Mabrouk Chetouane
Université Paris Dauphine
BFT-Investment Manager

Ecole Centrale de Paris - Supelec
Séries temporelles appliquées
Cours de Pascal Bondon
Février 2017

Estimation paramétrique
Méthode des moindres carrés

Votre fichier devra s'intituler "NOM1 NOM2 - CENTRALE - 0217.(pdf)" et il en va de même pour l'objet du mail . La date limite d'envoi est le vendredi 17 février à 00h00. Veuillez soigner la présentation de vos résultats ainsi que la rédaction. Il vous est également demandé de joindre vos codes dans un document text ou un fichier R. Veuillez éviter les docx et privilégier les pdf. L'adresse pour l'envoi du document est mabrouk.chetouane@gmail.com

Exercice 1 : Le Fed model

Cadre : Le Fed model est un modèle empirique que l'on peut classer dans la catégorie de modèle d'arbitrage. Le Fed model repose l'idée qu'il existe une relation d'arbitrage entre la rémunération d'une obligation et le taux de rendement d'une action. Autrement dit, il existerait une relation d'équilibre entre les deux variables. L'équation ci-dessous décrit cette relation :

$$\frac{E_{i,t}}{P_{i,t}} = \alpha + \beta . r_{i,t} \tag{1}$$

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

$$\frac{E_{i,t}}{P_{i,t}} = \alpha + \beta . r_{i,t} + \epsilon_t \tag{2}$$

où $E_{i,t}$ désigne les earnings payés par un titre ou un indice, $P_{i,t}$ le prix de ce titre ou de cet indice, $r_{i,t}$ un taux sans risque et ϵ_t le terme d'erreur du modèle et on suppose $\epsilon_t \sim \mathcal{N}\left(\beta, \sigma_\epsilon^2\right)$. On supposera des indices actions pour des raisons de simplicité (ici le S&P500) et nous considerons un taux de rendement d'une obligation à 10 ans en guise de taux sans risque.

Partie 1 : Estimation du modèle

Question 1 - Calculer le earning yield de l'indice. Tracer les deux séries sur un même graphique (abscisses temporelles). Tracer un nuage de points liant le earning yield au taux sans risque. L'ajustement linéaire est-il justifier ? Tracer les quantiles de l'endogène puis sa densité empirique. Que peut-on dire sur la normalité de cette distribution. Est ce un obstacle quant à l'application de méthodes d'estimation linéaire ?

Question 2 - Rappeler le principe des MCO. Montrer formellement que l'application des moindres carrés ordinaires à l'équation 2 permet de parvenir au résultat suivant :

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} \tag{3}$$

$$\hat{\beta} = \frac{Cov\,(r_i, r_m)}{\sigma_m^2} \tag{4}$$

où y_t désigne la variable endogène centrée et x_t la variable exogène centrée. En déduire que le coefficient $\hat{\beta}$ peut s'écrire sous la forme suivante :

$$\hat{\beta} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2} \tag{5}$$

Question 3 - Rappeler les principales propriétés de l'estimateur des MCO. Démontrer que $lim_{T\to\infty} V\left[\hat{\beta}\right] = 0$.

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

$$R^2 = \frac{cov\,(X_t, Y_t)^2}{V(X_t)V(Y_t)} \tag{6}$$

Partie 2: Estimation d'une nouvelle spécification et comparaison

Le *Fed model* repose cependant sur une erreur de spécification. En effet, ses auteurs tentent d'expliquer le comportement d'une variable réelle $E_{i,t}/P_{i,t}$ en utilisant un taux de rendement nominal. Pour corriger cette limite, on décide de calculer un taux d'intérêt réel en déflatant ce dernier de la croissance de l'indice des prix à la consommation (CPI) sur sur 12 mois.

$$\frac{E_{i,t}}{P_{i,t}} = \alpha + \beta.\,(r_{i,t} - \pi_t) + \epsilon_t \tag{7}$$

$$\frac{E_{i,t}}{P_{i,t}} = \alpha + \beta.rr_{i,t} + \epsilon_t \tag{8}$$

Question 5 - Calculer le taux d'intérêt réel. Estimer par la méthode des moindres carrés ordinaires cette nouvelle spécification. Améliore-t-elle le pouvoir explicatif du modèle? Justifier votre réponse.

Question 6 - 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.

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

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

Comparer les deux modèles selon ces critères. Quel est celui qui affiche la meilleure performance?

Question 7 **Bonus** - 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

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

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 3: Stabilité du modèle

Question 8 - 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 coeffcients 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 des paramètre où l'on sélectionne une période d'estimaion que l'on décale d'une période.

8.a - 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% [1]. 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.

8.b - Présenter le test CUSUM. Implémenter le test de CUSUM [2] et commenter vos résultats .

Question 9 - La pertinence de la relation d'abritrage suppose que celle ci reste valable sur longue période. Une manière de procéder à ce diagnostique consiste à étudier le résidus du modèle soit la relation susivante:

$$\hat{\epsilon}_{i,t} = \frac{E_{i,t}}{P_{i,t}} - \hat{\beta}.rr_{i,t} \tag{9}$$

Tracer la série η_t. Utiliser les fonctions d'autocorrélation et d'autocorrélation partielle pour qualifier la stationnarité de cette séries.

Bon Courage

1. 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"` .
2. Pour aller plus vite, charger puis installer le package `strucchange`. Le test CUSUM y est directement implémenté.