

# QRT Challenge 2022

## 1 Contexte

Un problème classique de finance est la prédiction du rendement du jour suivant. Soit un marché contenant  $N$  actions avec un rendement  $R_t \in \mathbb{R}^N$  au temps  $t$ , l'objectif de ce problème est de modéliser au temps  $t$  le vecteur  $S_{t+1} \in \mathbb{R}^N$  avec les informations disponibles jusqu'au temps  $t$ . La façon la plus simple d'obtenir les estimations est de considérer un modèle linéaire de la forme

$$S_{t+1} = \sum_{\ell=1}^F \beta_{\ell} F_{t,\ell} \quad (1)$$

où  $F_{t,\ell} \in \mathbb{R}^N$  sont des facteurs explicatifs et  $\beta_1, \dots, \beta_F \in \mathbb{R}$  sont les paramètres du modèle.

Les facteurs explicatifs peuvent être des indicateurs techniques utilisés en finance comme le Momentum ou le RSI. Ils peuvent aussi être construits de manière mathématique de façon à être des combinaisons linéaires des rendements passés

$$F_{t,\ell} = \sum_{k=1}^D A_{k\ell} R_{t+1-k} \quad (2)$$

avec  $A_{\ell} = (A_{k\ell}) \in \mathbb{R}^D$  un vecteur quelconque et  $D$  la profondeur de temps souhaitée. Pour s'assurer de l'indépendance des facteurs explicatifs, **nous assumons que vecteurs  $A_{\ell}$  sont orthonormés**. Notre modèle paramétrique a pour paramètres :

- Une matrice  $A = [A_1, \dots, A_F] \in \mathbb{R}^{D \times F}$  dont les colonnes sont orthonormées.
- Un vecteur  $\beta = (\beta_1, \dots, \beta_F) \in \mathbb{R}^F$ .

## 2 Objectif

L'objectif de ce challenge est d'apprendre des facteurs explicatifs pour prédire le rendement d'un marché. Les participants ont à disposition 3 années de données historiques de 50 actions du même marché (training data) et doivent proposer les paramètres du modèle  $(A, \beta)$  comme réponse. Ensuite le modèle prédictif est testé en utilisant ces paramètres pour prédire le retour de 50 autres actions sur la même période de temps (test data).

**Nous autorisons  $D = 250$  jours pour la profondeur de temps et  $F = 10$  pour le nombre de facteurs.**

**Métrique** Soit  $\tilde{R}_t \in \mathbb{R}^{50}$  le rendement des 50 actions des données test sur une période de trois ans ( $t = 0, \dots, 753$ ) et  $\tilde{S}_t = \tilde{S}_t(A, \beta)$  les prédictions du modèle du participant. La métrique à maximiser est définie par

$$\text{Metric}(A, \beta) = \frac{1}{504} \sum_{t=250}^{753} \frac{\langle \tilde{S}_t, \tilde{R}_t \rangle}{\|\tilde{S}_t\| \|\tilde{R}_t\|} \quad (3)$$

Par construction, la métrique (*Cosine Similarity*) est comprise dans l'intervalle  $[-1, 1]$  et égale à  $-1$  s'il y a une non-orthonormalité des colonnes de  $A$  n'est pas respectée.

### Informations

- Place 43e/161 (score : 0.0948).
- Meilleur score : 0.1051.
- Score compris entre 0.09 et 0.1 : 83e à 6e.