Sujet 6 — Arbitrage statistique par cointégration multivariée (Johansen)

Principe général

Ce sujet est une alternative au Sujet 5 : au lieu de construire des résidus mean-reverting à partir de facteurs de risque estimés par PCA sur les *rendements*, on cherche directement des combinaisons linéaires stationnaires des *prix* (ou des logs-prix) d'un ensemble d'actifs, en s'appuyant sur la notion de **cointégration**.

Deux séries de prix non stationnaires (typiquement intégrées d'ordre 1, notées $I(1)$: leur niveau suit une marche aléatoire mais leurs différences premières sont stationnaires) sont dites cointégrées s'il existe une combinaison linéaire d'entre elles qui, elle, est stationnaire. Intuitivement, cela signifie que même si chaque actif pris isolément dérive sans limite, un certain écart entre eux (le *spread*) reste borné et revient vers une moyenne : c'est exactement la propriété recherchée pour une stratégie de retour à la moyenne.

Limite de l'approche bivariée (Engle-Granger)

Le test de cointégration le plus simple, celui d'Engle-Granger, ne s'applique qu'à une paire d'actifs à la fois : on régresse le prix d'un actif sur celui d'un autre, et on teste la stationnarité du résidu. Cette approche devient vite limitante dès qu'on s'intéresse à un panier de plus de deux actifs, car elle ne permet pas d'identifier plusieurs relations de cointégration simultanées, ni de traiter tous les actifs de façon symétrique (le choix de la variable « expliquée » dans la régression n'est pas neutre).

Le test de Johansen

Le test de Johansen généralise cette idée à un ensemble de N actifs simultanément. On modélise le vecteur des prix $Y_t = (y_{1,t}, \ldots, y_{N,t})^\top$ par un modèle vectoriel à correction d'erreur (VECM) :

$$\Delta Y_t = \Pi Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + \epsilon_t, \tag{1}$$

où Π est une matrice $N \times N$. Le résultat central de la théorie de la cointégration (théorème de représentation de Granger) est que le rang r de la matrice Π correspond exactement au nombre de relations de cointégration indépendantes présentes parmi les N actifs :

- si $r = 0$, il n'existe aucune relation de cointégration (les séries ne partagent pas de tendance commune stationnaire) ;

— si $r = N$, l'ensemble du système est déjà stationnaire ;
— si $0 < r < N$, on peut factoriser $\Pi = \alpha\beta^\top$, où les colonnes de β (matrice $N \times r$) donnent les r vecteurs de cointégration : chaque colonne β_k définit un portefeuille $\beta_k^\top Y_t$ qui est stationnaire, et donc mean-reverting.

Le test de Johansen fournit deux statistiques (le test de la trace et le test de la valeur propre maximale) permettant de déterminer le rang r à partir des données, ainsi que l'estimation des matrices α (vitesses d'ajustement) et β (poids du portefeuille cointégrant).

De la cointégration au signal de trading

Une fois un vecteur de cointégration β_k identifié, le portefeuille $Z_t = \beta_k^\top Y_t$ est par construction (approximativement) stationnaire : on peut alors lui appliquer exactement la même logique que pour les résidus $X_{i,t}$ du Sujet 5, à savoir modéliser Z_t comme un processus d'Ornstein-Uhlenbeck (ou son analogue discret AR(1)), en déduire la vitesse de retour à la moyenne, la demi-vie, et construire un s-score permettant de décider quand entrer et sortir de la position. Les poids β_k donnent directement les quantités relatives de chaque actif à détenir pour former ce portefeuille stationnaire.

Différences pratiques avec l'approche PCA du Sujet 5

L'approche par cointégration travaille directement sur les niveaux de prix plutôt que sur les rendements, et ne suppose pas explicitement un modèle à facteurs communs sous-jacent : elle recherche directement, de façon purement statistique, les combinaisons de prix qui sont stationnaires, quel que soit le nombre de facteurs de risque réellement à l'œuvre. En contrepartie, l'estimation du rang de cointégration et des vecteurs β est connue pour être sensible au nombre de retards choisi dans le VECM, à la taille de l'échantillon, et à la stabilité de la relation de cointégration dans le temps (une relation de cointégration peut se rompre structurellement, ce qui doit être surveillé en continu).

Remarque générale

Ce qui précède est une trame indicative. Il est attendu que l'équipe creuse en profondeur la théorie de la cointégration (choix du nombre de retards, tests de robustesse de la relation de cointégration, gestion du risque de rupture de régime) ainsi que les aspects pratiques d'implémentation du test de Johansen, bien au-delà de ce qui est esquissé ici.