نعرف المتبقي بالفرق بين المتغيرة y_t والتوقع $\hat{y}_t$، ونكتب:

$$e_t = y_t - \hat{y}_t$$

يستهدف تحليل المتبقي التحقق من عدم وجود علاقات داخله. وجود علاقات في المتبقي يعني أن النموذج المستخدم لم يلتقط كل العلاقات الموجودة، وبالتالي وجود إمكانية لتحسين النموذج بتعديله أو استبداله بما يحقق التقاط كل العلاقات الموجودة في البيانات. يبسط هايدمان ما هو مطلوب تحققه في المتبقي في كتابه "التنبؤ: مبادئ وممارسات" بلغة مفهومة ومختصرة نستفيد منها هنا (Hyndman, 2018).

نستهدف في المتبقي أمرين أساسيين:

- أن يكون خاليا من الارتباط الذاتي، وجود ارتباط غير معدوم يعني أننا لم نصل بعد إلى النموذج الذي يستوعب كل المعلومة الموجودة في البيانات[1].
- أن يكون متوسطه معدوما. إختلاف متوسط المتبقي عن الصفر يعني أن التنبؤ منحاز.

هذين الشرطين هما ضروريان للوصول إلى النموذج الأمثل لكنهما ليسا كافيين للحكم على النموذج أو المفاضلة بين النماذج. كل نموذج لا يلبي هذين المعيارين أو أحدهما يمكن تحسينه، لكن أي نموذج يلبي هذين المعيارين ليس بالضرورة لا يمكن تحسينه، فقد نحصل على عدة نماذج تحقق هذين المعيارين معا، وعندها نفاصل بينها على أساس معايير أخرى.

التحيز الناتج عن اختلاف متوسط المتبقي عن الصفر يمكن تصحيحه بسهولة بإضافة هذا المتوسط إلى التنبؤ، على عكس مشكلة وجود ارتباط ذاتي. بالإضافة إلى الشرطين السابقين يستحسن (لكن ليس ضروريا) أن يحقق المتبقي الشرطين التاليين لتسهيل حساب مجال الثقة حوله:

- أن يكون تباين المتبقي ثابتا،
- أن يتبع المتبقي التوزيع الطبيعي.

---

[1] بالإضافة إلى عدم إمكانية بناء مجال ثقة حول التنبؤ ولو بطريقة البوتسترب.

النموذج الذي لا يحقق هذين الشرطين ليس بالضرورة قابلا للتحسين، وقد يفيد تحويل بوكس–كوكس في تحقيق ثبات تباين المتبقي واتباعه للتوزيع الطبيعي من أجل حساب مجال ثقة حول التقدير، وإلا فقد يتعين تغيير جذري للمقاربة المستخدمة.

كيفية التحقق من الاشتراطات بالرسم

هناك ثلاثة رسوم بيانية عادة ما نتفحصها من أجل التحقق من الاشتراطات الأربع المذكورة:

1. التمثيل البياني لسلسلة المتبقي
2. المدرج التكراري للمتبقي *(histogram)*
3. دالة الارتباط الذاتي

يفيد الرسم الأول، التمثيل البياني الخطي للمتبقي، في التحقق من تقييم مدى تماثل توزيع قيم المتبقي حول متوسط معدوم وثبات تباينه طوال تاريخ السلسلة. ما نريد أن نراه هو سحابة ممتدة أفقيا وعشوائيا على خط الصفر. يسمح الرسم الأول أيضا باستكشاف القيم الشاردة إن وجدت والبحث بالتالي عن تفسيرها.

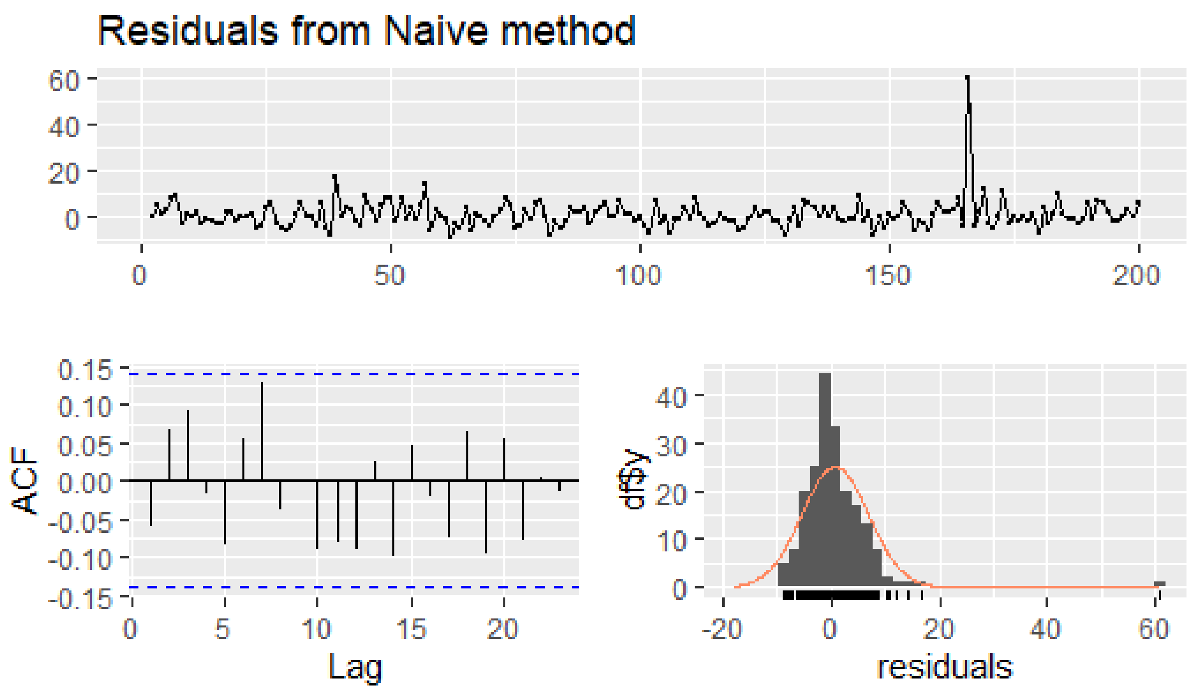
يفيد الرسم الثاني، المدرج التكراري للمتبقي، في تقييم مدى طبيعية توزيع المتبقي، بأن يكون له الشكل الجرسي المعروف، بدون أن يكون ممتدا من اليمين أو اليسار. تحقق هذا الشرط يفيد كما ذكرنا في حساب مجال الثقة حو التقدير وليس في قيمة التقدير نفسه، لذلك قد تكون التقديرات جيدة حتى إن لم يكن المتبقي طبيعيا.

يفيد الرسم الثالث، دالة الارتباط الذاتي *(ACF)*، في التحقق من عدم وجود ارتباطات ذاتية دالة، أي عدم وجود أعمدة صاعدة أو نازلة تخترق الخطين الأفقين المتقطعين حوالى الصفر. عدم تحقق هذا الشرط يعني أن النموذج يمكن تحسينه، وغالبا ما يفيدنا رسم دالة الارتباط الذاتي في فهم السلسلة وفي تحديد التعديلات التي يتعين إدخالها على النموذج.

مثال (Hyndman, 2018): نستخدم قاعدة البيانات الشهيرة goog200 من R وهي تحتاج إلى تحميل fpp2. تمثل هذه المتغيرة سعر سهم Google خلال 200 يوم، وهي جزء (200 يوم الأولى) من قاعدة أكبر بكثير هي goog .

فيما يلي نبين كيفية استخراج الرسوم البيانية الثلاث المذكورة في *R* دفعة واحدة، وكيف تفسر هذه الرسوم البيانية تجاه الاشتراطات الأربع المذكورة أعلاه. تسمح الدالة *checkresiduals()* في *R* بإظهار ثلاث رسوم بيانية (الرسوم المذكورة) بالإضافة إلى نتيجة اختبار إحصائي.

```
checkresiduals(naive(goog200))
```



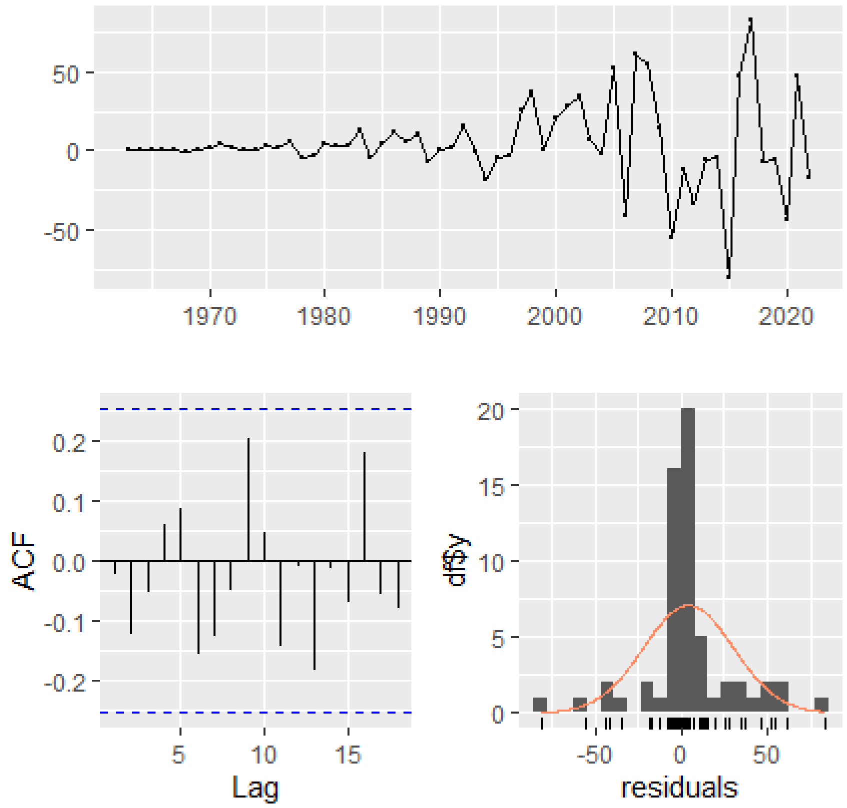

```
Ljung-Box test

data:  Residuals from Naive method
Q* = 11.031, df = 10, p-value = 0.3551

Model df: 0.   Total lags used: 10
```

يظهر الرسم الأول التمثيل البياني لقيم المتبقي، وهي هنا تظهر منتشرة بتماثل على خط الصفر ما عدا قيمة متطرفة واحدة يمكن الرجوع إليها. هذا الانتشار الأفقي المتوازن على جهتي المتوسط العليا والسفلى طيلة ماضي السلسلة يجعلنا نقبل بأن تباين المتبقي ثابت. استفدنا من هذا الرسم إذن عدم تحيز التقدير (المتوسط هو الصفر) وثبات التباين.

الرسم أسفله إلى اليمين يتفحص طبيعية المتبقي، ويظهر الرسم هنا ممتدا من اليمين حتى بدون احتساب القيمة المتطرفة (60). هذا يعني أن المتبقي ربما ليس طبيعيا، وهذا يأثر على مجال الثقة حول التقدير، لكن ليس على التقدير نفسه.

الرسم الثالث، وهو رسم دالة الارتباط الذاتي، وتظهر الدالة هنا عدم وجود ارتباطات دالة إحصائيا، ما كان سيدفعنا للعمل على تحسين النموذج. يمكن القول بأن النموذج المستخدم –على بساطته– ملائم (لكن ليس بالضرورة هو الأفضل) للتنبؤ بسعر سهم قوقل يوميا.

الرسم التالي يبين مخرجات برنامج R لاختبار المتبقي بالطرق الثلاث المذكورة لنموذج ARIMA مطبق على بيانات عدد الكلبة في التعليم العالي في الجزائر منذ الاستقلال. (الدالة المستخدمة في R للحصول عل هذا الرسم هي ()Checkresiduals):



نلاحظ في هذا الرسم أن التمثيل البياني للمتبقي (الرسم الأعلى) يظهر ثبات المتوسط لكن التباين غير ثابت، فهو يزيد مع مرور الزمن. دالة ACF تدل تظهر أن كل الارتباطات (الأعمدة الخارجة من خط الوسط) غير دالة (لا تقطع الخطين الأفقيين المتقطعين). المدرج التكراري يظهر إختلافا كبيرا عن المنحنى الطبيعي. إجمالا وجب في هذه الحالة تحسين النموذج أي البحث عن نموذج آخر.

2-5. المفاضلة بين النماذج من خلال (*BIC,AIC,AICc*)

يمكن أن يجد المحلل أكثر من نموذج يحقق عشوائية المتبقي، وفي هذه الحالة يقوم بالمفاضلة بينها من خلال مؤشرات الملائمة (fit indices) ومنها R^2 (أكبر يعني أفضل) ومؤشرات المعلومات **(BIC,AIC)**، حيث أدني قيمة لها هي الأفضل. مؤشرات المعلومات تحسب مدى التوازن بين المواءمة والاقتصاد في المعالم. AIC (Akaik Information Criterian)، BIC (Baesian Information Criterian)

6-2. التنبؤ

تقوم البرمجيات بالتنبؤ لمستقبل السلسلة المفرقة ثم بالتحويل من التفريق إلى السلسلة الأصلية.

عملية إعادة التحويل من التفريق إلى السلسلة الأصلية تتم رياضيا كما يلي: نفترض أن السلسلة الأصلية هي y_t والتفريق هو d_t حيث: $d_t = y_{t+1} - y_t$ ، وبالتالي $y_{t+1} = y_t + d_t$ نريد التنبؤ للتاريخ t+h إنطلاقا من t:

$$y_{t+h} = y_{t+h-1} + d_{t+h-1} = y_{t+h-2} + d_{t+h-2} + d_{t+h-1} = \ldots = y_{t+h-h} + d_{t+h-h} + d_{t+1} + d_{t+2} + \ldots + d_{t+h} = y_t + \sum_{i=0, h} d_{t+i}$$

3. دراسة حالة 1.

4. مراجع

Hyndman, R., & Athanasopoulos, G. (2018). *Forecasting: principles and practice* (éd. 2nd). https://otexts.com/. Consulté le August 1, 2023, sur https://www.otexts.org/fpp/8/1

عن ARIMA باستخدام R:

https://ademos.people.uic.edu/Chapter23.html

لتعلم استخدام R لتحليل السلاسل الزمنية عموما

http://a-little-book-of-r-for-time-series.readthedocs.io/en/latest/src/timeseries.html