

Identification de régimes de marché par un modèle de Markov caché multivarié : notice méthodologique

Introduction

Cette notice présente la méthodologie employée pour identifier des régimes de marché à partir de séries temporelles financières annuelles, notamment l'indice S&P 500 et les obligations d'État américaines (T. Bond). L'approche repose sur un *Hidden Markov Model* (HMM) multivarié appliqué à des rendements standardisés, permettant de détecter des régimes distincts par leurs caractéristiques statistiques.

L'objectif est de fournir une explication rigoureuse des notions mathématiques sous-jacentes au programme Python, ainsi que de justifier les choix de prétraitement et de modélisation.

1. Données et calcul des rendements

Les données disponibles comprennent notamment :

- Les prix de l'indice S&P 500 et ses dividendes annuels.
- Les taux des obligations d'État américaines (T. Bond) et autres obligations de référence.
- Des variables additionnelles : rendement immobilier, taux Aaa et Baa, dividend yield.

Le rendement total du S&P 500 à l'année t est calculé comme suit :

$$R_t^{\text{S&P}} = \frac{P_t - P_{t-1} + D_t}{P_{t-1}}, \quad (1)$$

où P_t est le prix de clôture annuel et D_t le dividende annuel.

Le rendement obligataire est directement fourni par les données (Return on Bond).

2. Standardisation des séries

Pour chaque série de rendement $R_t^{(i)}$, on effectue une standardisation :

$$\tilde{R}_t^{(i)} = \frac{R_t^{(i)} - \mu_i}{\sigma_i}, \quad (2)$$

avec μ_i et σ_i la moyenne et l'écart-type historiques de la série i .

Justification :

1. Les séries ont des échelles et des volatilités différentes. La standardisation permet d'éviter qu'une série domine l'apprentissage du HMM.
2. Le HMM multivarié suppose des distributions gaussiennes par état. La standardisation stabilise l'estimation des matrices de covariance.
3. Elle permet d'interpréter les états comme des écarts relatifs par rapport à la moyenne historique.

3. Modèle de Markov caché multivarié

On considère un HMM à K états cachés, $\{S_t\}_{t=1}^T$, régis par une chaîne de Markov homogène. La probabilité de transition entre états est donnée par la matrice $P \in \mathbb{R}^{K \times K}$:

$$\mathbb{P}(S_t = k | S_{t-1} = j) = P_{jk}, \quad \sum_{k=1}^K P_{jk} = 1. \quad (3)$$

Les observations standardisées $O_t = (\tilde{R}_t^{\text{S&P}}, \tilde{R}_t^{\text{Bond}})^\top$ suivent, conditionnellement à l'état $S_t = k$, une distribution normale multivariée :

$$O_t | S_t = k \sim \mathcal{N}(\mu_k, \Sigma_k), \quad (4)$$

où $\mu_k \in \mathbb{R}^2$ est le vecteur des moyennes et $\Sigma_k \in \mathbb{R}^{2 \times 2}$ la matrice de covariance associée à l'état k .

3.1 Estimation des paramètres

Les paramètres $\theta = \{\pi, P, \mu_k, \Sigma_k\}_{k=1}^K$ (probabilités initiales, matrice de transition, moyennes et covariances) sont estimés via l'algorithme EM (*Expectation-Maximization*) :

E-step : calculer les probabilités a posteriori des états :

$$\begin{aligned} \gamma_t(k) &= \mathbb{P}(S_t = k | O_1, \dots, O_T, \theta^{(\text{old})}) \\ \xi_t(j, k) &= \mathbb{P}(S_{t-1} = j, S_t = k | O_1, \dots, O_T, \theta^{(\text{old})}) \end{aligned}$$

M-step : mettre à jour les paramètres :

$$\pi_k = \gamma_1(k) \quad (5)$$

$$P_{jk} = \frac{\sum_{t=2}^T \xi_t(j, k)}{\sum_{t=2}^T \sum_{l=1}^K \xi_t(j, l)} \quad (6)$$

$$\mu_k = \frac{\sum_{t=1}^T \gamma_t(k) O_t}{\sum_{t=1}^T \gamma_t(k)} \quad (7)$$

$$\Sigma_k = \frac{\sum_{t=1}^T \gamma_t(k) (O_t - \mu_k)(O_t - \mu_k)^\top}{\sum_{t=1}^T \gamma_t(k)} \quad (8)$$

3.2 Décodage des régimes

Pour attribuer à chaque année un état spécifique, on utilise l'algorithme de Viterbi, qui maximise la probabilité conjointe des observations et des états :

$$\hat{S}_{1:T} = \arg \max_{S_{1:T}} \mathbb{P}(S_{1:T} | O_{1:T}, \hat{\theta}). \quad (9)$$

Le tableau présenté par le script contient exactement trois lignes (Régime 1/2/3) et les colonnes suivantes : **S&P 500**, **T.Bond**, **Aaa**, **Baa**, **Durée moyenne**. Chaque cellule des colonnes d'actifs contient la moyenne des rendements (standardisés dans la version fournie) pour les années assignées à ce régime.

Régime	S&P 500	T.Bond	Aaa	Baa	Durée moyenne
Régime 1
Régime 2
Régime 3

4. Justification des principaux choix de programmation

- **Standardisation** : justifiée pour les raisons numériques et d’interprétation déjà présentées ; dans le code, elle est implémentée via `StandardScaler` de `scikit-learn`.
- **HMM multivarié (covariance full)** : usage de covariance pleine (type ‘full’ dans `hmmlearn`) pour capturer les covariations croisées entre actions et différentes obligations.
- **Nombre d’états $K = 3$** : choix pragmatique visant un compromis entre parcimonie et richesse descriptive (haussier / neutre / crise). Il est recommandé d’évaluer des critères d’information (BIC/AIC) ou la validation croisée temporelle pour valider ce choix.
- **Visualisation + tableau** : combinaison graphique (zones colorées par régime) et tableau numérique (moyennes et durée moyenne) pour faciliter l’interprétation.

5. Analyse des résultats et validation historique

L’application du modèle HMM multivarié ($K = 3$) sur les rendements standardisés (S&P 500, T.Bond, Aaa, Baa) permet d’isoler trois régimes de marché distincts. La figure 1 présente la chronologie de ces régimes, superposée au rendement standardisé du S&P 500, ainsi que le tableau synthétique des caractéristiques statistiques de chaque état.

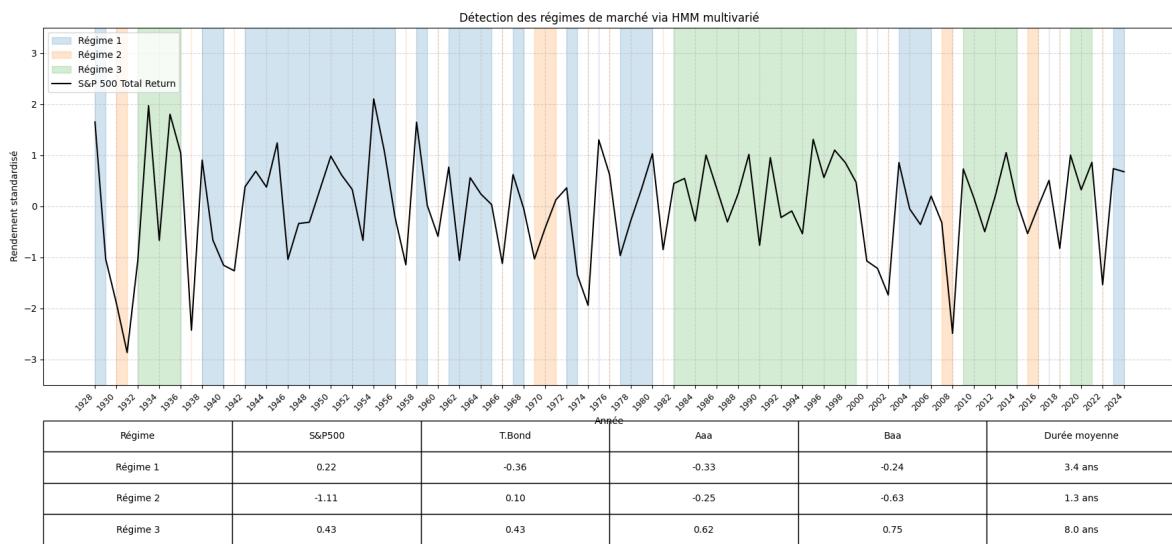


Figure 1: Détection des régimes de marché (HMM multivarié) et statistiques par état.

L’analyse combinée du graphique et du tableau permet de donner une interprétation économique robuste à chacun des trois états identifiés par l’algorithme.

Interprétation des signatures de régime

Les valeurs du tableau représentent les moyennes des rendements *standardisés* (écart-types par rapport à la moyenne historique) pour chaque actif, conditionnellement à un régime.

- **Régime 3 (Vert) : Le Boom "Risk-On" (Durée : 8.0 ans)**

Ce régime correspond à un **environnement macroéconomique porteur**, caractérisé par une forte confiance des investisseurs et un appétit marqué pour le risque. C’est la phase “Goldilocks” (croissance saine, inflation faible).

- *Signature statistique* : C'est un régime de "Everything Rally". Tous les actifs performent (S&P500 à +0.43, T.Bond à +0.43). L'appétit pour le risque est maximal : les obligations d'entreprises risquées (Baa à +0.75) surperforment largement les obligations d'État, indiquant une compression des "spreads" de crédit.
- *Cohérence historique* : Sa durée moyenne de 8 ans et sa position sur le graphique (Fig. 1) l'associent aux grands marchés haussiers : les années 1950, la "Grande Modération" (1982-1999) et la période de "Quantitative Easing" (2010-2017).

• Régime 2 (Orange) : La Crise "Risk-Off" (Durée : 1.3 ans)

Ce régime capture les **chocs financiers ou les récessions sévères**, provoquant une panique sur les marchés et une aversion totale pour le risque.

- *Signature statistique* : C'est le régime de crise classique. Les actions (S&P500 à -1.11) et les obligations d'entreprises (Baa à -0.63) s'effondrent. Les investisseurs se réfugient massivement vers l'actif sans risque (T.Bond à +0.10). C'est la "fuite vers la qualité" (flight-to-quality), provoquant un élargissement brutal des spreads de crédit.
- *Cohérence historique* : Sa durée courte (1.3 ans) et sa position confirment qu'il s'agit des chocs aigus : la Grande Dépression (1929-32), le choc pétrolier (1973-74), l'éclatement de la bulle Internet (2000-02), la crise financière (2008) et le choc inflationniste/taux de 2022.

• Régime 1 (Bleu) : Stagflation / Hausse des Taux (Durée : 3.4 ans)

Ce régime est caractéristique d'un **conflit entre l'inflation et la croissance**, souvent marqué par un resserrement monétaire qui pèse sur les actifs à revenu fixe.

- *Signature statistique* : C'est un régime où la corrélation actions-obligations devient positive. Les actions résistent (S&P500 à +0.22), mais toutes les classes d'actifs obligataires souffrent (T.Bond à -0.36), l'inflation érodant leur valeur réelle.
- *Cohérence historique* : Cette signature correspond parfaitement à un environnement de stagflation ou de hausse des taux. Le modèle identifie brillamment la période du "Choc Volcker" (fin 1970 - début 1980) comme appartenant à ce régime, ainsi que les périodes de normalisation monétaire d'après-guerre (années 1940).

Conclusion de l'analyse

La cohérence entre les signatures statistiques des trois régimes et leur alignement chronologique avec les événements majeurs de l'histoire économique (Grande Dépression, Choc Volcker, Crise de 2008) valide la pertinence du modèle HMM multivarié. L'algorithme a réussi à "désapprendre" et à classifier les grands "méta-régimes" qui dictent la performance et la corrélation des actifs financiers.

Conclusion

Cette notice fournit une base mathématique rigoureuse et une justification des choix de prétraitement et de modélisation. Le code Python associé (fonctionnant avec `hmmlearn`, `scikit-learn` et `matplotlib`) doit respecter les mêmes conventions (standardisation, gestion des NaN, index temporel) pour produire les sorties décrites.