



UNIVERSITÉ DE MONTRÉAL

Faculté des Arts et des Sciences

Rapport de Travail Dirigé

PRÉVISION DES RENDEMENTS DES ACTIFS FINANCIERS EN UTILISANT LES RÉGIMES ÉCONOMIQUES

Rédigé par :

OTMAN CHERRATI

MATRICULE - 20233031

Encadreur académique :

RÉNÉ GARCIA

Professeur titulaire

Université de Montréal

Juin 2024

Résumé

Ce projet se concentre sur l'analyse et la prévision des rendements des actifs financiers en combinant les caractéristiques des actifs avec les régimes économiques obtenus à partir des données macro-économiques. L'idée de ce projet est inspirée des travaux de Chen et al. et Gu et al., qui ont démontré que l'information macro-économique est indispensable dans la prévision des rendements. En conséquence, on a souhaité créer un modèle plus simple qui prend en compte cette information macro-économique. La méthode proposée combine l'analyse en composantes principales (PCA) et le filtre de Hamilton pour identifier les probabilités des régimes économiques. Ces probabilités sont ensuite intégrées avec les caractéristiques des actifs financiers pour modéliser et prédire les rendements.

Remerciements

Je tiens à exprimer ma profonde gratitude à toutes les personnes qui ont contribué à la réalisation de ce projet. Ce travail n'aurait pas été possible sans le soutien et l'encouragement de plusieurs individus précieux.

Tout d'abord, je souhaite remercier mon professeur encadreur, **M. René Garcia**, pour son enseignement, son encadrement, sa patience et ses conseils avisés. Votre expertise a été un pilier essentiel dans l'accomplissement de ce projet. Votre confiance en mes capacités m'a permis de repousser mes limites et d'atteindre mes objectifs.

Je tiens également à exprimer ma reconnaissance à mes professeurs, dont les enseignements ont largement contribué à enrichir mes connaissances et compétences.

Un remerciement spécial va à ma famille pour leur soutien tout au long de ce parcours. Parents, frères et sœurs pour leur patience, leurs encouragements et leur amour inconditionnel.

Je souhaite également remercier mes amis et collègues pour leur soutien moral, leurs conseils et leurs encouragements. Votre présence à mes côtés a été une grande source de motivation et de réconfort.

Enfin, je n'oublie pas toute personne qui, de près ou de loin, a contribué à la réalisation de ce projet. Que ce soit par des discussions constructives, des conseils pratiques ou simplement par leur présence bienveillante, chacune de vos contributions a été précieuse.

À toutes et à tous, merci du fond du cœur.

Otman

Table des matières

1	Revue de la Littérature	7
1.1	Prévision des Rendements des Actifs	7
1.2	Régimes économiques	8
1.3	Filtre de Kalman	9
2	Présentation des données	10
2.1	Caractéristiques des Actifs Financiers	10
2.2	Données Macro-économiques	11
3	PCA et filtre de Hamilton : Donnée macro	12
3.1	Analyse en Composantes Principales (PCA)	12
3.1.1	Étapes de la PCA	12
3.1.2	Résultats de la PCA	13
3.2	Modèle de changement de Régimes	14
3.2.1	Représentation État-Espace	14
3.2.2	Filtre de Kalman	15
3.2.3	Changement de régimes de Hamilton	16
4	Prédiction des Rendements des Actifs	20
4.1	Introduction aux Réseaux de Neurones	20
4.2	Structure et Fonctionnement d'un Réseau de Neurones	20
4.2.1	Couche d'entrée	21
4.2.2	Couches Cachées	21
4.2.3	Couche de Sortie	22
4.3	Entraînement du Modèle	22

4.3.1	Propagation Avant (Forward Propagation)	22
4.3.2	Calcul de la Perte (Loss)	22
4.3.3	Rétro propagation (Backpropagation)	23
4.3.4	Itérations et Convergence	24
4.4	Implémentation	24
4.4.1	Préparation des Données	24
4.4.2	Construction du Modèle	25
4.4.3	Entraînement du Modèle	25
4.4.4	Évaluation et Prédiction	26
5	Résultats	27
5.1	Probabilités de Régimes Économiques	27
5.2	Performance du Modèle de Prédiction	31
6	Discussion	35
6.1	Implications des Résultats	35
6.2	Limitations de l'étude	35
6.3	Suggestions	35
	Références	37

Liste des figures

4.1	Exemple de réseau à 2 couches cachées de 4 noeuds chacune (pour régression)	21
5.1	Probabilités du régime 1 et récessions NBER, modèle sans auto-régression . .	29
5.2	Probabilités du régime 1 et récessions NBER, modèle AR(1)	30
5.3	Ratios de Sharpe portefeuille Long Court (Hors Ech)	31
5.4	Rendement cumulatif de portefeuilles Long-court	32
5.5	Rendements cumulatifs de portefeuilles de déciles classé par prédiction de rendement	34

Introduction

La prédiction des rendements constitue un défi majeur pour les investisseurs et les analystes financiers, car elle leur permet de prendre des décisions éclairées sur la gestion de leur portefeuille et d'anticiper les mouvements du marché. Pour les investisseurs, il est crucial de pouvoir estimer avec précision les rendements futurs des actifs financiers afin de maximiser les gains potentiels tout en minimisant les risques associés.

Dans cette quête de prévision des rendements, une gamme de modèles théoriques ont vu le jour. Le CAPM est l'un des premiers modèles, il essaie d'évaluer le rendement d'un actif en fonction de son risque systématique, mesuré par le bêta. Plus tard Fama et French proposent les modèles à plusieurs facteurs. Chacun de ces modèles cherchent à expliquer les variations des rendements en fonction de différents facteurs économiques et financiers, mais tous ont des limites à prédire avec précision les mouvements du marché.

L'émergence des modèles d'apprentissage machine a ouvert de nouvelles possibilités dans le domaine de la prédiction des rendements. Ces modèles se basent sur des algorithmes qui exploitant la puissance de calcul des machines pour découvrir des schémas complexes dans les données. Les études ont pu montrer qu'ils sont plus précis, même si leur problème d'interprétabilité persiste.

En effet, des recherches telles que celles de Gu et al.(2020) et Chen et al.(2023), ont pu montré une amélioration significative par rapport aux méthodes traditionnelles de prévision des rendements.

Dans cette étude, on propose une approche pour estimer les rendements des actifs financiers en intégrant les probabilités de régimes économiques avec les caractéristiques des actifs

dans un modèle de réseau de neurones.

On utilise le filtre de Hamilton pour extraire les régimes économiques à partir des données macro-économiques historiques, puis on combine ces informations avec les caractéristiques des actifs pour obtenir une matrices enrichie de données. Cette méthodologie permet de capturer les variations des rendements des actifs dans différents régimes économiques.

D’abord, les données macro-économiques sont classées en groupes et une analyse en composantes principales est effectuée pour chaque groupe. La première composante principale (PC1) de chaque groupe, représentant la majorité de la variance, est extraite et utilisée pour capturer les dynamiques macro-économiques pertinentes.

Ensuite, les séries temporelles de PC1 sont passer à travers le filtre de Hamilton pour identifier les régimes économiques et calculer les probabilités associées à chaque régime.

Enfin, un modèle de régression (Réseau de neurones) est ensuite utilisé pour estimer les rendements des actifs en fonction de la matrice de données enrichies.

Ce projet vise à démontrer que l’intégration des régimes économiques, obtenus via une combinaison de PCA et du filtre de Hamilton, peut améliorer la compréhension et la prévision des rendements des actifs financiers. En utilisant les informations macro-économiques et en les combinant avec les caractéristiques spécifiques des actifs, il est anticipé que les modèles de prévision pourront capturer plus efficacement les dynamiques économiques complexes et fournir des estimations de rendement plus précises.

Chapitre 1

Revue de la Littérature

1.1 Prédiction des Rendements des Actifs

L'estimation des rendements des actifs financiers a toujours été un sujet central en finance. Les modèles traditionnels, tels que le modèle d'évaluation des actifs financiers (CAPM) , le modèle à trois facteurs (1993) et le medel à cinq facteurs (2014) de Fama et French , se concentrent principalement sur les caractéristiques spécifiques des actifs et les facteurs de marché ref : (Common risk factors in the returns on stocks and bonds, A five-factor asset pricing model).

Avec l'avènement des techniques de machine learning, de nouvelles approches ont émergé pour améliorer la prédiction des rendements des actifs.

Gu et al. (2020), dans leur étude ref : "Empirical Asset Pricing via Machine Learning", démontrent des gains économiques significatifs pour les investisseurs utilisant des prévisions basées sur l'apprentissage machine, doublant parfois les performances des stratégies de régression traditionnelles, grâce à leur capacité à capturer des interactions non linéaires entre les prédicteurs.

Chen et al. (2023), dans l' étude "Deep Learning in Asset Pricing", utilisent des réseaux de neurones profonds pour estimer les rendements des actions individuelles. Leur approche utilise une vaste quantité d'informations de conditionnement et la condition fondamentale de non-arbitrage comme fonction de critère pour construire les actifs de test les plus informatifs

avec une approche "adversariale ou antagoniste" (Réseau génératif antagoniste).

Ces avancées dans l'utilisation de l'apprentissage machine pour la prévision des rendements des actifs montrent que les modèles traditionnels peuvent être considérablement améliorés en incorporant des techniques modernes d'apprentissage automatique, ouvrant ainsi de nouvelles perspectives pour la recherche en finance.

1.2 Régimes économiques

Les régimes économiques jouent un rôle crucial dans l'estimation des rendements des actifs financiers, car ils permettent de capturer les changements structurels dans l'économie, changement qui peuvent influencer bien sur les rendements. Les régimes économiques peuvent être modélisés à l'aide de processus de changement de régime, tels que les modèles de Markov "Switching", qui permettent de passer entre différents états de l'économie de manière probabiliste.

L'utilisation des régimes économiques dans l'analyse financière est bien illustrée par les travaux de René Garcia et Pierre Perron (1996) dans leur étude "An Analysis of Real Interest Rates Under Regime Shifts". Ils montrent comment les taux d'intérêt réels peuvent être influencés par des changements de régime, ce qui permet une meilleure compréhension et prévision de leurs dynamiques.

Douglas T.F. Chau "Constructing an Observable Regime Indicator for Asset Allocation 2015 " propose un indicateur unique basé sur un modèle de facteur dynamique et un modèle de Markov à trois états pour classifier les régimes économiques. Cet indicateur identifie les marchés normaux, les récessions modérées et sévères, et capture les changements de volatilité et de corrélation des rendements des actifs.

Hamilton (1989), dans son étude pionnière "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", démontre l'importance des régimes économiques pour comprendre les cycles économiques et leurs impacts sur les rendements des

actifs.

Ces études montrent que les régimes économiques, et en particulier les modèles de changement de régime de Markov, sont des outils puissants pour améliorer la prévision et l'estimation des rendements des actifs financiers en tenant compte des changements structurels et des dynamiques économiques complexes.

1.3 Filtre de Kalman

Le Filtre de Kalman, développé par Rudolf E. Kalman, est une technique statistique fondamentale utilisée pour estimer les états non observés d'un système dynamique à partir de mesures bruitées. Introduit dans son article pionnier "A New Approach to Linear Filtering and Prediction Problems" (1960), le filtre de Kalman est largement utilisé dans divers domaines, y compris l'économie et la finance, en raison de sa capacité à gérer les dynamiques complexes et les changements structurels dans les séries temporelles.

Dans son étude de 1989, "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", James D. Hamilton applique le filtre de Kalman pour modéliser et analyser les séries temporelles économiques. Hamilton utilise le filtre de Kalman pour estimer les composantes non observées telles que les tendances économiques et les cycles d'affaires. Cette méthode permet d'extraire les tendances sous-jacentes des données observées et de filtrer le bruit, améliorant ainsi la précision de l'analyse des dynamiques économiques.

En gros le Filtre est utilisé pour :

- Détection des Régimes : Identifier les périodes de croissance et de récession dans les données macro-économiques.
- Prévision Économique : Améliorer les prévisions économiques en tenant compte des changements de régime.
- Analyse des Cycles Économiques : Comprendre les cycles économiques et leur impact sur les rendements des actifs financiers.

Chapitre 2

Présentation des données

Pour cette étude, on utilise les données que Chen et al. (2023), disponible sur le drive de Chen ([lien ici](#)) qui ont été collectées et traitées par eux dans le cadre de leur article "Deep Learning in Asset Pricing".

2.1 Caractéristiques des Actifs Financiers

Les données comprennent les rendements mensuels des actions américaines, collectées à partir de la base de données CRSP (Center for Research in Security Prices). La période d'échantillonnage s'étend de janvier 1967 à décembre 2016, couvrant ainsi 50 ans.

Les données sont divisés en trois échantillons : un échantillon d'entraînement de 20 ans (1967-1986), un échantillon de validation de 5 ans (1987-1991), et un échantillon de test hors échantillon de 25 ans (1992-2016).

Le reste de la description des données est disponible sur l'article de Chen et al. (2020) : [lien](#) :

Le nombre total d'actions utilisés est d'environ l'ensemble à environ 10 000 action, représentant les entreprises pour lesquelles toutes les informations caractéristiques de l'entreprise étaient disponibles à un moment donné.

2.2 Données Macro-économiques

Enfin, Chen et al. ont également collecté 178 séries temporelles macro-économiques provenant de trois sources : 124 prédictors macro-économiques de la base de données FRED-MD (détaillés par McCracken et Ng, 2016), les séries temporelles médianes transversales pour chacune des 46 caractéristiques des entreprises, et huit prédictors macro-économiques supplémentaires suggérés par Welch et Goyal (2007).

En outre, pour faciliter l'interprétation de l'impacte des différentes données macro-économiques sur l'estimation des rendements des actifs, on a classé les 178 séries temporelles en neuf catégories distinctes. Ces catégories, basées sur la classification de la base de données FRED-MD, sont les suivantes : Production et revenus, Travail et marché, Logement, Consommation, monnaie et crédit, taux d'intérêt et de change, Prix, Marchés boursiers et autres variables . Cette catégorisation permet de regrouper les séries temporelles similaires et de mieux comprendre les relations entre les différentes variables macroéconomiques. La description des différentes variables et des catégories est accessible en ligne dans le travail de McCracken et Ng (2020) sur FRED-MD, [lien \(ici\)](#),

Chapitre 3

PCA et filtre de Hamilton : Donnée macro

Cette section détaille la méthodologie employée pour effectuer l'analyse en composantes principales (PCA) et appliquer le filtre de Hamilton pour obtenir les probabilité des régimes.

3.1 Analyse en Composantes Principales (PCA)

L'Analyse en Composantes Principales (PCA) est une technique de réduction de dimensionnalité largement utilisée en statistique et en apprentissage automatique. Elle permet de transformer un ensemble de variables corrélées en un ensemble de variables non corrélées appelées composantes principales, tout en conservant autant que possible de la variance présente dans les données d'origine.

La PCA vise à identifier les directions dans lesquelles les données varient le plus. Ces directions sont représentées par des vecteurs appelés composantes principales. L'objectif est de réduire la complexité des données tout en conservant leur structure essentielle.

3.1.1 Étapes de la PCA

1. **Standardisation des données** : Pour garantir que chaque variable contribue équitablement à l'analyse, il est courant de standardiser les données, en les centrant (soustraction de la moyenne) et en les réduisant (division par l'écart-type).

2. **Calcul de la matrice de covariance** : La matrice de covariance, notée Σ , mesure les relations de covariance entre les différentes variables. Pour un ensemble de données X avec n observations et p variables, la matrice de covariance est calculée comme suit :

$$\Sigma = \frac{1}{n-1} X^T X$$

3. **Calcul des vecteurs propres et des valeurs propres** : Les vecteurs propres (\mathbf{v}) et les valeurs propres (λ) de la matrice de covariance sont calculés pour déterminer les directions principales et l'importance de ces directions respectivement. Ces vecteurs propres forment les composantes principales et les valeurs propres indiquent la quantité de variance expliquée par chaque composante. Mathématiquement, cela se traduit par :

$$\Sigma \mathbf{v} = \lambda \mathbf{v}$$

4. **Projection des données** : Les données d'origine sont projetées dans le nouvel espace formé par les composantes principales, transformant ainsi les données initiales en un ensemble de nouvelles variables non corrélées :

$$X' = X \mathbf{V}$$

où \mathbf{V} est la matrice des vecteurs propres.

3.1.2 Résultats de la PCA

- **Variance expliquée** : Les valeurs propres (λ) de la matrice de covariance représentent la quantité de variance expliquée par chaque composante principale. La somme des valeurs propres est égale à la variance totale des données d'origine, et le pourcentage de variance expliquée par chaque composante principale est donné par :

$$\text{Variance expliquée par } \mathbf{v}_i = \frac{\lambda_i}{\sum_{j=1}^p \lambda_j}$$

- **Choix des composantes principales** : Les composantes principales sont ordonnées par la quantité de variance qu'elles expliquent. En général, seules les premières composantes principales (celles avec les plus grandes valeurs propres) sont conservées, car

elles capturent la majorité de la variance dans les données.

- **Dé-corrélation des variables** : Les composantes principales sont orthogonales entre elles, ce qui signifie qu'elles sont non corrélées. Cela simplifie l'analyse et permet d'éviter la redondance d'information.

Dans le contexte de ce projet , on regroupe les données macro-économiques en plusieurs groupes, puis on effectue une PCA sur chacun de ces groupes. Cela permet de :

- **Réduire la dimensionnalité** : En résumant chaque groupe de variables macro-économiques par un petit nombre de composantes principales on passe de 178 variables à 8 groupes, cela nous facilite l'utilisation du filtre de Hamilton sans perdre d'information cruciale, autrement le bruit de ces données rendra le filtre inutilisable (dans le filtre il y a une partie d'optimisation où on doit minimiser une fonction objective, avec 178 variables le nombre de paramètres de la fonction explose et trouver un optimum devient presque impossible)
- **Éliminer les redondances** : Les composantes principales étant non corrélées, elles éliminent la redondance présente dans les données initiales, ce qui peut améliorer la performance des algorithmes de prédiction.
- **Améliorer l'interprétabilité** : Avec un nombre limité de groupe de données macro l'interprétation devient plus facile que les 178 variables d'origine, facilitant ainsi l'analyse des résultats.

En résumé, on groupe nos données en neuf catégories et on ajoute un dixième groupe qui contient toutes les données et on applique la PCA sur chaque groupe, on garde que la première composante de chaque groupe. On obtient ainsi les séries chronologiques de chaque groupe de données macro économiques.

3.2 Modèle de changement de Régimes

3.2.1 Représentation État-Espace

Un système linéaire peut être représenté en forme état-espace avec les équations suivantes :

Équation d'observation

$$\mathbf{y}_t = \mathbf{A}'\mathbf{x}_t + \mathbf{H}'\xi_t + \omega_t \quad (3.1)$$

avec $\mathbf{E}[\omega_t \omega_t] = \mathbf{R}$ de dimension $n \times n$

où :

- \mathbf{y}_t est le vecteur d'observation à l'instant t (dimension $n \times 1$).
- \mathbf{H}' est la matrice de coefficients appelée matrice d'émission (dimension $n \times r$).
- ξ est le vecteur d'états du système (dimension $r \times 1$).
- \mathbf{x}_t vecteur de variables exogènes prédéterminées (dimension $k \times 1$).
- \mathbf{A}' est la matrice de coefficients pour les variables exogènes (dimension $n \times k$).
- ω_t un vecteur d'erreur de mesure ou de bruit (dimension $n \times 1$).

Équation d'état

$$\xi_{t+1} = \mathbf{F}\xi_t + \mathbf{v}_{t+1} = \quad (3.2)$$

avec $\mathbf{E}[\mathbf{v}_t \mathbf{v}_t] = \mathbf{Q}$ de dimension $r \times r$

où :

- ξ_t est le vecteur d'état à l'instant t (dimension $n \times 1$).
- \mathbf{F} est la matrice de transition d'état (dimension $n \times n$).
- \mathbf{v}_t est le bruit de processus (dimension $n \times 1$).

3.2.2 Filtre de Kalman

Le filtre de Kalman est une méthode récursive pour estimer l'état d'un processus dynamique en minimisant l'erreur quadratique moyenne. Il fonctionne en deux étapes principales : la prédiction et la mise à jour.

Étape de prédiction

$$\hat{\xi}_{t+1|t} = \mathbf{F}\hat{\xi}_{t|t} \quad (3.3)$$

$$\mathbf{P}_{t+1|t} = \mathbf{F}\mathbf{P}_{t|t}\mathbf{F}' + \mathbf{Q} \quad (3.4)$$

où :

- $\hat{\xi}_{t+1|t}$ est l'estimation a priori de l'état à l'instant $t + 1$.

— $\mathbf{P}_{t|t-1}$ est la covariance a priori de l'erreur d'estimation.

Étape de mise à jour

$$\begin{aligned}\mathbf{K}_t &= \mathbf{P}_{t|t-1} \mathbf{H} (\mathbf{H}' \mathbf{P}_{t|t-1} \mathbf{H} + \mathbf{R})^{-1} \\ \hat{\xi}_{t|t} &= \hat{\xi}_{t|t-1} + \mathbf{K}_t (\mathbf{y}_t - \mathbf{A}' \mathbf{x}_t - \mathbf{H}' \hat{\xi}_{t|t-1}) \\ \mathbf{P}_{t|t} &= (\mathbf{I} - \mathbf{K}_t \mathbf{H}') \mathbf{P}_{t|t-1}\end{aligned}$$

où :

- \mathbf{K}_t est le gain de Kalman.
- $\hat{\xi}_{t|t}$ est l'estimation a posteriori de l'état à l'instant t .
- $\mathbf{P}_{t|t}$ est la matrice de covariance de la distribution conditionnelle de ξ_t (covariance a posteriori de l'erreur d'estimation).

3.2.3 Changement de régimes de Hamilton

Hamilton a étendu le filtre de Kalman pour modéliser les changements de régimes économiques en utilisant des chaînes de Markov à états cachés.

Par exemple un Changement de régime AR(1) à la forme :

$$y_t = \mu_{st} + \beta_{st} y_{t-1} + \omega_{st}$$

ou $s_t = 1, 2, 3, \dots, N$ sont les régimes possibles.

Dans le filtre de Kalman les états ξ_t sont généralement des variables continues et leur distribution peut être modélisée par des distributions continues, comme la distribution normale. Cependant, lorsqu'on modélise des régimes économiques, les états peuvent être des variables discrètes. Dans ce cas, les états représentent des régimes ou des conditions spécifiques qui sont qualitativement distincts, comme des périodes de croissance économique, de récession, ou de stagnation, donc ces états sont des variables discrètes et leur distribution conditionnelle se définit par :

$$P(s_t = i | \zeta_{t-1})$$

avec $i = 1, 2, \dots$ et ζ_{t-1} représente toute l'information (variables et paramètres) connue jusqu'au temps $(t-1)$.

Dans le cas d'une chaîne de Markov

$$P(s_t = i | \zeta_{t-1}) = P(s_t = i | s_{t-1})$$

ce qui donne

$$P(s_t = j | s_{t-1} = i) = P_{ij}$$

Ainsi dans le cas de deux régimes la matrice des probabilités de transitions est :

$$P(s_t = j | s_{t-1} = i) = P_{ij} = \begin{bmatrix} p_{00} & p_{10} \\ p_{01} & p_{11} \end{bmatrix}$$

Hamilton propose la même technique que Kalman : prédiction puis mise à jour pour trouver :

$$\xi_{i,t} = P(s_t = i | \zeta_t)$$

ou $\zeta_t = y_t, y_{t-1}, \dots$

en partant de :

$$\xi_{i,t-1} = P(s_{t-1} = i | \zeta_{t-1})$$

avec l'assomption :

$$f(y_t | s_t = i, \zeta_{t-1}) \sim N(\mu_{y_t}, \sigma)$$

Algorithme :

L'algorithme proposé par Hamilton est comme suite :

Étape1 : Poser

$$\xi_{i,t-1} = P(s_{t-1} = i \mid \zeta_{t-1})$$

Étape2 : Calculer la distribution de densité conditionnelle conjointe de y_t et de s_t :

$$f(y_t, s_t = i \mid \zeta_{t-1}) = f(y_t \mid s_t = i, \zeta_{t-1}) \cdot P(s_t = i \mid \zeta_{t-1})$$

ou :

$$f(y_t \mid s_t = j, \zeta_{t-1}) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp \left[-\frac{(y_t - \mu_j - \beta_j y_{t-1})^2}{2\sigma_j^2} \right]$$

Étape3 : On utilise la densité conditionnelle et les probabilité à priori pour obtenir la vraisemblance de l'observation :

$$f(y_t \mid \zeta_{t-1}) = \sum_{j=0}^1 f(y_t, s_t = j \mid \zeta_{t-1}) = \sum_{j=0}^1 f(y_t \mid s_t = j, \zeta_{t-1}) P(s_t = j \mid \zeta_{t-1})$$

$$f(y_t \mid \zeta_{t-1}) = \sum_{i=0}^1 \sum_{j=0}^1 f(y_t \mid s_t = i, \zeta_{t-1}) P_{ij} P(s_{t-1} = i \mid \zeta_{t-1}) = \sum_{i=0}^1 \sum_{j=0}^1 \xi_{i,t-1} p_{ij} f(y_t \mid s_{t-1} = i, \zeta_{t-1})$$

En détails :

$$f(y_t \mid \zeta_{t-1}) = \xi_{0,t-1} p_{00} \eta_{0t} + \xi_{0,t-1} p_{01} f(y_t \mid s_t = i, \zeta_{t-1}) + \xi_{1,t-1} p_{10} \eta_{0t} + \xi_{1,t-1} p_{11} f(y_t \mid s_t = i, \zeta_{t-1})$$

Étape4 : On utilise la vraisemblance pour mettre à jour le nouveau état :

$$\xi_{jt} = P(s_t = j \mid \zeta_t) = \frac{\sum_{i=0}^1 \xi_{i,t-1} p_{ij} \eta_{jt}}{f(y_t \mid \zeta_{t-1})}$$

Finalement on répète de : $t = 1$ à T pour obtenir les probabilités à toutes les périodes.

Optimisation :

Le log vraisemblance des données peut être calculer comme suit :

$$\log(f(y_T, y_{T-1}, \dots, y_1 \mid y_0)) = \sum_{t=1}^T \log(f(y_t \mid \zeta_{t-1}))$$

Pour trouver les paramètres optimaux, on doit maximiser cette log-vraisemblance ou minimiser $-\log L(\theta)$. Cela peut être accompli en utilisant des méthodes d'optimisation numérique comme l'algorithme de descente de gradient, Newton-Raphson, ou BFGS.

L'optimisation nécessite de choisir des paramètres initiaux :

Pour l'état initial ξ_{i0} , Hamilton propose d'utiliser les probabilités non conditionnelle :

$$\xi_{i0} = P(s_0 = i) = \frac{1 - p_{jj}}{2 - p_{ii} - p_{jj}}$$

Ces probabilités sont obtenu sous l'assomption que la chaîne de Markov est ergodique.

Pour les autres paramètres on choisit des valeurs arbitraire raisonnables.

A partir de ces paramètres initiaux on utilise l'optimisation numérique (BFGS) pour obtenir les paramètres optimaux $(\mu_0, \mu_1, \beta_0, \beta_1, \sigma_0, \sigma_1, p, q)$.

A noté qu'on peut s'en passer du β pour obtenir un modèle sans auto régression :

$$y_t = \mu_{s_t} + \sigma_{s_t}$$

.

On a implémenté les deux cas pour le filtre de Hamilton et après comparaison des résultats, on a choisit les probabilités issues du modèle sans auto-régression pour être intégrées dans le modèle de prédiction des rendements.

Chapitre 4

Prédiction des Rendements des Actifs

4.1 Introduction aux Réseaux de Neurones

En apprentissage machine les réseaux de neurones artificiels (ANN - Artificial Neural Networks) sont des modèles conçus pour reconnaître des motifs complexes dans les données et les utiliser pour effectuer des tâches tel que la prédiction.

Ils sont composés de couches de noeuds inter-connectés qui traitent les informations de manière non linéaire, permettant ainsi de modéliser des relations complexes entre les variables d'entrée et de sortie.

4.2 Structure et Fonctionnement d'un Réseau de Neurones

Un réseau de neurones peut être décrit comme une succession de fonctions appliquées aux données d'entrée pour produire une sortie. Le but est de trouver une fonction complexe f telle que $y = f(X)$, où X représente les données d'entrée (caractéristiques des actifs et états des régimes macro par exemple) et y représente la sortie (rendement prédit dans cet exemple).

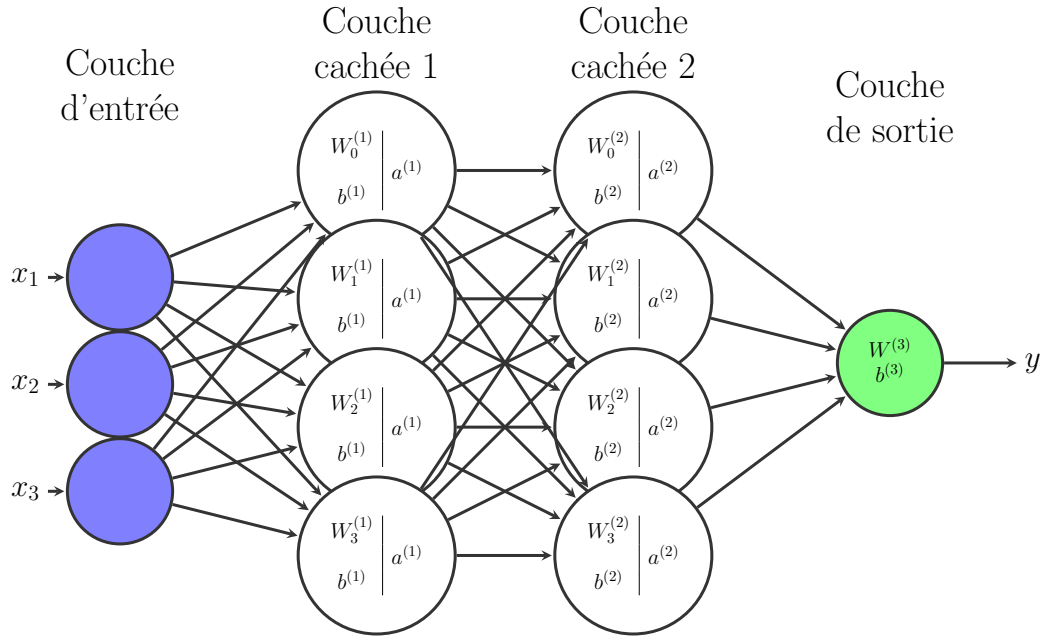


FIGURE 4.1 – Exemple de réseau à 2 couches cachées de 4 noeuds chacune (pour régression)

4.2.1 Couche d'entrée

- **Description** : La couche d'entrée reçoit les données initiales. Chaque noeud de cette couche représente une caractéristique spécifique de l'ensemble de données.
- **Formule** : $\mathbf{a}^{(0)} = \mathbf{x}$, où \mathbf{x} est le vecteur des caractéristiques d'entrée. La couche d'entrée ne contient ni poids ni biais car elle sert seulement comme support pour recevoir les entrées.

4.2.2 Couches Cachées

- **Description** : Les couches cachées sont composées de noeuds qui effectuent des transformations non linéaires sur les données. Chaque couche cachée l reçoit les activations

de la couche précédente $l - 1$, les transforme, et produit des activations pour la couche suivante.

- **Formule** : $\mathbf{z}^{(l)} = \mathbf{W}^{(l)}\mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}$, où $\mathbf{W}^{(l)}$ est la matrice de poids, $\mathbf{b}^{(l)}$ est le vecteur de biais, et $\mathbf{z}^{(l)}$ est le vecteur de valeurs linéaires.
- **Fonction d'activation** : $\mathbf{a}^{(l)} = g(\mathbf{z}^{(l)})$, où g est une fonction d'activation non linéaire (par exemple, ReLU, Sigmoid), cette fonction permet de capter les interactions non linéaire dans les données.

4.2.3 Couche de Sortie

- **Description** : La couche de sortie produit le résultat final du réseau. Pour la prédiction des rendements des actifs, cette couche génère une valeur continue représentant le rendement prédit.
- **Formule** : $\mathbf{y}_{\text{prédit}} = \mathbf{a}^{(L)}$, où L est le nombre total de couches.

4.3 Entraînement du Modèle

L'objectif de l'entraînement est de trouver les poids \mathbf{W} et les biais \mathbf{b} qui minimisent l'erreur entre les prédictions du modèle et les valeurs réelles. Ce processus se déroule en plusieurs étapes :

4.3.1 Propagation Avant (Forward Propagation)

Les données d'entrée traversent le réseau couche par couche. Chaque neurone applique la fonction $\mathbf{z}^{(l)} = \mathbf{W}^{(l)}\mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}$ suivie de l'activation $\mathbf{a}^{(l)} = g(\mathbf{z}^{(l)})$. Cela permet de calculer la sortie prédite $\mathbf{y}_{\text{prédit}}$.

4.3.2 Calcul de la Perte (Loss)

La fonction de perte $L(\mathbf{y}_{\text{réel}}, \mathbf{y}_{\text{prédit}})$ mesure la différence entre les prédictions et les valeurs réelles. Pour un problème de régression, l'erreur quadratique moyenne (Mean Squared Error, MSE) est couramment utilisée :

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

4.3.3 Rétro propagation (Backpropagation)

L'erreur calculée est propagée en arrière à travers le réseau pour mettre à jour les poids et les biais. Les gradients de la fonction de perte par rapport aux poids et aux biais sont calculés en utilisant la règle de la chaîne (A noté que les graphes de calculs permettent d'exécuter la différentiation automatique d'une manière rapide et précise). Les poids et les biais sont mis à jour en utilisant un algorithme d'optimisation de la descente de gradient :

$$\mathbf{W}^{(l)} \leftarrow \mathbf{W}^{(l)} - \eta \frac{\partial L}{\partial \mathbf{W}^{(l)}}$$

et

$$\mathbf{b}^{(l)} \leftarrow \mathbf{b}^{(l)} - \eta \frac{\partial L}{\partial \mathbf{b}^{(l)}},$$

où η est le taux d'apprentissage.

Il existe plusieurs algorithmes pour implémenter la descente de gradient, chacun ayant ses propres avantages. Parmi eux, Adam et WAdam sont particulièrement notables.

Adam (Adaptive Moment Estimation) est un algorithme d'optimisation qui adapte les taux d'apprentissage pour chaque paramètre individuellement. Il utilise des moyennes mobiles des gradients et des carrés des gradients pour stabiliser et accélérer la convergence. À chaque itération, Adam met à jour les paramètres selon les formules suivantes :

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

$$\theta_t = \theta_{t-1} - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}$$

où g_t est le gradient à l'instant t , β_1 et β_2 sont des coefficients de décroissance, α est le taux d'apprentissage, et ϵ est une petite constante pour éviter la division par zéro.

WAdam (Weighted Adam) est une variante de l'optimiseur Adam qui intègre un terme de décroissance de poids pour améliorer la régularisation des modèles. En plus des mises à jour standard d'Adam, WAdam ajoute un terme de régularisation L2, ce qui aide à réduire le surapprentissage. La règle de mise à jour de WAdam est :

$$\theta_t = \theta_{t-1} - \alpha \left(\frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} + \lambda \theta_{t-1} \right)$$

où λ est le coefficient de décroissance de poids. Cette régularisation explicite permet d'obtenir des modèles plus robustes et performants, en particulier lorsqu'il y a un grand nombre de paramètres à optimiser.

Dans ce projet on utilise l'algorithme d'optimisation WAdam.

4.3.4 Itérations et Convergence

Le processus de propagation avant et de rétro propagation est répété pour de nombreuses itérations (époques) jusqu'à ce que l'erreur soit minimisée et que le modèle converge vers une solution optimale.

4.4 Implémentation

4.4.1 Préparation des Données

Après application du filtre de Hamilton pour déterminer les probabilités de régime macro-économiques, celles ci sont combinées avec les caractéristiques des actifs pour donner une matrice de caractéristiques augmentée.

En effet pour prendre en compte toutes les interactions linéaire et non linéaire entre les données on construit une matrice de caractéristiques enrichies et on implémente le modèle suivant en utilisant la même idée que Gu. et al. (2020) :

$$E[r_{i,t+1}] = f(x_{it} \otimes c_t) = f(z_{it})$$

ou c_t est constitué de la probabilité et d'un $c_t = [1, p_t]$ et puisque nous avons que deux régimes on utilise les probabilités d'un seul régime (l'autre probabilité n'ajoute aucune infor-

mation $p_2 = 1 - p_1$), ainsi le schéma de nos données se simplifie à

$$E[r_{i,t+1}] = f(x_{it} * c_t) = f(z_{it})$$

En gros, les données d'entrées contiennent les caractéristiques et les interactions entre ces caractéristiques avec les régimes économiques (ces régimes étant communs à toutes les caractéristiques).

4.4.2 Construction du Modèle

- Les caractéristiques augmentées servent de données d'entrée au modèle. Afin de prendre en compte la dépendance temporelle entre les données on fait une prédiction cross sectionnelle, c'est à dire que chaque entrée comprend toutes les caractéristiques qui correspondent à une période spécifique et chaque erreur calculée correspond à une période.
- La fonction activation utilisée est Relu.
- : Pour évaluer l'erreur on utilise la fonction objective qui prend en compte la dépendance temporelle entre les données :

$$L(\theta) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (r_{it+1} - f(z_{it}; \theta))$$

4.4.3 Entraînement du Modèle

Une partie des données est utilisée pour entraîner le modèle afin de trouver les paramètres optimaux (poids et biais). La rétro propagation est utilisée pour ajuster les poids en fonction de l'erreur d'entraînement.

Malgré leur puissance les modèles de réseaux de neurones peuvent souffrir de problèmes de sur entraînement (surtout si les données sont insuffisantes), et leur convergence dépend aussi des points initiaux, la validation est utilisée pour trouver les hyper paramètres optimaux dans ce cas.

Pour trouver les hyper paramètres optimaux on utilise l'ensemble de données de validation

et on essaie un ensemble d'hyper paramètres : deux pas d'apprentissage, deux ensembles de paramètres initiaux , deux 'dropping rates' et deux architectures différentes , on choisit les hyper paramètres qui donnent l'erreur minimale sur cet ensemble.

4.4.4 Évaluation et Prédiction

Le modèle est validé en utilisant un ensemble de données distinct pour vérifier sa performance. Les métriques telles que l'erreur quadratique moyenne (MSE) ou le coefficient de détermination (R^2) et le Ratio de Sharpe sont utilisées pour évaluer la qualité des prédictions. Une fois le modèle entraîné et validé, il est utilisé pour prédire les rendements des actifs sur de nouvelles données.

Chapitre 5

Résultats

Cette section présente les résultats de l’analyse, y compris les probabilités de régimes économiques obtenues, et les performances du modèle.

5.1 Probabilités de Régimes Économiques

Dans les deux figures ci-après, nous observons les probabilités des états du régime 1 obtenues en utilisant les composantes principales de différents groupes de données, y compris les cas où l’on utilise tous les groupes et le cas où l’on fait le PCA directement sur toutes les données sans regroupement préalable.

Étant donné qu’il n’y a que deux régimes, nous présentons uniquement les probabilités du régime 1, car les probabilités du régime 2 peuvent être déduites de celles du régime 1 et n’apportent pas d’informations supplémentaires.

Les probabilités issues du filtre de Hamilton sont des probabilités filtrées et lissées pour les données d’entraînement et de validation (avant 1991), ainsi que des probabilités filtrées pour les données de test (après 1991), on a concaténé les deux séries temporelles pour obtenir une seule série temporelle le long de toute la période.

Les figures montrent que le filtre de Hamilton parvient à capturer les changements de régimes. Toutefois, le modèle sans auto-régression $y_t = \mu_t + \sigma_t$ s’adapte mieux aux données historiques de récessions du Bureau National de Recherche Économique (NBER). Cela suggère que l’inclusion des termes auto-régressifs peut introduire du bruit ou des dynamiques complexes qui ne sont pas présentes dans les données NBER.

Les différents groupes de données ne captent pas les régimes de la même manière. Le

modèle construit avec le groupe "Marché du travail" s'adapte le mieux aux données du NBER, alors que dans certains groupes, les changements de régime sont captés en retard ou persistent. Cela peut être dû à la nature des données macro-économiques de chaque groupe et à leur interaction avec les changements de régime économique. Par exemple, les données du marché du travail peuvent réagir plus rapidement aux changements économiques par rapport à d'autres indicateurs macro-économiques qui pourraient avoir des délais de réaction plus longs.

De plus, le cas où l'on exécute le filtre de Hamilton sur la première composante de toutes les données donne de meilleurs résultats que lorsque l'on applique le filtre sur les 8 premières composantes des groupes constitués. Cela peut être dû à des problèmes d'optimisation : lorsqu'on augmente le nombre de séries chronologiques dans le filtre de Hamilton, trouver les paramètres optimaux devient difficile. Ici, nous avons supposé que les composantes principales des groupes sont indépendantes. En cas de dépendance, le nombre de paramètres augmente, rendant l'optimisation encore plus complexe.

En somme, ces observations montrent que le filtre de Hamilton et le modèle sans auto-régression captent bien les changements de régime économique. En supposant que les rendements des actifs sont sensibles à ces régimes, les probabilités des régimes seront des indicateurs pertinents pour la prédiction des rendements.

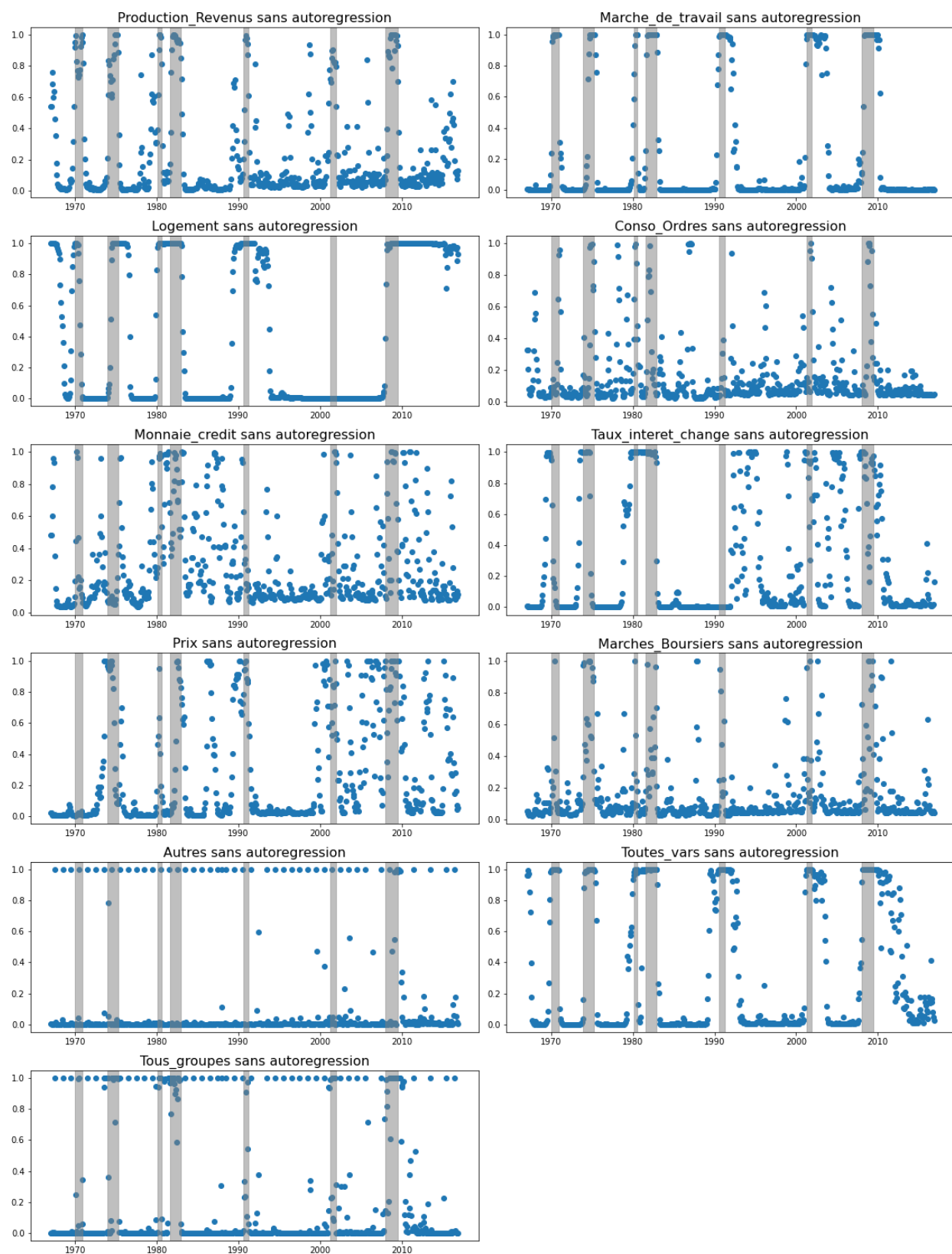


FIGURE 5.1 – Probabilités du régime 1 et récessions NBER, modèle sans auto-régression

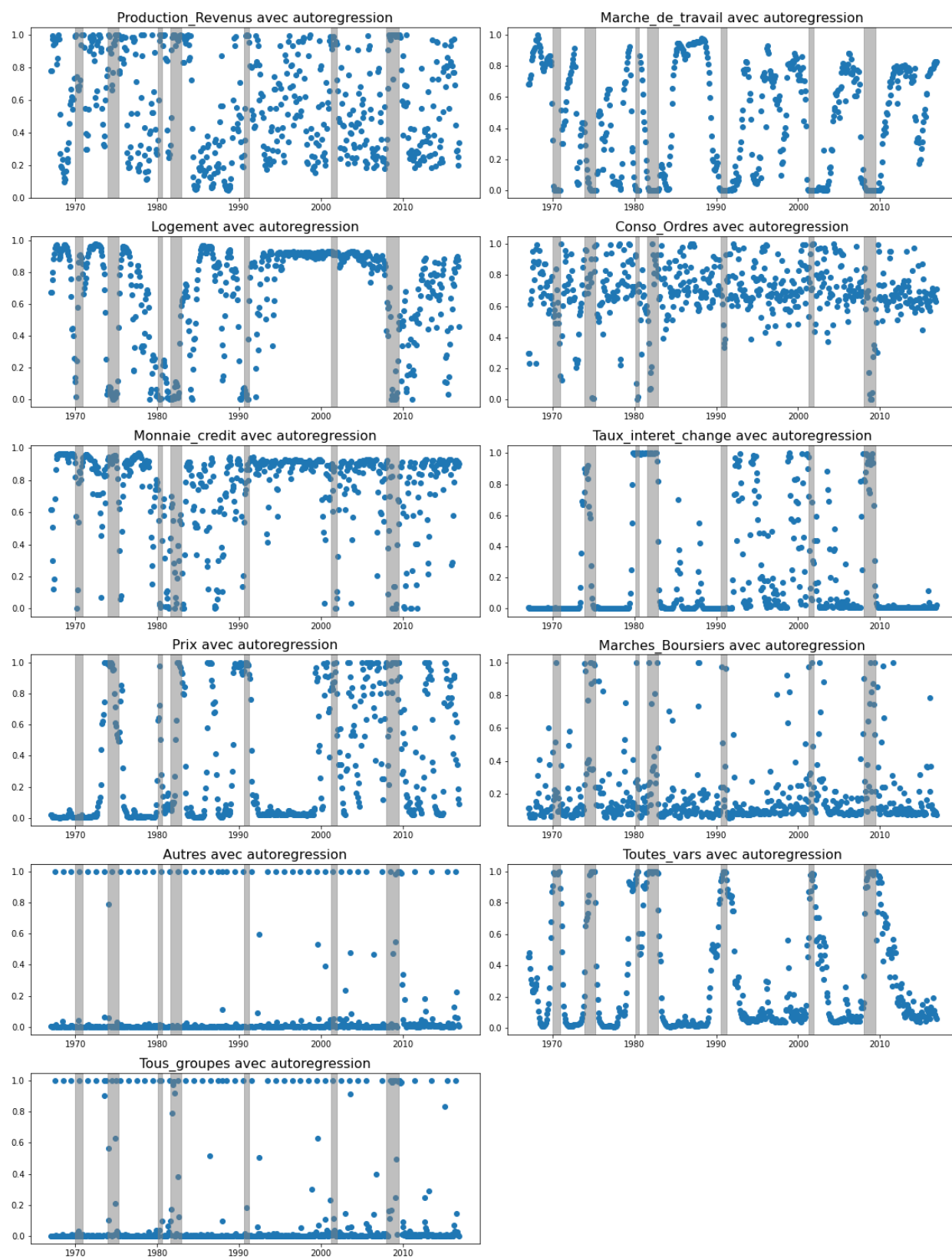


FIGURE 5.2 – Probabilités du régime 1 et récessions NBER, modèle AR(1)

5.2 Performance du Modèle de Prédiction

Pour chaque groupe macro-économique construit par l'approche intégrant le PCA et le filtre de Hamilton, on évalue la performance du modèle de prévision de rendements sur des données hors échantillon (out-of-sample). On présente les résultats pour chaque groupe de données macro.

groupe macro	sharpe
Marche de travail	0.699
Toutes vars	0.682
Tous groupes	0.68
Marches Boursiers	0.669
Conso Ordres	0.645
Production Revenus	0.643
Autres	0.636
Logement	0.627
Monnaie credit	0.612
Sans macro	0.573
Prix	0.562
Taux interet change	0.542

FIGURE 5.3 – Ratios de Sharpe portefeuille Long Court (Hors Ech)

Pour chaque groupe macro-économique, on a construit des portefeuilles long-courts en utilisant les prédictions de rendement du modèle. La stratégie long-court implique de prendre des positions longues sur les actifs avec des rendements prédits élevés et des positions courtes sur ceux avec des rendements prédits faibles. Cette approche permet de bénéficier des gains sur les actifs performants tout en couvrant les risques avec les positions courtes.

Sur la figure 5.3 on remarque que les ratios de Sharpe indiquent que les portefeuilles construits à partir des prédictions obtenus en utilisant les données construites en incluant les régimes macro offrent un meilleur rendement ajusté en fonction du risque.

Les prédictions obtenus en incluant les régimes obtenus à partir du groupe "Marché de travail" offre la meilleur performance suivi par les régimes construits à partir de la composante principales de toutes les séries chronologiques (sans groupement), par les régimes obtenus en utilisant un filtre de Hamilton multi-varié sur les composantes principales des 9 groupes de données macro et enfin du régimes obtenus en utilisant le groupe "Marchés boursiers".

Les données construites en incluant les régimes obtenus à partir des composantes principales des groupes de séries chronologiques "Prix" et "Taux d'intérêts et Taux de Change" ainsi que les données n'incluant pas de régimes macro donnent les pires ratios de Sharpe.

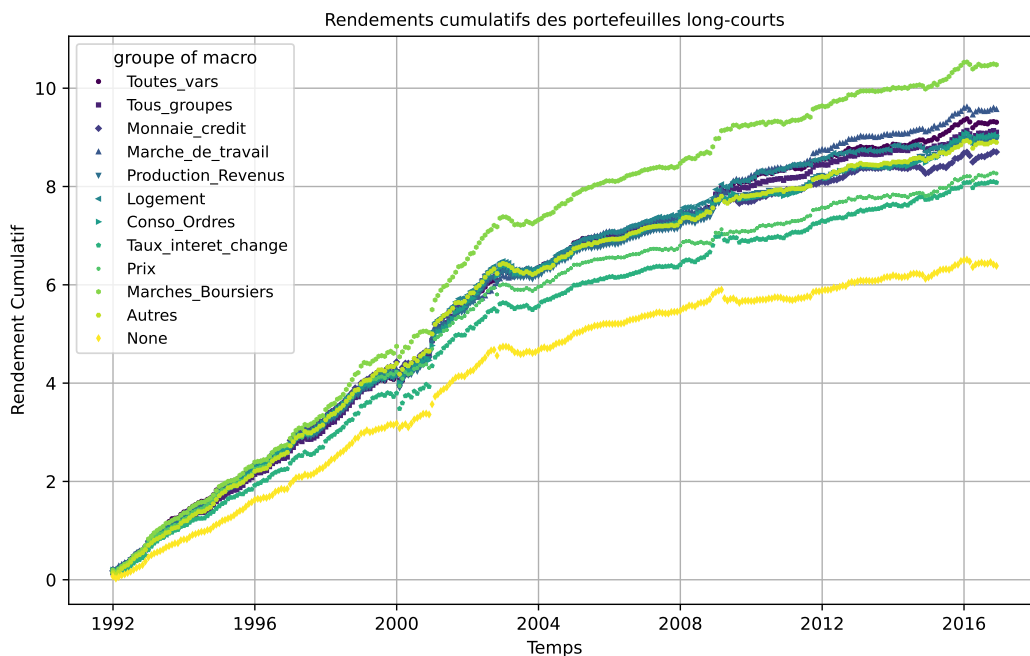


FIGURE 5.4 – Rendement cumulatif de portefeuilles Long-court

Pour mieux visualiser la différence entre les rendements obtenus en utilisant les régimes extraits de différents groupes de données macro et aussi le cas où on utilise pas d'information macro, on calcule les rendements cumulatifs des portefeuilles long-court dans chaque cas de figure.

Sur la figure 5.4 on remarque que l'intégration des régimes dans la prédiction des rendements

des actifs améliore nettement les rendements des portefeuilles. Le rendement du portefeuille long-court dans le cas où on inclut pas le régime macro dans les données est nettement inférieur aux autres rendements. Le rendement obtenu en utilisant les données du "Marchés boursiers" est nettement supérieur. Ces résultats sont en accord avec les précédentes et confirment que l'utilisation des données macro dans la prédiction offrent des gains meilleurs.

On classe aussi les actifs par prédiction de rendement et on construit des portefeuilles de décile. Les actifs sont répartis en dix groupes, ou déciles, allant de ceux avec les rendements prédits les plus bas à ceux avec les rendements prédits les plus élevés. On calcule le rendement cumulé de chaque portefeuille de décile pour chaque groupe macro-économique inclus dans l'analyse des régimes.

La figure 5.4 montre le rendement cumulé des portefeuilles de déciles selon le groupe de données macro utilisé y compris le cas où on utilise juste les caractéristiques des actifs sans données de régimes macro.

Dans tous les cas on remarque une divergence entre les déciles inférieurs et supérieurs en termes de rendement cumulé. Les déciles supérieurs, composés des actifs ayant les prédictions de rendement les plus élevées, ont constamment surpassé les déciles inférieurs. On compare les résultats, la divergence entre les rendements des déciles inférieur et supérieur est plus grande lorsqu'on inclut les régimes macro dans la prédiction que lorsqu'on utilise pas d'information macro, les rendements obtenus en utilisant les groupes "Marché de travail", "Toutes variables économiques" et "Tous les groupes" offrent la meilleure prédiction en termes de divergence, tant que les groupes "Taux d'intérêt et taux de change", "Prix" et aussi le cas où on inclut pas de régime donnent la pire prédiction.

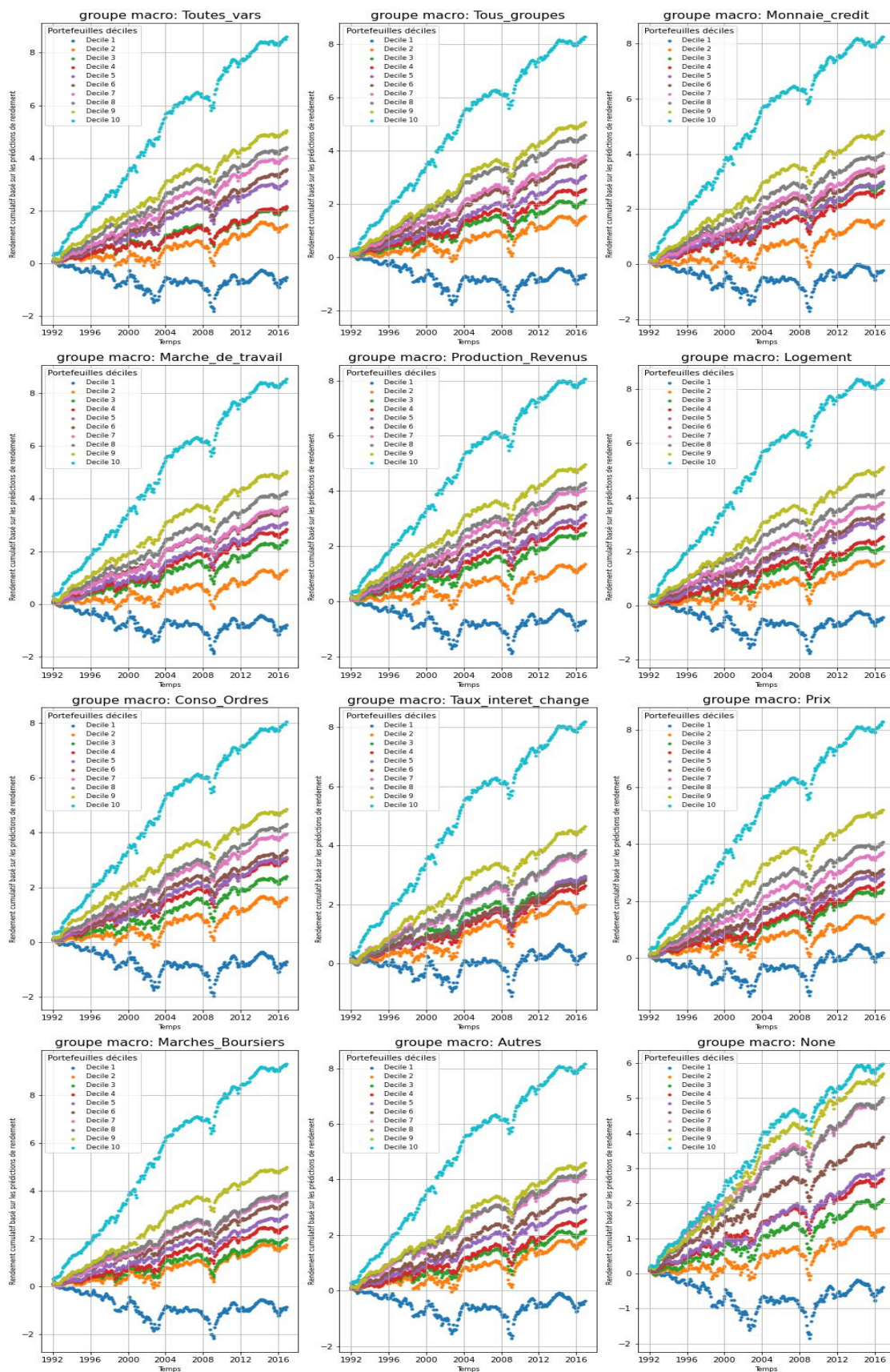


FIGURE 5.5 – Rendements cumulatifs de portefeuilles de déciles classé par prédiction de rendement

Chapitre 6

Discussion

Cette section discute des implications des résultats, des limitations de l'étude, et des suggestions pour des recherches futures.

6.1 Implications des Résultats

6.2 Limitations de l'étude

6.3 Suggestions

Conclusion

Références

- [1] Hamilton, James D., *A new approach to the economic analysis of nonstationary time series and the business cycle*, Econometrica : Journal of the Econometric Society Vol 57 No 2, pages 366–375, 1989.
- [2] Hamilton, James D., *Regime-Switching Models*, Palgrave Dictionary of Economics, pages 1–6, 2005.
- [3] Hamilton, James D., *Time series analysis*, Princeton University Press, pages 372–400, 1994.
- [4] Kálmán, Rudolf E., *A New Approach to Linear Filtering and Prediction Problems*, Transactions of the ASME–Journal of Basic Engineering, volume 82, number Series D, pages 35–45, 1960,
- [5] Douglas T.F. Chau., *Constructing an Observable Regime Indicator for Asset Allocation*, Working Paper, pages 1–49, 2015.
- [6] Luyang Chen, Markus Pelger, Jason Zhu, *Deep Learning in Asset Pricing*, Management Science Articles in Advance, pp. 1–37, 2023.
- [7] Michael W. McCracken, and Serena Ng, *FRED-MD : A Monthly Database for Macroeconomic Research*, Federal Reserve Bank of St. Louis Working Paper 2015-012., 2015.
- [8] Shihao Gu, Bryan Kelly, Dacheng Xiu, *Empirical Asset Pricing via Machine Learning*, The Review of Financial Studies 33 (2020) 2223–2273, 2020.
- [9] Engle, R. F., & McFadden, D. L., *Handbook of Econometrics*, Volume 1 V, Edited by R.F. Engle and D.L. McFadden , Chapter 50 State-Space Models , Pages 3041–3071, 1994
- [10] G. Kapetanios, F. Kempf, *Interpretable Machine Learning Modelling for Asset Pricing*, Working paper No. 2022/1 | February 2022 | ISSN 2516-5933, pages 1–36, 2022.

- [11] James D. Hamilton, Baldev Raj, *New directions in business cycle research and financial analysis*, Empirical Economics (2002) 27 :149–162, 2002.

.....

- [12] Fama, Eugene F., and Kenneth R. French, *Common risk factors in the returns on stocks and bonds*, Journal of Financial Economics, volume 33, number 1, pages 3–56,