

Approche de Box-Jenkins pour l'identification et l'estimation des modèles ARIMA -Partie théorique-

Yassine HALLAL

2023-01-29

Contents

Introduction	2
1. Approche de Box-Jenkins :	2
1.1. Stationnarité :	2
1.2. Identification :	2
ACF et PACF	2
1.3. Estimation :	3
1.4. Diagnostic :	3
1.5. Forecasting :	4
Références :	4

Introduction

Au XXe siècle, à l'époque où les ordinateurs étaient lents, la tâche était ardue. George Box et Gwilym Jenkins (Box et Jenkins, 1976) ont donc mis au point une méthodologie pour identifier et estimer les modèles ARIMA. Bien qu'il existe aujourd'hui des méthodes plus efficaces de sélection de l'ordre pour les ARIMA, certains de leurs principes sont toujours utilisés dans l'analyse des séries chronologiques et dans la prévision.

1. Approche de Box-Jenkins :

1.1. Stationnarité :

Avant d'effectuer toute analyse de séries temporelles, nous devons rendre les données stationnaires, ce qui est fait via les différences dans le contexte de l'ARIMA. Mais avant de faire quoi que ce soit, nous devons d'abord savoir si les données sont stationnaires ou non : une surdifférenciation est généralement nuisible au modèle et entraînerait des problèmes de mauvaise spécification. En même temps, dans le cas d'un sous-différenciation, il pourrait être impossible d'identifier le modèle correctement.

Il existe différentes façons de savoir si les données sont stationnaires ou non. La plus simple consiste à examiner les données : dans certains cas, il est évident que la moyenne des données change ou qu'il existe une tendance dans les données, la conclusion est donc relativement simple. Si les données ne sont pas stationnaires, l'étape suivante consisterait à faire des différences et à analyser à nouveau les données différenciées pour s'assurer que les secondes différences ne sont pas nécessaires.

L'approche plus formelle consisterait à effectuer des tests statistiques, tels que ADF ou KPSS. Notez qu'ils testent des hypothèses différentes :

1. Dans le cas du test ADF :

- H_0 : Les données **ne sont pas** stationnaires.
- H_1 : Les données sont stationnaires.

2. Dans le cas du test KPSS :

- H_0 : Les données sont stationnaires.
- H_1 : Les données **ne sont pas** stationnaires.

Les deux tests ont leurs avantages et leurs inconvénients et peuvent parfois se contredire.

1.2. Identification :

La première étape de l'approche de Box-Jenkins est "l'identification". Cette étape permet de déterminer l'ordre du modèle (p,q) en utilisant l'ACF & PACF. Ce n'est qu'après s'être assuré que les données sont stationnaires que nous pouvons passer à l'identification des ordres AR et MA.

ACF et PACF La fonction d'autocorrélation (ACF) et la fonction d'autocorrélation partielle (PACF) sont couramment utilisées pour déterminer les paramètres d'un modèle ARMA (Autoregressive Moving Average).

L'ACF montre la corrélation entre la série temporelle et ses valeurs retardées. Le PACF, quant à lui, montre la corrélation entre la série temporelle et ses valeurs retardées après avoir éliminé les effets des retards intermédiaires.

Les directives suivantes peuvent être utilisées pour déterminer les paramètres d'un modèle ARMA en utilisant ACF et PACF :

1. Pour le processus $AR(p)$, l'ACF diminuera soit de manière exponentielle, soit de manière alternée (en fonction des valeurs des paramètres), à partir du décalage p ;
2. Pour le processus $AR(p)$, le PACF diminuera brusquement juste après le décalage p ;
3. Pour un processus $MA(q)$, l'ACF chutera brusquement juste après le décalage q ;
4. Pour le processus $MA(q)$, le PACF diminuera soit de manière exponentielle, soit de manière alternée (en fonction des valeurs spécifiques des paramètres), à partir du décalage q .

Paramètre	ACF	PACF
AR	Geométrique	significatif jusqu'à p lag
MA	significatif jusqu'à q lag	Geométrique
ARMA	Geométrique	Geométrique

1.3. Estimation :

Cela implique l'estimation des paramètres du modèle spécifié à l'étape 1. Cette estimation peut être faite en utilisant le Maximum likelihood (MLE), ABC estimation, Bootstrap, Critères d'informations (AIC, AICc, BIC...) et Méthodes des moments. Le choix est fait en fonction du modèle.

1.4. Diagnostic :

Le diagnostic d'un modèle ARIMA se réfère au processus d'évaluation de l'adéquation du modèle aux données et aux hypothèses du modèle. Il existe plusieurs méthodes pour diagnostiquer l'adéquation d'un modèle ARIMA, notamment :

- **L'analyse des résidus** : Il s'agit d'examiner les résidus (c'est-à-dire les différences entre les valeurs observées et les valeurs prédites) pour vérifier la présence de modèles ou de tendances. Idéalement, les résidus devraient être un bruit blanc, ce qui signifie qu'ils devraient avoir une moyenne nulle, une variance constante et aucune autocorrélation.
- **Test de Ljung-Box** : Il s'agit d'un test statistique qui vérifie s'il existe une autocorrélation significative dans les résidus. Si le test montre une autocorrélation significative, cela indique que le modèle peut ne pas capturer de manière adéquate la dépendance temporelle des données.
- **Invertibilité** : Les inverses des racines AR et MA de l'équation caractéristique peuvent être utilisées pour vérifier si le processus impliqué par le modèle est stationnaire et inversible. Pour que les parties AR et MA du processus soient stationnaires et inversibles, respectivement, les racines inversées doivent dans chaque cas être inférieures à 1 en valeur absolue.
- **Test de normalité** : Il s'agit de tester si les résidus suivent une distribution normale. Si les résidus ne sont pas distribués normalement, cela peut indiquer que le modèle est mal spécifié ou qu'il y a des valeurs aberrantes dans les données.
- **Détection des valeurs aberrantes** : Il s'agit de vérifier la présence de valeurs aberrantes ou d'observations influentes dans les données qui peuvent affecter l'ajustement du modèle.
- **Évaluation des prévisions** : Il s'agit d'évaluer la précision des prévisions du modèle en les comparant aux valeurs observées. Les mesures courantes d'évaluation de la précision des prévisions comprennent l'erreur absolue moyenne, l'erreur quadratique moyenne et le coefficient de détermination.

En effectuant ces diagnostics, on peut déterminer si le modèle ARIMA capture correctement les données et peut identifier tout problème potentiel d'ajustement du modèle. Si des problèmes sont identifiés, le modèle peut être affiné ou des modèles alternatifs peuvent être envisagés.

1.5. Forecasting :

Lorsqu'on identifie le modèle qui respecte l'ensemble des points envisagés, on peut utiliser ce modèle pour la prévision.

Références :

- [1] Box, George E. P., Gwilym M. Jenkins, and Gregory C. Reinsel. Time Series Analysis: Forecasting and Control. 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 1994.
- [2] Brooks, C. (2012) Introductory Econometrics for Finance. 3rd edn. Cambridge University Press.