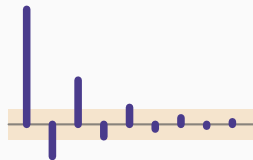


Modélisation ARMA des rentabilités financières

Chapitre 15 : La validation, et ce que les résidus révèlent



Avant de commencer

Les tests sur les coefficients

Les tests sur les résidus

Application aux rentabilités financières

Le fil conducteur, et la suite

Synthèse

Avant de commencer

Le critère, et il est unique

Un modèle ARMA est validé si ses **résidus sont un bruit blanc**. Tout ce qui reste structuré dans les résidus est une information que le modèle n'a pas su capter.

Deux séries de contrôles

Sur les **coefficients** : sont-ils significatifs, et le modèle est-il admissible ?

Sur les **résidus** : ne reste-t-il rien à expliquer ?

Les tests sur les coefficients

Test de Student

$$t = \frac{\hat{\alpha}_i}{\hat{\sigma}_{\hat{\alpha}_i}}, \quad H_0 : \alpha_i = 0.$$

On rejette au seuil de 5 % si $|t| > 1,96$.

Le cas du coefficient de rang le plus élevé

Si le dernier coefficient d'un AR(3) n'est pas significatif, l'ordre est trop grand : il faut passer à un AR(2). Les coefficients intermédiaires, en revanche, se conservent même non significatifs — on ne « trouve » pas un polynôme de retard sans raison.

Trois vérifications

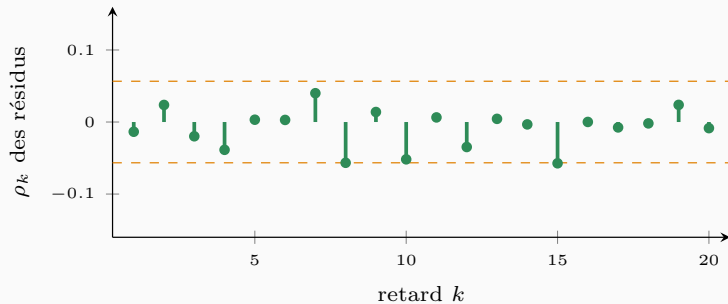
1. Racines de $\hat{A}$ hors du cercle unité : stationnarité.
2. Racines de $\hat{B}$ hors du cercle unité : inversibilité.
3. Aucune racine commune aux deux polynômes : identifiabilité.

Le symptôme d'un problème

Des écarts-types anormalement grands accompagnent presque toujours une racine commune ou une racine proche du cercle. Le remède est de réduire les ordres, jamais de forcer l'algorithme.

Les tests sur les résidus

L'autocorrélogramme des résidus



Toutes les autocorrélations dans les bandes : rien ne reste à expliquer linéairement.

Statistique de Box et Pierce (1970)

$$Q_{\text{BP}}(m) = T \sum_{k=1}^m \hat{\rho}_k^2 \sim \chi_{m-p-q}^2 \text{ sous } H_0.$$

H_0 : les m premières autocorrélations des résidus sont **toutes** nulles.

D'où sort cette forme

Sous H_0 , chaque $\hat{\rho}_k$ est approximativement $\mathcal{N}(0, 1/T)$, donc $T\hat{\rho}_k^2$ suit un χ_1^2 . En sommant m termes indépendants, on obtient un χ_m^2 – diminué de $p + q$ degrés de liberté parce que les résidus proviennent d'un modèle **estimé**, non des vrais chocs.

Statistique de Ljung et Box (1978)

$$Q_{\text{LB}}(m) = T(T+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{T-k} \sim \chi_{m-p-q}^2.$$

Ce que la correction change

$\hat{\rho}_k$ n'est calculée que sur $T - k$ paires : sa variance vaut en réalité $(T - k)^{-1}$ plutôt que T^{-1} .

Box-Pierce sous-pondère donc les retards élevés, et **rejette trop rarement** sur échantillon fini.

Le facteur $T(T+2)/(T-k)$ rétablit la pondération. L'écart s'annule quand T devient grand, mais il est sensible dès que m n'est pas petit devant T .

Lequel employer

Ljung-Box, sans hésiter : c'est le standard de tous les logiciels. Box-Pierce reste utile à connaître — c'est l'idée d'origine, et sa forme rend le raisonnement transparent.

Combien de retards tester

$m \approx \ln T$ selon Tsay, ou $m = 20$ par convention. Trop peu de retards : on manque une autocorrélation lointaine. Trop : la puissance du test se dilue.

Une précaution de lecture

On **ne rejette pas** H_0 quand le modèle est bon. C'est le cas rare d'un test où la non-significativité est le résultat recherché — donc où l'absence de puissance passe pour un succès. D'où l'intérêt de le compléter par l'examen visuel du corrélogramme.

Test de Jarque-Bera

$$JB = \frac{T}{6} \left(S^2 + \frac{(K - 3)^2}{4} \right) \sim \chi^2_2,$$

avec S l'asymétrie et K le kurtosis.

Sur données financières, ce test est toujours rejeté

Le kurtosis des rentabilités quotidiennes dépasse couramment 6, contre 3 pour une gaussienne. Ce rejet n'invalide pas le modèle : l'estimateur reste convergent (quasi-vraisemblance, chapitre 14). Il invalide les **intervalles de prévision** calculés sous hypothèse normale.

Une vérification simple et efficace

Estimer délibérément un modèle plus grand — ARMA(2,1) au lieu d'ARMA(1,1). Si les coefficients ajoutés ne sont pas significatifs et si les autres restent stables, le modèle initial est confirmé.

Application aux rentabilités financières

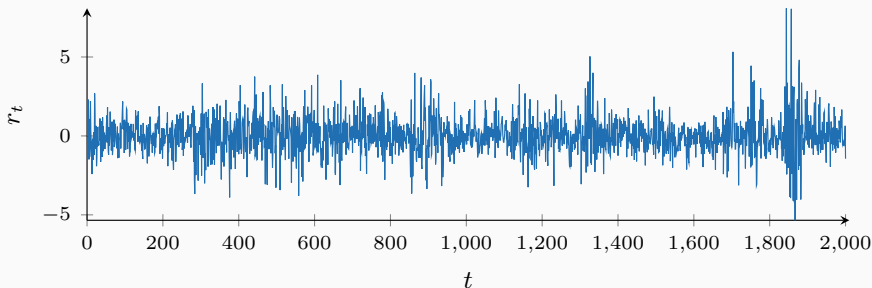
Le résultat, sur pratiquement toute série de rentabilités quotidiennes

- ACF et PACF des rentabilités : **presque entièrement dans les bandes.**
- Critères d'information : minimum atteint par un modèle très petit — bruit blanc, AR(1) ou MA(1).
- Coefficients estimés : de l'ordre de 0,02 à 0,05, parfois significatifs, jamais exploitables.
- Ljung-Box sur les résidus : non rejeté.

Comment lire ce résultat

Ce n'est pas un échec de la méthode : c'est une **confirmation empirique de l'efficience** sous sa forme faible. Le passé linéaire des rentabilités ne permet pas de prévoir leur avenir.

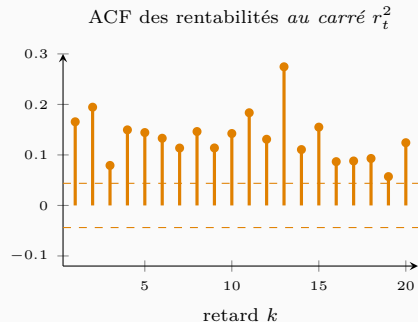
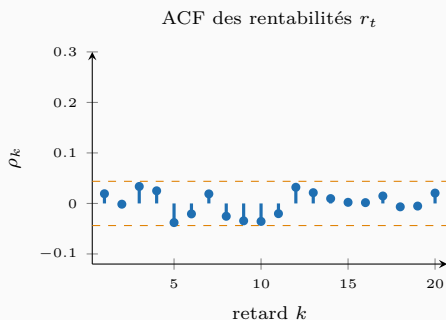
La trajectoire, pourtant, n'a rien d'un bruit blanc



Ce que l'œil voit et que l'ACF ne voit pas

Des périodes calmes, des périodes agitées, qui se succèdent par blocs. Il y a manifestement de la structure — mais elle ne porte pas sur le **signe** des rentabilités, elle porte sur leur **amplitude**.

La figure décisive



Le contraste

À gauche, l'ACF des rentabilités : nulle. À droite, l'ACF des rentabilités **au carré** : significative sur vingt retards et au-delà.

La conclusion du cours, en une phrase

Les rentabilités sont non corrélées, mais **pas indépendantes**. Elles sont un bruit blanc **faible** et non un bruit blanc **fort** — la distinction posée au chapitre 3, et dont on comprend maintenant qu'elle n'était

Test ARCH de Engle

Régresser $\hat{\varepsilon}_t^2$ sur ses m retards :

$$\hat{\varepsilon}_t^2 = \alpha_0 + \alpha_1 \hat{\varepsilon}_{t-1}^2 + \dots + \alpha_m \hat{\varepsilon}_{t-m}^2 + u_t.$$

Sous H_0 d'homoscédasticité conditionnelle, $T \cdot R^2 \sim \chi_m^2$.

Le résultat, sans exception

Sur données financières quotidiennes, ce test est rejeté massivement — souvent avec une p -valeur inférieure à 10^{-20} . L'hétéroscédasticité conditionnelle n'est pas une anomalie d'échantillon : c'est un fait établi.

Retour au chapitre 2

La condition (2) — variance constante — n'est pas violée au sens **inconditionnel** : la variance globale de la série existe et est finie. Elle l'est au sens **conditionnel** : la variance de demain, sachant le passé, change d'un jour à l'autre.

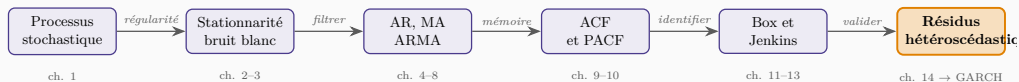
$$V(r_t) = \sigma^2 \text{ constante,} \quad V(r_t \mid \Omega_{t-1}) = h_t \text{ variable.}$$

Ce qu'il faut donc modéliser

Un ARMA modélise l'**espérance** conditionnelle. Il ne dit rien de la **variance** conditionnelle. Il faut un second modèle, pour h_t .

Le fil conducteur, et la suite

Le parcours complet



Modélisation GARCH

$$r_t = \mu_t + \varepsilon_t, \quad \varepsilon_t = \sqrt{h_t} z_t, \quad h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}.$$

L'équation de moyenne μ_t est l'ARMA de ce cours. L'équation de variance h_t est le cours suivant. Les deux se complètent : on ne remplace pas l'ARMA, on lui adjoint ce qui lui manque.

Ce que l'ACF des carrés annonçait

Une ACF des r_t^2 décroissant lentement est précisément la signature d'un GARCH(1,1) à forte persistance. La figure de ce chapitre n'est donc pas un constat d'échec : c'est le corrélogramme d'identification du cours suivant.

Synthèse

L'essentiel du chapitre 15

- Un modèle est validé si ses résidus sont un bruit blanc.
- Coefficients : test de Student, racines hors du cercle unité, absence de racine commune.
- Résidus : **Box-Pierce** et sa version corrigée **Ljung-Box** (H_0 que l'on souhaite ne pas rejeter, χ^2_{m-p-q}), Jarque-Bera, sur-ajustement.
- Les rentabilités n'ont **presque aucune mémoire linéaire** : l'efficiencia faible est confirmée.
- Mais l'ACF de leurs **carrés** est fortement significative : elles ne sont pas indépendantes.
- Bruit blanc faible, pas fort. La variance **conditionnelle** varie : c'est l'objet des modèles GARCH.

Le mot de la fin

Ce cours a construit tout l'appareil nécessaire pour modéliser la moyenne conditionnelle d'une série — et l'a appliqué à des données sur lesquelles cet appareil ne trouve presque rien. C'est un résultat scientifique de premier ordre : il dit où se trouve, et où ne se trouve pas, la prévisibilité des marchés.