Mabrouk Chetouane
Université Paris Dauphine
BFT - Investment Manager

Ecole Centrale de Paris - Supelec
Séries temporelles appliquées
Cours de Pascal Bondon
13 novembre 2020

Modèle $ARMA(p, d, q)$ et Méthode des moindres carrés ordinaires

Votre fichier devra s'intituler "NOM1 NOM2 - CENTRALE - 061120.(pdf)" et il en va de même pour l'objet du mail . La date limite d'envoi est le dimanche 22 novembre à 00h00. Veuillez soigner la présentation de vos résultats ainsi que la rédaction (en Tex uniquement...tout document word sera refusé). Il vous est également demandé de joindre vos code dans un fichier R. L'adresse pour l'envoi du document est mabrouk.chetouane@gmail.com

Exercice : modèle d'arbitrage

Cadre : Le Fed model est un modèle empirique que l'on peut classer dans la catégorie des modèles d'évaluation fondamentale. Le Fed model repose sur l'idée qu'il existe une relation d'arbitrage entre la rémunération d'une obligation d'état, jugée sans risque, et le taux de rendement d'une action plus risquée. Le prolongement de cette logique induit l'existence d'une relation d'équilibre entre les deux variables. L'équation ci-dessous permettrait de décrire cette relation :

$$\frac{E_t}{P_t} = \alpha + \beta . r_t \tag{1}$$

Le modèle économétrique est donné par la relation ci-dessous :

$$\frac{E_t}{P_t} = \alpha + \beta . r_t + \epsilon_t \tag{2}$$

où E_t désigne les earnings d'un indice ou d'une entreprise, P_t le prix de cette action ou de l'indice, r_t un taux sans risque et ϵ_t le terme d'erreur du modèle et on suppose $\epsilon_t \sim \mathcal{BB}\left(0, \sigma_\epsilon^2\right)$. On supposera des indices actions pour des raisons de simplicité (ici le S&P500) et nous considérons un taux de rendement d'une obligation d'état à 10 ans en guise de taux sans risque.

Partie 1 : Estimation du modèle

Question 1 - Calculer le earning yield du S&P500. Tracer les deux courbes en faisant apparaître les dates sur l'axe des abscisses. Tracer un nuage de points liant le earning yield au taux de rendement des obligations à 10 ans. L'ajustement linéaire est-il justifier ? Utiliser la commande `"abline"` pour tracer cet ajustement.

Question 2 - Rappeler le principe de la méthode des moindres carrés ordinaires. Rappeler les hypothèses sous-jacentes de l'estimateur des MCO.

Question 3 - Montrer que l'application des moindres carrés ordinaires à l'équation 2 permet de parvenir au résultat suivant:

$$\hat{\alpha} = \left(\frac{\bar{E}}{P}\right) - \hat{\beta}\bar{r} \tag{3}$$

$$\hat{\beta} = \frac{Cov\left(\frac{E}{P}, r\right)}{\sigma_r^2} \tag{4}$$

Montrer que $lim_{T\to\infty} V\left[\hat{\beta}\right] = 0$.

Question 4 - A l'aide la fonction `"lm"` , estimer par les MCO les coefficients de l'équation 2. Commenter vos résultats en particulier le signe du coefficient et sa significativité. Qu'indiquent les statistiques de Student et de Fisher ainsi que le coefficient de détermination ?

Question 5 - Réaliser une étude complète des résidus estimés : tracer leur densité, étudier leur normalité et vérifier l'existence/l'absence d'autocorrélation et d'hétéroscédasticité à l'aide des tests adéquats. Quel serait le modèle le plus adéquat si l'on venait à détecter de l'hétéroscédasticité dans les résidus. Détailler votre réponse.

Question 6 - L'observation des séries et du nuage de point laisse penser qu'une rupture s'est produite. Existe-t-il des tests qui permettraient d'indentifier des ruptures les plus probables?

A l'aide de la commande `"window "` découper votre dataset en deux parties : depuis 1961 jusqu'à décembre 2001 puis de janvier 2002 à mars 2019. Représenter côte-à-côte les nuages de points de ses deux sous-échantillons. Que constatez-vous? Estimer le modèle décrit par l'équation 2 pour chacune des sous-périodes. Commenter vos résultats. Existe-t-il des familles de modèles capables de prendre en charge ce type de configuration de données.

Partie 2: Estimation d'une nouvelle spécification et comparaison

Le *Fed model* comprend cependant une limite dans sa spécification. En effet, ses auteurs tentent d'expliquer le comportement d'une variable réelle E_t/P_t en utilisant un taux de rendement nominal. Pour corriger cette limite, on décide de calculer un taux d'intérêt réel rr_t en déflatant ce dernier de la croissance de l'indice des prix à la consommation (CPI) sur 12 mois soit

$$\frac{E_t}{P_t} = \alpha + \beta.\left(r_t - \pi_t\right) + \epsilon_t \tag{5}$$

$$\frac{E_t}{P_t} = \alpha + \beta.rr_t + \epsilon_t \tag{6}$$

$$rr_t = r_t - \pi \; avec \; \pi = \frac{CPI_t - CPI_{t-12}}{CPI_{t-12}}$$

Question 7 - Calculer le taux d'intérêt réel. Estimer par la méthode des moindres carrés ordinaires cette nouvelle spécification sur l'ensemble de la période d'une part puis sur les deux sous périodes d'autre part. Ce changement de spécification améliore-t-il le pouvoir explicatif du modèle notamment sur la seconde période? Justifier votre réponse.

Partie 3: Estimation d'une nouvelle spécification : modèle $ARMA(p, d, q)$

On propose de comparer ce modèle augmenté à une approche naïve basée sur un modèle $ARIMA(p, d, q)$ sur l'ensemble de période d'étude c'est à dire depuis 1961.

Question 8 - Présenter le modèle ARIMA et ses principales propriétés de manière précise et succinctes.

Question 9 - Identifier l'ordre du modèle ARIMA à l'aide de deux méthodes différentes (on privilégiera une approche parcimonieuse). L'introduction d'une partie intégrée dans votre modèle ARIMA (soit un modèle de type ARIMA) vous parait nécessaire. Tester la stationnarité de la série endogène à l'aide des tests d'ADF [1], PP et KPSS.

Question 10 - Estimer le modèle identifié à l'aide de la fonction `"auto.arima"` et vérifier la qualité de votre estimation.

Question 11 - Effectuer une prévision à l'aide de la commande `"predict"` sur horizon de trois périodes. Donner l'intervalle de confiance de votre prévision à 95% et tracer le sur le même graphique que votre prévision.

Une manière de sélectionner un modèle estimé consiste à évaluer sa capacité prédictive. On a alors recours à la Root Mean Square Error et de la Mean Absolute Error.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

Question 12 - Comparer les deux derniers modèles estimés en utilisant ces deux critères. Quel est celui qui affiche la meilleure performance?

Question 11 bis : - Une autre manière de comparer la qualité prédictive de deux modèles économétriques consiste à mettre en oeuvre le test de Diebold et Mariano (1995). On note $r^a_{i,t+h|t}$ la prévision du taux de rendement issue du Fed model (première version) et $r^b_{i,t+h|t}$ celle issue de la seconde spécification, pour $t = t_0, \ldots, T$. Dès lors on a

$$\epsilon^a_{i,t+h|t} = r_{i,t+h} - r^a_{i,t+h|t}$$

$$\epsilon^b_{i,t+h|t} = r_{i,t+h} - r^b_{i,t+h|t}$$

On définie une fonction de perte, notée $L\left(r_{i,t+h}, r^j_{i,t+h|t}\right) = L\left(\epsilon^j_{i,t+h|t}\right),\ avec\ j = a, b$. On retient habituellement la fonction de perte suivante $L\left(\epsilon^j_{i,t+h|t}\right) = \left(\epsilon^j_{i,t+h|t}\right)^2$. Pour savoir si un modèle est plus performant qu'un autre modèle, on teste alors l'hypothèse nulle suivante:

$$H_0:\ \mathbb{E}\left[L\left(\epsilon^a_{i,t+h|t}\right)\right] = \mathbb{E}\left[L\left(\epsilon^b_{i,t+h|t}\right)\right]$$

contre l'hypothèse alternative

1. Prenez soin de déterminer le bon PGD avant de réaliser les tests

$$H_0 : \mathbb{E}\left[L\left(\epsilon^a_{i,t+h|t}\right)\right] \neq \mathbb{E}\left[L\left(\epsilon^b_{i,t+h|t}\right)\right]$$

Le test de Diebold et Mariano (1995) est basé sur la différence des fonctions de perte. Formellement on a,

$$d_t = L\left(\epsilon^a_{i,t+h|t}\right) - L\left(\epsilon^b_{i,t+h|t}\right)$$

L'hypothèse nulle devient alors $H_0 = \mathbb{E}\left[d_t\right] = 0$. La statistique de Diebold et Mariano (1995) est donnée par

$$S_{DM} = \frac{\bar{d}}{\sqrt{\omega}} \sim \mathcal{N}(0,1) \text{ avec } \bar{d} = \frac{1}{T_0}\sum_{t=t_0}^{T} d_j$$

$$\text{avec } \hat{\omega} = \gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j \; et \; \gamma_j = cov(d_t, d_{t-j})$$

Diebold et Mariano (1995) montrent que sous l'hypothèse nulle (prévisions équivalentes) alors $S_{DM} \sim \mathcal{N}(0,1)$. On rejettera l'hypothèse nulle, au seuil de 5% si $|S_{DM}| > 1.96$.

Développer une routine sous R permettant de calculer la statistique de test puis de conclure sur la qualité prédictive du modèle.

Partie 4: Stabilité du modèle

On s'intéresse désormais à la stabilité du coefficient de l'équation du *FED model*. Plusieurs outils sont disponibles pour juger de la stabilité du modèle autrement dit des coefficients estimés :

1. sur une estimation récursive des coefficients de l'équation. Autrement dit, on estime le modèle en ajoutant à chaque nouvelle régression une observation supplémentaire.

2. sur une estimation glissante du paramètre où l'on sélectionne une période d'estimation que l'on décale d'une période.

Question 13 - A l'aide d'une routine que vous développez sous R, estimer les coefficients conformément à la méthode glissante. Tracer les courbes des coefficients β_i estimés ainsi que leur intervalle de confiance au seuil de 95% [2]. Commenter.

La seconde étape pour étudier la stabilité globale d'un modèle consiste à analyser les résidus issus des estimations récursives (cf. étape 1). Ces résidus sont ensuite utilisés pour calculer la statistique CUSUM qui permet de conclure en matière sur la stabilité du modèle estimé.

Question 14 - Présenter le test CUSUM. Implémenter le test de CUSUM [3] et commenter vos résultats.

Bon Courage

2. L'intervalle de confiance au niveau $\alpha = 0.95$ pour X_{t+h} peut être calculé en utilisant la fonction `"predict"` et l'option `"conf"` .

3. Pour aller plus vite, charger puis installer le package `strucchange`. Dans votre scripte, `install.packages("strucchange");` `library("strucchange")`. Le test CUSUM y est directement implémenté.