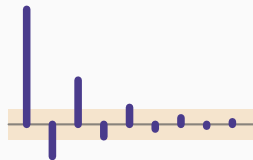


Modélisation ARMA des rentabilités financières

Chapitre 7 : Les processus mixtes ARMA(p,q)



Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger · 27 août 2026

Avant de commencer

La définition

Le théorème de Wold

Le principe de parcimonie

Le piège des racines communes

La famille au complet

Synthèse

Avant de commencer

Ce n'est pas un goût de la généralité

On pourrait toujours approcher une série par un AR d'ordre élevé, ou par un MA d'ordre élevé. Mais chaque retard ajouté coûte un paramètre à estimer, donc de la précision perdue. Le modèle mixte est une affaire d'**économie de paramètres**.

La définition

ARMA(p,q)

$$y_t = \theta + \sum_{i=1}^p \alpha_i y_{t-i} + \varepsilon_t + \sum_{j=1}^q \beta_j \varepsilon_{t-j}, \quad \boxed{\alpha_p \neq 0 \text{ et } \beta_q \neq 0}.$$

Une partie autorégressive, une partie moyenne mobile, un seul bruit blanc.

Écriture compacte

$$A(L) y_t = \theta + B(L) \varepsilon_t.$$

Deux polynômes, deux rôles

A gouverne la **stationnarité**; B gouverne l'**invertibilité**. Chacun garde exactement le rôle qu'il avait dans sa famille d'origine — c'est ce qui rend le cas mixte simple à traiter.

Le modèle mixte minimal

$$y_t = \theta + \alpha y_{t-1} + \varepsilon_t + \beta \varepsilon_{t-1}, \quad (1 - \alpha L)y_t = \theta + (1 + \beta L)\varepsilon_t.$$

Deux paramètres seulement, et pourtant : une mémoire infinie, dont le **premier** retard peut être ajusté indépendamment du rythme de décroissance.

Le théorème de Wold

La question se pose légitimement : en s'enfermant dans les modèles AR, MA et ARMA, ne se prive-t-on pas de tout ce que les séries stationnaires peuvent faire d'autre ?

Théorème de décomposition de Wold (1938)

Tout processus stationnaire au sens faible, centré et non déterministe, s'écrit de façon **unique**

$$y_t = \underbrace{\sum_{j \geq 0} \psi_j \varepsilon_{t-j}}_{\text{partie stochastique}} + \underbrace{\tilde{d}_t}_{\text{partie déterministe}},$$

avec $\psi_0 = 1$, $\sum_j \psi_j^2 < \infty$, $\varepsilon_t \sim \text{WN}(0, \sigma_\varepsilon^2)$, et d_t parfaitement prévisible à partir du passé lointain.

Ce qu'il affirme, et c'est considérable

Toute série stationnaire est un bruit blanc filtré. Pas « peut être approchée par », mais **est** — exactement, et d'une seule façon.

Le filtre du chapitre 3 n'était donc pas une hypothèse de travail commode : c'est la forme générale de la stationnarité.

Les ε_t ont un nom

Ce sont les **innovations** : ε_t est l'erreur de la meilleure prévision linéaire de y_t à partir de tout son passé. Ce qui n'était pas prévisible, et rien d'autre.

Trois conséquences pour la pratique

- La forme $MA(\infty)$ est **universelle** : chercher un modèle, c'est chercher les ψ_j .
- Un $MA(\infty)$ a une infinité de paramètres, donc n'est pas estimable tel quel. Un **ARMA(p,q)** en donne une approximation à quelques paramètres, aussi fine qu'on veut.
- Autrement dit : la classe ARMA est **dense** dans l'ensemble des processus stationnaires. On ne perd rien d'essentiel à s'y restreindre.

Ce que le théorème ne dit pas

Il porte sur la **prévision linéaire**, et sur elle seule. Les ε_t sont non corrélés — pas indépendants. Une dépendance non linéaire peut donc subsister intégralement dans la partie « imprévisible ».

Retenez cette réserve : c'est exactement l'espace où logeront les modèles GARCH, et le chapitre 15 y reviendra.

Le principe de parcimonie

Parcimonie

Entre deux modèles qui ajustent également bien les données, on retient celui qui compte le **moins de paramètres**.

Ce n'est pas une préférence esthétique

Chaque paramètre estimé consomme de l'information. Un modèle surchargé ajuste très bien l'échantillon — y compris son bruit — et prévoit mal hors échantillon. C'est le sur-ajustement, et c'est ce que les critères d'information du chapitre 13 pénaliseront explicitement.

Ce qu'un ARMA(1,1) économise

Considérons un ARMA(1,1) avec $\alpha = 0,8$. Sa représentation MA infinie a des poids qui décroissent en $0,8^j$.

Pour approcher cette mémoire par un MA pur avec une erreur inférieure à 1 %, il faudrait aller jusqu'à $q \simeq 20$ — donc estimer vingt paramètres au lieu de deux.

Le piège des racines communes

Redondance

Si $A(L)$ et $B(L)$ ont une racine commune, elle se simplifie et le modèle se ramène à un ARMA d'ordres inférieurs.

Un cas d'école

$$(1 - 0,5L)(1 - \alpha L)y_t = (1 - 0,5L)(1 + \beta L)\varepsilon_t$$

est un ARMA(2,2) en apparence. En simplifiant par $(1 - 0,5L)$, c'est un ARMA(1,1). Les deux écritures décrivent **le même processus**.

Le symptôme au moment de l'estimation

Le modèle n'est plus identifiable : une infinité de couples de paramètres donnent le même ajustement. L'algorithme d'estimation peine à converger, et les écarts-types des coefficients explosent.

La règle pratique

Si un $\text{ARMA}(2,2)$ donne des coefficients très mal estimés alors qu'un $\text{ARMA}(1,1)$ ajuste presque aussi bien, il faut soupçonner une racine commune — et retenir le modèle réduit. C'est le principe de parcimonie, sous sa forme la plus concrète.

La famille au complet

ARMA contient tout ce qui précède

Modèle	p	q	Écriture
Bruit blanc	0	0	$y_t = \varepsilon_t$
AR(p)	p	0	$A(L)y_t = \varepsilon_t$
MA(q)	0	q	$y_t = B(L)\varepsilon_t$
ARMA(p,q)	p	q	$A(L)y_t = B(L)\varepsilon_t$

ARIMA(p,d,q)

Lorsqu'il faut différencier d fois avant de modéliser :

$$A(L) (1 - L)^d y_t = \theta + B(L) \varepsilon_t.$$

Le « I » signifie *integrated* : la série d'origine est l'intégrale — la somme cumulée — d'un processus ARMA.

Le cas des rentabilités

Le log-prix est typiquement $I(1)$: il faut le différencier une fois. Modéliser la rentabilité par un ARMA(p,q) revient donc à modéliser le log-prix par un ARIMA(p,1,q). Les deux formulations sont équivalentes, et l'on travaillera sur la rentabilité.

Synthèse

L'essentiel du chapitre 7

- ARMA(p,q) : $A(L)y_t = \theta + B(L)\varepsilon_t$.
- A décide de la stationnarité, B de l'inversibilité.
- La **parcimonie** n'est pas esthétique : chaque paramètre superflu dégrade la prévision.
- Une racine commune à A et B rend le modèle non identifiable; il faut simplifier.
- Bruit blanc, AR et MA sont des cas particuliers d'ARMA; ARIMA ajoute la différenciation.

Vers la partie III

Nous avons les modèles. Il faut maintenant établir sous quelles conditions ils sont utilisables, puis calculer ce qu'ils prédisent — moments, autocovariances, autocorrélations. C'est ce qui rendra l'identification possible.