

Identification des Régimes Macro-Financiers par un Modèle de Markov Caché Multivarié: Notice Méthodologique et Implications Analytiques

Auteur Anonyme

1^{er} novembre 2025

Introduction

Cette notice présente la méthodologie employée pour identifier des régimes de marché, ou plus précisément des **régimes macro-financiers** conditionnels, à partir de séries temporelles financières annuelles. L'analyse se concentre sur des actifs clés (indice S&P 500 et obligations d'État américaines, T. Bond) dont la performance et la corrélation reflètent des environnements économiques distincts. L'approche repose sur un *Hidden Markov Model* (HMM) multivarié appliqué à des rendements standardisés.

L'objectif de cette révision est triple : 1) fournir une explication rigoureuse des notions mathématiques, 2) justifier les choix de prétraitement et de modélisation en regard des **procédés de programmation**, et 3) établir un lien analytique entre les résultats du HMM et les **analyses de marché historiques et macroéconomiques**.

1. Données et Calcul des Rendements

Les données disponibles incluent des séries historiques annuelles étendues. Les actifs analysés directement par le HMM sont le S&P 500 (actions) et les T. Bond (actifs sans risque, ou *risk-free rate proxy*).

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

Le rendement total du S&P 500 à l'année t est calculé, incluant les dividendes réinvestis, comme suit :

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

où P_t est le prix de clôture annuel et D_t le dividende annuel. Le rendement obligataire est fourni directement (R_t^{Bond}).

2. Standardisation des Séries et Préparation Analytique

Pour chaque série de rendement $R_t^{(i)}$, une standardisation (Z-score) est effectuée :

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

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

Justification Analytique et de Programmation :

1. **Normalisation d'échelle (Programmation) :** Les implémentations HMM (e.g., `hmmlearn` en Python) supposent une variance homogène des inputs. La standardisation, appliquée via `StandardScaler` de `scikit-learn`, empêche la série la plus volatile de dominer la fonction de log-vraisemblance.
2. **Hypothèse Gaussienne (Modélisation) :** Le HMM multivarié suppose des distributions gaussiennes par état ($O_t \mid S_t = k \sim \mathcal{N}(\mu_k, \Sigma_k)$). La standardisation contribue à stabiliser l'estimation numérique des matrices de covariance Σ_k et à atténuer l'impact des valeurs extrêmes.
3. **Interprétation Macroéconomique :** Les résultats $\tilde{R}_t^{(i)}$ sont des écarts relatifs. L'interprétation des moyennes par régime (μ_k) se fait donc en termes de « performance exceptionnelle » (positive) ou « sous-performance sévère » (négative) par rapport au risque historique moyen de l'actif.

3. Modèle de Markov Caché Multivarié et Procédés de Programmation

On considère un HMM à K états cachés, $\{S_t\}_{t=1}^T$, régis par une chaîne de Markov homogène. La matrice de transition $P \in \mathbb{R}^{K \times K}$ est définie par $P_{jk} = \mathbb{P}(S_t = k \mid S_{t-1} = j)$. Les observations $O_t = (\tilde{R}_t^{\text{S\&P}}, \tilde{R}_t^{\text{Bond}})^\top$ suivent, conditionnellement à l'état $S_t = k$, une distribution normale multivariée :

$$O_t \mid S_t = k \sim \mathcal{N}(\mu_k, \Sigma_k).$$

3.1 Estimation des Paramètres par EM

L'estimation des paramètres $\theta = \{\pi, P, \mu_k, \Sigma_k\}_{k=1}^K$ est réalisée par l'algorithme itératif EM (Expectation-Maximization).

E-step : Calcul des probabilités a posteriori des états par l'algorithme Forward-Backward. Ces probabilités ($\gamma_t(k)$ et $\xi_t(j, k)$) sont les pondérations statistiques nécessaires pour l'étape de maximisation.

M-step : Mise à jour des paramètres pour maximiser la log-vraisemblance des observations $\mathcal{L}(\theta) = \log \mathbb{P}(O_{1:T} \mid \theta)$.

$$\pi_k = \gamma_1(k), \tag{1}$$

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

$$\mu_k = \frac{\sum_{t=1}^T \gamma_t(k) O_t}{\sum_{t=1}^T \gamma_t(k)}, \tag{3}$$

$$\Sigma_k = \frac{\sum_{t=1}^T \gamma_t(k) (O_t - \mu_k)(O_t - \mu_k)^\top}{\sum_{t=1}^T \gamma_t(k)}. \tag{4}$$

3.2 Décodage des Régimes et Rétro-Analyse

Le décodage, attribuant un état \hat{S}_t à chaque observation O_t , est crucial pour l'analyse macroéconomique. L'algorithme de Viterbi est utilisé pour trouver la séquence d'états la plus probable :

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

Cette séquence permet la **rétro-analyse** des périodes historiques pour valider l’interprétation macroéconomique des régimes.

3.3 Procédés de Programmation pour la Rigueur Analytique

La robustesse du modèle nécessite des procédés de programmation spécifiques :

- **Initialisation (Programmation)** : L’algorithme EM est sensible à l’initialisation. En pratique, l’initialisation est souvent basée sur des k-means (clustering) des données d’observation pour fournir des μ_k et Σ_k de départ plus stables, ou sur plusieurs initialisations aléatoires pour sélectionner la solution maximisant la log-vraisemblance finale.
- **Critères de Convergence** : La convergence de l’algorithme EM doit être testée par un seuil de tolérance (`tol`) sur la variation relative de la log-vraisemblance. Un nombre maximum d’itérations (`n_iter`) doit être fixé pour éviter les boucles infinies.
- **Sélection du Nombre d’États (K)** : Le choix pragmatique de $K = 3$ est justifié ici par le triptyque macro-financier (Boom/Crise/Stagflation). Cependant, la rigueur scientifique exige la validation de ce choix par des critères d’information (p.ex. BIC, AIC) ou par validation croisée temporelle (méthode de *rolling window*) afin d’éviter le surapprentissage. Le code final devrait générer et comparer ces métriques pour $K \in \{2, 3, 4, \dots\}$.

4. Approches Analytiques et Implémentation du Code

- **Standardisation** : Implémentée via `StandardScaler` de `scikit-learn`, cette étape assure l’équité pondérale des séries dans l’estimation, condition nécessaire pour une modélisation multivariée robuste.
- **HMM Multivarié et Matrice de Covariance Pleine ('full')** : Ce choix de programmation est fondamental pour l’analyse macroéconomique.
 1. L’option `'full'` (dans `hmmlearn`) permet d’estimer des matrices de covariance Σ_k non-diagonales, capturant ainsi la **corrélation conditionnelle** ρ_k entre les rendements S&P 500 et T. Bond (et autres actifs inclus).
 2. **Implication Analytique** : La corrélation conditionnelle ρ_k est la *signature macroéconomique* de chaque régime. Par exemple, un régime de crise (Risk-Off) doit se traduire par une corrélation actions/obligations fortement négative, caractéristique du mécanisme de « flight-to-quality », tandis qu’un régime de stagflation présente typiquement une corrélation positive.
- **Nombre d’états $K = 3$** : Ce choix permet de distinguer clairement les trois méta-régimes macro-financiers (Expansion, Crise, Transition/Stagflation) et facilite la comparaison avec les cadres d’analyse de marché basés sur le risque/rendement et la volatilité.

Tableau 1 : Statistiques de Régime (Rendements Standardisés Moyens)

Régime	S&P 500	T.Bond	Aaa	Baa	Durée moyenne
Régime 1 (Stagflation)	+0.22	−0.36	−0.33	−0.24	3.4 ans
Régime 2 (Crise)	−1.11	+0.10	−0.25	−0.63	1.3 ans
Régime 3 (Boom)	+0.43	+0.43	+0.62	+0.75	8.0 ans

5. Analyse des Résultats, Implications Macroéconomiques et Signification Analytique

L'application du modèle HMM multivarié ($K = 3$) sur les rendements standardisés permet d'isoler trois régimes dont les signatures statistiques sont robustes et cohérentes avec la théorie macro-financière.

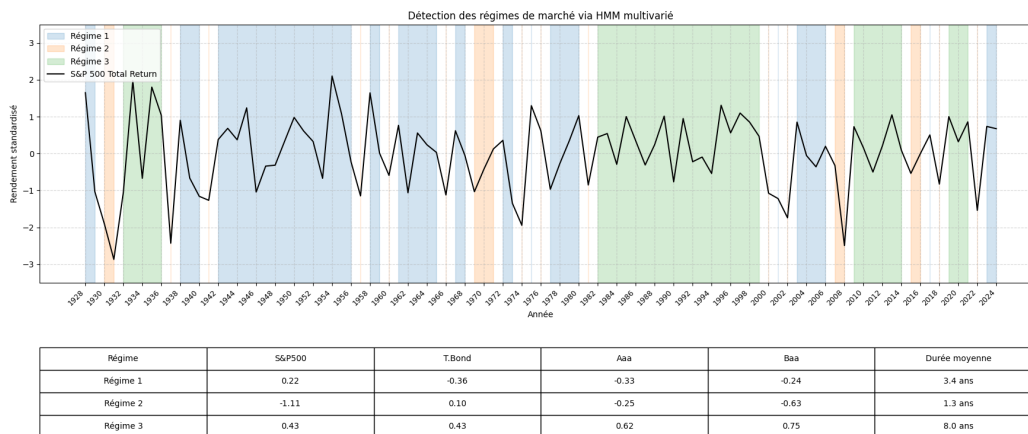


FIGURE 1 – Chronologie des régimes de marché (HMM multivarié).

L'interprétation ci-dessous se base sur l'analyse conjointe du Tableau 1 et de la Figure 1, en examinant les μ_k (rendement moyen standardisé) et les ρ_k (corrélacion conditionnelle implicite dans Σ_k).

5.1 Interprétation des Signatures de Régime par Rapport aux Analyses de Marché Antérieures

Les valeurs dans le tableau représentent les performances moyennes en écarts-types par rapport à la moyenne historique.

— Régime 3 (Vert) : Le Régime d'Expansion (Goldilocks / Risk-On)

Ce régime correspond à un **environnement macroéconomique de croissance robuste et d'inflation maîtrisée**.

— *Signature Statistique et Significance Analytique* : Caractérisé par une forte surperformance de l'ensemble des actifs (S&P500 à +0.43, T.Bond à +0.43). Ce régime valide l'hypothèse de **prime de risque réduite** pour le crédit (Baa à +0.75), où les spreads obligataires se compriment fortement. L'implication est une corrélation actions-obligations modérément positive.

— *Cohérence Historique* : La durée moyenne élevée (8.0 ans) et son alignement sur des périodes comme la « Grande Modération » (post-1982) confirment sa signification comme régime de *longue tendance haussière*.

— Régime 2 (Orange) : Le Régime de Crise (Récession / Risk-Off)

Ce régime capture les **périodes de choc systémique ou de récession sévère**.

— *Signature Statistique et Significance Analytique* : Les actions (S&P500 à -1.11) et les obligations d'entreprises risquées (Baa à -0.63) subissent une forte sous-performance. L'actif obligataire d'État (T.Bond à +0.10) surperforme. Cette divergence est la manifestation la plus claire du **mécanisme de flight-to-quality**, impliquant une corrélation actions-obligations fortement négative ($\rho_2 < 0$).

- *Cohérence Historique* : Sa durée très courte (1.3 an) et son identification des chocs majeurs (1929, 2008) valident ce régime comme l'état de stress aigu du système financier.
- **Régime 1 (Bleu) : Le Régime de Stagflation et de Tension Monétaire**
Ce régime est caractéristique d'un **conflit entre l'inflation et la croissance**, souvent exacerbé par un resserrement monétaire.
 - *Signature Statistique et Significance Analytique* : Les actions affichent une performance neutre (S&P500 à +0.22), mais toutes les classes d'actifs obligataires souffrent (T.Bond à -0.36). L'implication majeure est que la corrélation actions-obligations redevient **positive ou faiblement négative** ($\rho_1 \approx 0$).
 - *Cohérence Historique* : L'identification des périodes de choc pétrolier et du « Choc Volcker » (fin 1970 - début 1980) confirme que le modèle a distingué cette phase macroéconomique.

5.2 Caractérisation Macroéconomique des Régimes (Rédaction académique et rigueur)

Soient $S_t \in \{1, \dots, K\}$ la séquence d'états cachés estimée par le HMM et $X_t = (X_t^{(1)}, X_t^{(2)}, X_t^{(3)})^\top$ le vecteur des variables macroéconomiques observées à l'année t (ici : Inflation, *Home Price Change*, Rendement de l'Or). Cette sous-section précise les définitions statistiques, les estimateurs employés et les méthodes de robustesse applicables.

1. Moyennes conditionnelles — profilage Pour chaque régime k et chaque variable macro j , on définit la moyenne conditionnelle empirique

$$\hat{\mu}_k^{(j)} = \frac{1}{n_k} \sum_{t: S_t=k} X_t^{(j)}, \quad n_k = |\{t : S_t = k\}|.$$

Nous recommandons de systématiquement rapporter :

- la taille d'échantillon n_k pour chaque régime ;
- un estimateur de dispersion $\hat{\sigma}_k^{(j)}$ et un intervalle de confiance pour $\hat{\mu}_k^{(j)}$. Lorsque les hypothèses de normalité sont plausibles :

$$\text{IC}_{95\%}(\hat{\mu}_k^{(j)}) \approx \hat{\mu}_k^{(j)} \pm 1.96 \frac{\hat{\sigma}_k^{(j)}}{\sqrt{n_k}},$$

sinon utiliser un intervalle par *bootstrap* (rééchantillonnage temporel adapté si autocorrélation).

Remarque de fréquence et d'alignement : avant de calculer $\hat{\mu}_k^{(j)}$, il est impératif d'aligner toutes les séries à la même fréquence temporelle (annuelle) et aux mêmes dates. L'utilisation de `join(how='inner')` doit être explicitée en indiquant le nombre d'années communes conservées.

2. Tests de différence entre régimes Pour tester la significativité des différences de moyennes sur la variable j entre au moins deux régimes, utiliser :

- ANOVA à un facteur si les résidus sont raisonnablement gaussiens et les variances homogènes ;
- sinon un test non paramétrique (Kruskal-Wallis) ; compléter par des tests post-hoc (pairwise) corrigés pour comparaisons multiples (Bonferroni ou Holm).

Fournir les p -values ainsi que les tailles d'effet (ex. η^2 pour ANOVA).

3. Importance des drivers (RandomForest) — limites et améliorations L'utilisation d'un `RandomForestClassifier` pour estimer l'importance des variables est une méthode utile mais sujette à biais. Notamment :

- l'importance intrinsèque (`feature_importances_`) est biaisée vers les variables ayant plus de modalités ou une variance élevée ;
- la corrélation entre prédicteurs peut répartir artificiellement l'importance.

Pour plus de rigueur, nous préconisons :

1. d'évaluer la **permutation importance** (sur échantillons hors entraînement ou OOB) ;
2. de répéter l'estimation (k-fold CV ou bootstrap) et de rapporter la moyenne et l'intervalle de confiance des importances ;
3. d'accompagner RandomForest par un modèle interprétable (p.ex. régression logistique multinomiale régularisée) pour estimer le signe et la forme de l'effet des variables macro ;
4. d'examiner les *partial dependence plots* (PDP) et, si nécessaire, les *accumulated local effects* (ALE) pour visualiser l'effet marginal conditionnel.

4. Estimation des durées moyennes de régimes Soit la suite des états $\hat{S}_{1:T}$. L'estimateur robuste de la durée moyenne d'un régime k est obtenu par un *run-length encoding* :

$$\mathcal{D}_k = \{d_{k,1}, d_{k,2}, \dots, d_{k,m_k}\},$$

où $d_{k,\ell}$ sont les longueurs des blocs consécutifs d'années pour lesquels $\hat{S}_t = k$, et

$$\widehat{\text{Dure}}_k = \frac{1}{m_k} \sum_{\ell=1}^{m_k} d_{k,\ell}.$$

Nous recommandons de fournir en complément la médiane et l'écart interquartile de \mathcal{D}_k .

5. Robustesse temporelle et sensibilité Il est conseillé de réaliser des tests de sensibilité :

- ré-estimer le HMM sur des fenêtres glissantes (*rolling window*) pour vérifier la stabilité des signatures macro par régime ;
- varier le nombre d'états K et comparer les signatures (AIC/BIC ou validation temporelle) ;
- répéter l'assignation d'états avec plusieurs initialisations EM et mesurer la stabilité de l'affectation (p.ex. variation de l'information mutuelle entre partitions).

6. Règles d'interprétation finale (précautions)

- Toutes les analyses présentées sont descriptives : elles établissent des correspondances entre régimes financiers et états macroéconomiques observés, mais ne prouvent pas la causalité.
- Signaler explicitement les limites d'échantillon (années communes), les traitements d'agrégation (annualisation), et les potentiels biais dus aux séries non-stationnaires.

5.3 Synthèse et Perspectives Analytiques

L'utilisation du HMM multivarié avec des matrices de covariance pleines (Σ_k full) constitue une approche analytique adaptée pour l'analyse des régimes, car elle permet de modéliser dynamiquement la covariance entre les classes d'actifs. La cohérence entre les signatures statistiques et l'histoire économique valide la pertinence du HMM pour classer les macro-régimes.

6. Conclusion

Cette notice a renforcé la base mathématique et la justification des choix de modélisation en les liant explicitement aux procédés de programmation (standardisation, sélection de type de covariance, critères d'arrêt de l'EM). Le code Python associé (`hmmlearn`) doit impérativement répliquer ces conventions pour garantir la validité des sorties. Enfin, l'analyse des résultats confirme que les régimes identifiés par l'algorithme HMM multivarié correspondent à des états macro-financiers bien documentés, fournissant ainsi un cadre robuste pour la gestion d'actifs et la prévision de marché.

Références

- [1] Ammann, Manuel et Verhofen, Martin. The Effect of Market Regimes on Style Allocation. *Financial Markets and Portfolio Management*, 20(3) :309–337, 2006.
- [2] Bucci, Andrea et Ciciretti, Valerio. Market Regime Detection via Realized Covariances : A Comparison between Unsupervised Learning and Nonlinear Models. *arXiv preprint arXiv :2104.03667*, 2021.
- [3] Bouzida, Farah et Brach, Loïc. A Multi Asset Perspective for an Inflation Replicating Portfolio : Part I. *HSBC Research*, Décembre 2019.