UNIVERSITÉ DU QUÉBEC À MONTRÉAL

PRÉVISION DE LA DENSITÉ AVEC UN RÉSEAU DE NEURONES EN HÉMISPHÈRE

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN ÉCONOMIQUE

PAR

MIKAEL FRENETTE

Août 2024

"Take data and statistical techniques, throw them at a computer and what you have is machine learning" - Paul Wilmott

REMERCIEMENTS

J'aimerais tout d'abord remercier mon directeur de mémoire, M. Philippe Goulet Coulombe, pour son appuie et pour la qualité de son enseignement durant ces deux dernières années.

Je tiens également à remercier ma famille et mes amis pour leur soutien et leurs conseils qui m'ont aidé tout au long de ce parcours. Je remercie également Lance L'archevêque pour sa contribution dans l'élaboration des graphiques utilisés dans ce mémoire.

Enfin, je souhaite exprimer ma gratitude envers le restaurant Ginko, non seulement pour son soutien moral, mais aussi pour avoir stimulé la créativité celle de moi mes collègues tout au long de notre parcours universitaire.

RÉSUMÉ

Les économètres font face à une multitude de défis lorsqu'ils effectuent des prévisions. L'un d'entre eux est la quantification au moment opportun de l'incertitude entourant celles-ci. L'application récente des réseaux de neurones en macroéconomie a permis une amélioration notable des prévisions ponctuelles. Par contre, la modélisation proactive de leur volatilité reste relativement inexplorée. Dans le cadre de ce mémoire, nous proposons un *Hemisphere Neural Network* (HNN) qui estimera les processus sous-jacents aux deux premiers moments conditionnels d'une série macroéconomique. Nous évaluons le HNN dans un vaste exercice hors échantillon pour plusieurs variables à multiples horizons, le comparant à plusieurs modèles de référence provenant de la littérature en apprentissage automatique et en économétrie. Nous observons que le HNN améliore à la fois la prévision ponctuelle et la prévision de densité, notamment en tirant parti de transformations non linéaires de signaux précurseurs d'incertitude. De plus, l'utilisation de réseaux récurrents – mieux adaptés pour les séries persistantes – améliore les résultats obtenus pour les horizons plus longs.

Table des matières

REMERCIEMENTS II

RÉSUMÉ III

LISTE DES TABLEAUX VI

LISTE DES ILLUSTRATIONS IX

CHAPITRE 1 MODÉLISATION 1

1.1 Introduction 1

1.2 Littérature 2

1.3 Le modèle 7

1.4 Structure neurale 12

1.4.1 Optimisation 14

1.4.2 Mieux capturer la dépendance temporelle avec un réseau récurrent 17

CHAPITRE 2 EXERCICE EMPIRIQUE 19

2.1 Introduction 19

2.1.1 Métrique d'évaluation 22

2.2 Résultats 24

2.2.1 Prévision un an à l'avance (h=4) 28

2.2.2 Est-ce que le HRNN peut améliorer les prévisions ? 30

2.3 Importance des variables 32

2.4 Conclusion 36

CHAPITRE 3 ANNEXES 38

A Sommaire des résultats statistiques 39

B Graphiques de prévision 41

C Importance de variable 46

BIBLIOGRAPHIE 49

Liste des tableaux

TABLE 1 : Modèles de références 22
TABLE 2 : Hyperparamètres 23
TABLE 3 : HNN - Sommaire des résultats statistiques des prévisions ($h = 1$) 26
TABLE 4 : HRNN - Sommaire des résultats statistiques des prévisions ($h = 4$) 31
TABLE 5 : HNN - Sommaire des résultats statistiques des prévisions ($h = 4$) 39
TABLE 6 : HRNN - Sommaire des résultats statistiques des prévisions ($h = 1$) 40

Liste des figures

FIGURE 1 : Architecture du réseau en hémisphère. L'hémisphère en jaune modélise la moyenne conditionnelle et l'hémisphère en bleu modélise la variance conditionnelle. Les deux hémisphères partagent l'hémisphère en rouge à l'entrée du réseau. 9
FIGURE 2 : Réseau de neurones 13
FIGURE 3 : Représentation d'une couche récurrente 17
FIGURE 4 : HNN - Produit Intérieur Brut (h=1) 25
FIGURE 5 : HNN - Inflation (h=1) 27
FIGURE 6 : HNN - Mise en chantier de résidences privées (h=1) 28
FIGURE 7 : HNN - Inflation (h=4) 29
FIGURE 8 : HNN - S&P 500 (h=4) 29
FIGURE 9 : HRNN - Produit Intérieur Brut (h=4) 30
FIGURE 10 : HRNN - Taux de chômage (h=4) 32
FIGURE 11 : Importance des variables pour le PIB. Dans le panneau de gauche, on présente l'importance des variables pour l'hémisphère de la moyenne conditionnelle, et dans le panneau de droite, celle de la variance conditionnelle. 33
FIGURE 12 : Graphique des variables importantes par hémisphère pour le PIB. Dans le panneau de gauche, on trouve les variables qui ont le plus d'influence sur le taux d'inflation durant la période allant de 2007Q1 à 2019Q4 pour l'hémisphère de la moyenne conditionnelle. À droite, celles de l'hémisphère de la variance conditionnelle. 33
FIGURE 13 : Importance des variables pour l'inflation. Dans le panneau de gauche, on présente l'importance des variables pour l'hémisphère de la moyenne conditionnelle, et dans le panneau de droite, celle de la variance conditionnelle. 34

FIGURE 14 : Graphique des variables importantes par hémisphère pour l'inflation. Dans le panneau de gauche, on trouve les variables qui ont le plus d'influence sur le taux de croissance du PIB durant la période allant de 2007Q1 à 2019Q4 pour l'hémisphère de la moyenne conditionnelle. À droite, celles de l'hémisphère de la variance conditionnelle. 35
FIGURE 15 : HNN - Taux de chômage (h=1) 41
FIGURE 16 : HNN - S&P 500 (h=1) 41
FIGURE 17 : HNN - Produit Intérieur Brut (h=4) 42
FIGURE 18 : HNN - Taux de chômage (h=4) 42
FIGURE 19 : HRNN - Taux de chômage (h=1) 42
FIGURE 20 : HRNN - Inflation (h=1) 43
FIGURE 21 : HRNN - Produit Intérieur Brut (h=1) 43
FIGURE 22 : HRNN - Mise en chantier de résidence privé (h=1) 43
FIGURE 23 : HRNN - S&P 500 (h=1) 44
FIGURE 24 : HRNN - Inflation (h=4) 44
FIGURE 25 : HRNN - Mise en chantier de résidence privé (h=4) 44
FIGURE 26 : HRNN - S&P 500 (h=4) 45
FIGURE 27 : Importance des variables pour le taux de chômage. Dans le panneau de gauche, on présente l'importance des variables pour l'hémisphère de la moyenne conditionnelle, et dans le panneau de droite, celle de la variance conditionnelle. 46
FIGURE 28 : Graphique des variables importantes par hémisphère pour le taux de chômage. Dans le panneau de gauche, on trouve les variables qui ont le plus d'influence sur le taux de chômage durant la période allant de 2007Q1 à 2019Q4 pour l'hémisphère de la moyenne conditionnelle. À droite, celles de l'hémisphère de la variance conditionnelle. 46
FIGURE 29 : Importance des variables pour la mise en chantier de résidences privées. Dans le panneau de gauche, on présente l'importance des variables pour l'hémisphère de la moyenne conditionnelle, et dans le panneau de droite, celle de la variance conditionnelle. 47

FIGURE 30 : Graphique des variables importantes par hémisphère pour la mise en chantier de résidences privées. Dans le panneau de gauche, on trouve les variables qui ont le plus d'influence sur la mise en chantier de résidences privées durant la période allant de 2007Q1 à 2019Q4 pour l'hémisphère de la moyenne conditionnelle. À droite, celles de l'hémisphère de la variance conditionnelle. 47
FIGURE 31 : Importance des variables pour le S&P 500. Dans le panneau de gauche, on présente l'importance des variables pour l'hémisphère de la moyenne conditionnelle, et dans le panneau de droite, celle de la variance conditionnelle. 48
FIGURE 32 : Graphique des variables importantes par hémisphère pour le S&P 500. Dans le panneau de gauche, on trouve les variables qui ont le plus d'influence sur le S&P 500 durant la période allant de 2007Q1 à 2019Q4 pour l'hémisphère de la moyenne conditionnelle. À droite, celles de l'hémisphère de la variance conditionnelle. 48

CHAPITRE 1
MODÉLISATION

1.1 Introduction

La prévision macroéconomique est sujette à une multitude de facteurs qui peuvent influencer sa précision. Ces facteurs incluent notamment le manque de données, la complexité du modèle ainsi que sa spécification. Ces facteurs doivent absolument être pris en considération par l'économètre. Malheureusement, advenant le cas où l'on choisit une fonction de perte quadratique, la prévision ponctuelle capte seulement la moyenne de la distribution conditionnelle et ne prend pas en considération l'intervalle de confiance autour de cette moyenne. La prévision de densité, d'autre part, fait l'estimation de la distribution des valeurs possibles de la variable aléatoire. Elle prend en considération des moments supérieurs d'une distribution et permet de nous renseigner sur le niveau d'incertitude de nos prévisions. Ce sujet a été discuté par Diebold et Lopez (1996); Granger et Pesaran (2000) et Berkowitz (2001) qui précisent que l'ensemble d'informations pertinentes n'est pas entièrement capté lorsqu'on analyse seulement la moyenne conditionnelle. De plus, puisque la prévision de densité implique une analyse plus rigoureuse notamment en termes de calibration, de bonnes pratiques ont été proposées dans Diebold *et al.* (1997); Gneiting *et al.* (2007) et plus récemment dans Bassetti *et al.* (2020); Diebold *et al.* (2022).

Les récents développements en intelligence artificielle, ou plus précisément, en *deep learning*[1] (Hinton *et al.*, 2006; LeCun *et al.*, 2015; Goodfellow *et al.*, 2016) alimentés par l'essor computationnel provenant de divers milieux, ont permis à une variété de domaines d'améliorer leur capacité prédictive. En science économique, ils ont permis aux économètres d'explorer les caractéristiques profondément non-linéaires qui décrivent notre économie (Goulet Coulombe, 2022) tout en permettant d'obtenir de meilleures prévisions (Smalter Hall et Cook, 2017). Un des inconvénients des réseaux de neurones est qu'il est difficile d'identifier ce qui justifie leur prévision. Ils sont donc connus pour être une *boîte noire*[2]. Il est connu que les modèles linéaires et les méthodes à base d'arbre fournissent une meilleure interprétabilité que les modèles plus complexes tels que les réseaux de

[1] Le *deep learning* fait référence à l'apprentissage automatique faisant usage de réseaux de neurones.

[2] Une *boîte noire* est un système pour lequel on observe uniquement les entrées et les sorties.

neurones. Heureusement, plusieurs méthodes ont été proposées pour permettre d'interpréter cette famille d'algorithmes. Bien que la littérature entourant cette problématique soit relativement petite, (Montavon *et al.*, 2018) discute des derniers avancements sur le sujet.

Dans le cadre de ce mémoire, nous proposons un modèle de réseau de neurones pour prévoir les deux premiers moments de la distribution conditionnelle d'une série macroéconomique. Nous réorganisons les hémisphères comme décrit dans (Goulet Coulombe, 2022), permettant ainsi d'estimer simultanément le processus de la moyenne et de la variance conditionnelles d'une variable macroéconomique sous hypothèse que celui-ci est distribué normalement. Cette approche nous permettra non seulement de faire des prévisions, mais aussi de quantifier l'incertitude entourant ces prévisions. Nous comparons deux types de réseaux: un réseau de neurones simple et un réseau récurrent, qui a attiré beaucoup d'attention ces dernières années pour les prévisions. Nos résultats indiquent que notre modèle est capable d'identifier l'hétéroscédasticité plus rapidement que les modèles standard en économétrie et de fournir des prévisions plus précises pour des séries macroéconomiques. De plus, nous utilisons un algorithme qui nous permet de déterminer l'importance de la contribution de chaque variable à la prévision obtenue. Bien que les réseaux de neurones soient connus pour leur opacité, cet algorithme suggère que notre modèle a pu identifier des relations en accord avec la théorie macroéconomique. Avant d'aller plus loin dans la discussion des résultats, abordés dans la deuxième partie de ce mémoire, nous présentons les travaux importants qui justifient le développement de notre modèle.

1.2 Littérature

Il y a 40 ans, les modèles paramétriques de type *Autoregressive Conditional Heteroskedasticity* (ARCH) (Engle, 1982) et *Generalized Autoregressive Conditional Heteroskedastic* (GARCH) (Bollerslev, 1986) ont déclenché une avalanche d'extensions dans l'estimation de la volatilité. Les spécifications telles que *Nonlinear Asymetric* (NAGARCH), *Integrated* (IGARCH), *Exponential* (EGARCH), *Threshold* (T-GARCH) - pour n'en nommer que quelques-unes - ont été conçues pour capter différents faits stylisés de la volatilité. Ces modèles ont attiré beaucoup d'attention, en particulier dans le secteur financier où la valeur d'un actif dépend fortement de son risque. Une autre classe de modèle estime la volatilité de manière stochastique. Dans ce cas, la volatilité est estimée par un processus autorégressif latent. Cette spécification est mieux connue sous le nom de *stochastic*

volatility (SV) (Taylor, 1982; Hull et White, 1987; Jacquier *et al.*, 2002).

Un des inconvénients des modèles mentionnés ci-dessus est que la volatilité estimée est simplement le résultat des résidus au carré provenant d'une moyenne conditionnelle qui est utilisée comme filtre. Intuitivement, GARCH et SV utilisent la volatilité de la période précédente pour prévoir celle d'aujourd'hui. En d'autres termes, leur nature autorégressive ne fait que propager des chocs à des périodes ultérieures. Ce comportement les rend moins adaptés pour la prévision, car ils sont réactifs. Ces deux modèles ont également l'inconvénient d'être difficiles à estimer, car ils sont optimisés numériquement. Si le modèle GARCH est estimé en utilisant le maximum de vraisemblance, un problème de convergence peut survenir s'il est mal spécifié. Dans le cas du modèle SV, son optimisation peut prendre beaucoup de temps, car elle requiert plusieurs échantillonnages à la *Markov Chain Monte Carlo* (MCMC) (Kroese et Rubinstein, 2012). Néanmoins, les deux modèles ont posé les bases de la littérature sur l'estimation de l'hétéroscédasticité dans un contexte de série temporelle.

Au lieu d'être réactif, on peut mettre l'accent sur la modélisation de la distribution pour être proactif. Par exemple, Adams *et al.* (2021) ont étendu les travaux d'Adrian *et al.* (2019) qui utilisent la régression quantile pour prévoir la distribution d'une prévision. Les auteurs mettent l'accent sur le fait que les quantiles les plus petits de la distribution sont plus sensibles aux conditions économiques actuelles. En outre, ils soutiennent que les banques centrales sont davantage concernées par le *downside risk* de certains évènements et donc, les plus petits quantiles de la distribution sont plus d'intérêt. D'autres approches telles que le modèle de changement de régime (*Markov Switching Model*) de Hamilton (1989) ont été employées pour capter les changements dans le processus conditionnel dépendamment des conditions macroéconomiques actuelles (Caldara *et al.*, 2021).

***Deep Learning* et prévision**. L'apprentissage profond est de plus en plus populaire dans une variété de milieux en raison de sa capacité à traiter des données alternatives[3]. La nature non linéaire des réseaux de neurones, combinée à sa spécification non paramétrique qui implique une grande quantité de paramètres, conduit à un système complexe et particulièrement opaque. Cependant, ceux-ci sont dotés d'une grande flexibilité et leur rapidité d'optimisation est désormais possible grâce aux récents développements dans le secteur des jeux vidéo.

[3] Les données alternatives sont des données qui ne sont pas obtenues de façon conventionnelle, comme des images ou du texte, ou toute autre structure qui n'est pas représentée dans une base de données traditionnelle.

En science économique, la littérature entourant l'utilisation des réseaux de neurones pour faire de l'inférence en science économique remonte, étonnamment, jusqu'à White (1989). Plus récemment, Farrell *et al.* (2021) a exploré une nouvelle méthodologie pour effectuer de l'inférence semi-paramétrique avec un réseau de neurones. Encore plus récemment, (Goulet Coulombe, 2022) a proposé une architecture de réseau en hémisphères où chacune d'entre elles est un bloc de neurones indépendant qui modélise une composante de la courbe de Phillips. Ces composantes sont représentées sous forme de facteur latent qui sont ensuite combinés linéairement afin d'estimer la courbe de Phillips. Cette idée de conception est intéressante puisqu'elle permet d'alléger le fardeau de la complexité de cette famille d'algorithmes tout en produisant des prévisions intéressantes d'un point de vue macroéconomique.

Du côté de la prévision macroéconomique, les réseaux de neurones ont été comparés à plusieurs reprises aux modèles linéaires dans le passé (Swanson et White, 1997; Stock et Watson, 1998; Medeiros *et al.*, 2006; Moshiri et Cameron, 2000), mais sans grandes promesses. Néanmoins, pour certaines variables telles que l'inflation, on peut obtenir de meilleurs résultats (Nakamura, 2005). Plus récemment, Smalter Hall et Cook (2017) ont testé différents types de réseaux de neurones pour leur capacité à prévoir le taux de chômage. Leurs résultats démontrent que, malgré la petite quantité de données macroéconomiques disponibles, certaines structures de neurones sont en mesure de battre l'indice de référence *Survey of Professional Forecaster* (SPF)[4]. En effet, les travaux récents dans le développement de structures de neurones plus sophistiquées ont permis d'obtenir des prévisions plus précises. Paranhos (2021) démontre que les LSTM[5] sont de puissants algorithmes capables de prévoir l'inflation avec davantage de précision qu'un réseau simple. Plus récemment encore, Almosova et Andresen (2023) démontrent que cette même structure de neurones permet d'obtenir des prévisions qui surpassent celles des modèles autorégressifs (AR) ou naïfs tels que la marche aléatoire.

Prévision de densité. L'analyse de l'incertitude entourant la prévision remonte au moins à Box et Jenkins (1970) qui précise que chaque prévision doit-être accompagnée d'un indicateur d'incertitude. En pratique, elle vise à faire la prévision de la distribution de probabilité d'une variable aléatoire

[4] L'Enquête des prévisionnistes professionnels est une enquête trimestrielle sur les prévisions macroéconomiques pour l'économie des États-Unis publiée par la Banque fédérale de réserve de Philadelphie.

[5] Long Short-Term Memory est une structure de neurones pour les réseaux récurrents qui permet de filtrer l'information à préserver en mémoire par le réseau.

au lieu de prévoir une seule valeur. Cette approche est utilisée dans une grande variété de milieux, notamment la météorologie, la finance, l'économie ainsi que la consommation d'énergie. Au départ, la prévision par intervalles était la principale méthodologie utilisé pour estimer l'incertitude. Cette dernière est largement discutée dans Chatfield (1993). L'inconvénient des prévisions de densité est qu'elles peuvent être plus difficiles à interpréter en raison du fait qu'elles modélisent une distribution entière. En revanche, elles permettent d'obtenir une estimation de l'incertitude et donc favorisent une prise de décision plus nuancée. Cela est particulièrement important dans des situations où le coût d'une mauvaise décision pourrait être dommageable (Delle Monache *et al.*, 2020). À la lumière des récents événements concernant la Covid-19, les séries macroéconomiques et financières sont devenues particulièrement difficiles à prévoir, ce qui augmente l'intérêt à prévoir l'intervalle dans lequel une valeur prédite pourrait se retrouver (Guidolin *et al.*, 2021). On peut estimer l'incertitude sous l'hypothèse que la distribution de l'erreur est normale et donc symétrique. Il est possible de relâcher cette hypothèse en calculant l'intervalle de confiance avec ses quantiles conditionnels Goulet Coulombe *et al.* (2022). Cela permet de nous renseigner sur le niveau d'asymétrie qu'une variable macroéconomique pourrait subir suite à un choc.

La prévision de densité comporte certains défis, notamment au niveau computationnel ainsi que l'utilisation d'hypothèses restrictives. Heureusement, plusieurs méthodes ont été proposées pour surmonter ces problèmes. L'utilisation de la méthode bootstrap (Tibshirani et Efron, 1993) et de la statistique bayésienne (Gelman *et al.*, 1995) permettent de créer un environnement flexible qui incorpore des connaissances et des incertitudes préalables dans le modèle. Il est bon de mentionner que l'évaluation de la prévision de densité est particulièrement importante dans la littérature. Elle implique notamment l'analyse d'une métrique nommée *scoring rule*[6] (Gneiting et Raftery, 2007; Diebold *et al.*, 2022), qui doit être adaptée au problème en question. D'autres éléments, tels que l'analyse de la couverture nominale, le *Probability integral transform* (Diebold *et al.*, 1997) ainsi que le critère d'information *Kullback-Leibler* (Bao *et al.*, 2007), ont été proposés dans le but d'évaluer la calibration de la distribution conditionnelle.

Plusieurs modèles d'apprentissage automatique ont été proposés récemment dans le but d'obtenir des prévisions de l'incertitude. Récemment, (Salinas *et al.*, 2020) ont proposé le *DeepAR*, un réseau

[6] Les *scoring rules* sont utilisées pour mesurer l'exactitude ou la qualité des prédictions probabilistes. Plus précisément, elles mesurent la différence entre la distribution de probabilité prédite et la vraie distribution du résultat observé.

récurrent autorégressif entraîné à faire des prévisions en utilisant la log-vraisemblance d'une loi normale. Les auteurs ont démontré qu'en entraînant l'algorithme sur plusieurs séries chronologiques similaires, ils sont en mesure d'obtenir des prévisions significativement plus précises que les modèles à la fine pointe de l'art. De plus, leur fonction objectif leur permet d'obtenir une estimation de l'incertitude autour de leur prévision. Même si leur application est sur des données à très hautes fréquences, leur cadre d'analyse mérite d'être exploré dans différents environnements.

Les réseaux de neurones ne sont pas les seuls algorithmes utilisés pour faire la prévision de densité. Récemment, (Clark *et al.*, 2022) ont utilisé les *Bayesian Additive Regression Trees* (BART) de (Chipman *et al.*, 2010) pour effectuer des prévisions de variables macroéconomiques. Globalement, leur modèle démontre beaucoup de flexibilité dans la modélisation des probabilités dans les queues de la distribution.

Le problème de la quantification de l'incertitude. La littérature entourant la prévision de la volatilité avec un réseau de neurones est relativement jeune puisqu'elle fait face à une multitude de problèmes en lien avec son estimation. Bien que l'apprentissage profond ait su démontrer une bonne performance dans plusieurs applications, la prévision de l'incertitude semble être dans leur angle mort. En effet, ceux-ci ont tendance à produire des prévisions surconfiantes (Amodei *et al.*, 2016; Lakshminarayanan *et al.*, 2017; Gawlikowski *et al.*, 2021). On définit une prévision surconfiante lorsque le modèle fait une large erreur de prévision accompagné d'une très petite volatilité. Naturellement, les résultats s'apparentent à une mauvaise estimation de la distribution hors échantillon et à une calibration de mauvaise qualité. Le sujet est discuté en profondeur dans Guo *et al.* (2017). Évidemment, la nature opaque de ces algorithmes rend difficile la détermination de la cause derrière ces résultats. Elle peut provenir autant de la spécification du modèle que de la structure des données sous-jacente. Néanmoins, Luo *et al.* (2018) et Yin et Barucca (2022) proposent de modéliser l'hétéroscédasticité en tirant parti, des réseaux récurrents. Les deux papiers modélisent la distribution des paramètres avec un réseau génératif avant d'effectuer une prévision avec un réseau complémentaire et ils ont démontré qu'il est possible de faire mieux que les modèles GARCH et SV en suivant cette méthodologie.

Afin de pouvoir produire un signal d'incertitude qui ne disparaît pas avec la surconfiance de l'algorithme, l'utilisation de régularisations devient un ingrédient très important. Une des tech-

niques de régularisation les plus populaires pour les réseaux de neurones est appelée *Dropout* (Srivastava *et al.*, 2014; Hinton *et al.*, 2012). Cette régularisation fonctionne de la façon suivante : À chaque début d'itération, toutes les neurones (incluant celles à l'entrée du réseau) ont une probabilité p d'être ignorées durant cette itération seulement. Cela force celles qui ne sont pas ignorées à compenser pour celles qui le sont durant la modification des paramètres. Une autre méthode consiste à ajouter un terme de pénalité ℓ_1 et ℓ_2 à la fonction objectif durant sa minimisation. Cela permet d'obtenir les mêmes effets que les modèles *Least Absolute Shrinkage and Selection Operator* (LASSO) (Tibshirani, 1996) et *Ridge regression* (Hoerl et Kennard, 1970). D'autres méthodes de régularisation sont mentionnées à travers la littérature sur la modélisation de l'incertitude telles que l'utilisation de *Batch Normalization* (Ioffe et Szegedy, 2015) et *adversarial training* (Goodfellow *et al.*, 2014).

1.3 Le modèle

Le modèle proposé exploite l'avantage de la flexibilité et de la non-linéarité des réseaux de neurones pour effectuer la prévision de densité de séries temporelles. Nous utilisons un *Hemisphere Neural Network* (HNN) tel que proposé par Goulet Coulombe (2022) qui force chaque processus à être traité de façon idiosyncratique. De cette façon, nous pouvons personnaliser le flux de transformation des régresseurs avec l'objectif d'obtenir de meilleurs résultats et de favoriser son interprétabilité.

Fonction objective. Pour notre modèle, les deux processus conditionnels sont obtenus en maximisant leur log vraisemblance sous hypothèse de normalité (Salinas *et al.*, 2020). Supposons une variable aléatoire avec une moyenne conditionnelle $\hat{y}_t$ et un écart-type conditionnelle $\hat{\sigma}_t$ à estimer à travers le temps. La fonction de densité de probabilité s'écrit.

$$p(y_t; \hat{y}_t, \hat{\sigma}_t) = \frac{1}{\hat{\sigma}_t \sqrt{2\pi}} \exp\left(-\frac{(y_t - \hat{y}_t)^2}{2\hat{\sigma}_t^2}\right) \tag{1.1}$$

Prendre le logarithme facilite sa dérivation:

$$\ln(p(y_t)) = -\frac{1}{2}\ln(2\pi) - \ln(\hat{\sigma}_t) - \frac{1}{2\hat{\sigma}_t^2}(y_t - \hat{y}_t)^2. \tag{1.2}$$

Si nous avons T observations indépendantes, la vraisemblance est:

$$(p(y_1))(p(y_2))...(p(y_t)) = \prod_{t=1}^{T} p(y_t) \tag{1.3}$$

et la log-vraisemblance devient:

$$\ln\left(\prod_{t=1}^{T} p(y_t)\right) = \sum_{t=1}^{T} \ln(p(y_t)). \tag{1.4}$$

À partir de (1.2) on retrouve la fonction de log-vraisemblance

$$\ln\left(\prod_{t=1}^{T} p(y_t)\right) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\frac{y_t - \hat{y}_t}{\hat{\sigma}_t}\right)^2 - \sum_{t=1}^{T}\ln(\hat{\sigma}_t) \tag{1.5}$$

Puisque le premier terme n'affecte pas la valeur de l'optimum, on peut l'éliminer sans perte de généralité. Finalement, on obtient la fonction de perte suivante:

$$L(\hat{y}_t, \hat{\sigma}_t; w_m, w_v) = -\sum_{t=1}^{T}\left(\frac{y_t - \hat{y}_t}{\hat{\sigma}_t}\right)^2 - \sum_{t=1}^{T}\ln(\hat{\sigma}_t) \tag{1.6}$$

Notre objectif sera de minimiser l'équation (1.6) en calculant les paramètres du réseau w_m et w_v représentant la sortie de l'hémisphère de la moyenne conditionnelle $\mathtt{H}_m$ et celle de la variance conditionnelle $\mathtt{H}_v$ respectivement:

$$\min_{w_m, w_v} -\sum_{t=1}^{T}\left(\frac{y_t - \mathtt{H}_m(\mathbf{X}_{t-1}; w_m)}{\mathtt{H}_v(\mathbf{X}_{t-1}; w_v)}\right)^2 - \sum_{t=1}^{T}\ln(\mathtt{H}_v(\mathbf{X}_{t-1}; w_v)) \tag{1.7}$$

Où p est le nombre de retards utilisés par le modèle.

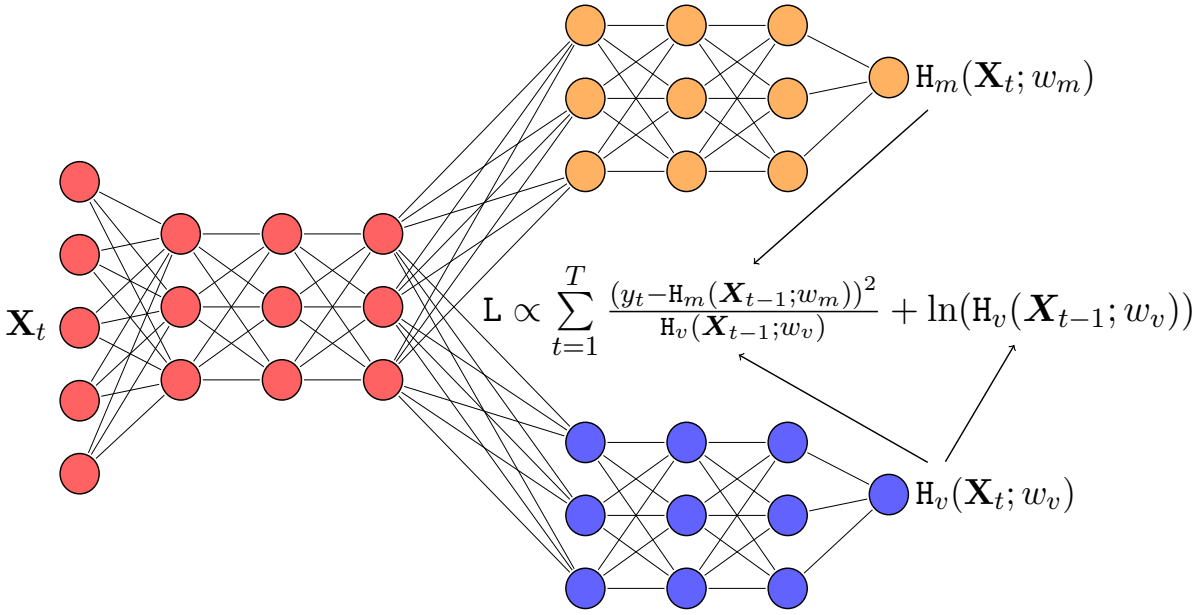

Figure 1: Architecture du réseau en hémisphère. L'hémisphère en jaune modélise la moyenne conditionnelle et l'hémisphère en bleu modélise la variance conditionnelle. Les deux hémisphères partagent l'hémisphère en rouge à l'entrée du réseau.

Nous transformons la sortie de l'hémisphère de volatilité en utilisant la fonction d'activation *softplus*:

$$f(x) = \ln(1 + e^x) \tag{1.8}$$

De cette façon on s'assure que la volatilité estimée est positive. Notre modèle est constitué de 4 ingrédients qui nous permettent d'obtenir des résultats qui sont à la fine pointe de l'art de la prévision.

Ingrédient 1 : Hémisphère commune. Les hémisphères de la moyenne et de la variance conditionnelle partagent un bloc de couches de neurones dans la première partie du réseau tel que présenté à la Figure 1. L'intuition est la suivante : dans les modèles ARCH et SV, la prévision de la moyenne et de la variance conditionnelle implique l'estimation de paramètres qui sont partagés à travers leurs équations respectives. Dans le modèle ARCH, on a :

$$y_t = X_t'\beta + \epsilon_t$$
$$\sigma_t = c + \sum_{k=1}^{K} \alpha_k \underbrace{\epsilon_{t-k}^2}_{y_{t-k} - X_{t-k}'\beta}$$

On observe que β est partagé à travers les deux équations. Pour cette raison, il est logique d'estimer les deux processus conjointement (Gouriéroux, 1997). Notre réseau contraint les deux processus conditionnels à partager la même information latente qui est obtenue par l'hémisphère commune. Cette hémisphère a pour rôle d'extraire l'information pertinente provenant des régresseurs, qui sera ensuite transférée aux deux autres.

Ingrédient 2 : L'emphase sur la volatilité. Comme mentionné précédemment, modéliser l'incertitude peut être un véritable défi lorsqu'on utilise des réseaux de neurones. C'est pourquoi la régularisation est un ingrédient important dans notre cas. Nous régularisons l'algorithme en spécifiant l'importance que le modèle accorde à la volatilité. Intuitivement, certaines séries, telles que le rendement de l'indice boursier S&P 500, ont une moyenne conditionnelle moins persistante, donc modéliser leur variance conditionnelle pourrait être plus informatif. À l'inverse, des séries telles que l'inflation et le PIB, sont plus persistantes. Modéliser leur moyenne conditionnelle est beaucoup plus facile que la volatilité. Afin de pouvoir contrôler l'importance que l'algorithme accorde à la modélisation de l'incertitude, nous devons effectuer quelques transformations à la sortie de son hémisphère. Tout d'abord, nous divisons celle-ci par sa moyenne pour obtenir une volatilité normalisée à 1, notée $\bar{\mathtt{H}}_{v,t}$. Ensuite, nous multiplions $\bar{\mathtt{H}}_{v,t}$ par un hyperparamètre $\boldsymbol{\lambda} \in]0, 1]$ déterminé préalablement, qui va fixer sa magnitude moyenne.

$$\mu_v = \frac{1}{T} \sum\nolimits_{t=1}^{T} \mathtt{H}_{v,t} \tag{1.9}$$

$$\bar{\mathtt{H}}_{v,t} = \frac{\mathtt{H}_{v,t}}{\mu_v}. \tag{1.10}$$

Notre fonction objective à minimiser devient:

$$\min_{w_m, w_v} \sum_{t=1}^{T} \left(\frac{y_t - \mathtt{H}_m(X_{t-1}; w_m)}{\boldsymbol{\lambda} \bar{\mathtt{H}}_v(X_{t-1}; w_v)} \right)^2 + \sum_{t=1}^{T} \ln(\boldsymbol{\lambda} \bar{\mathtt{H}}_v(X_{t-1}; w_v)). \tag{1.11}$$

Intuitivement, on fixe la volatilité moyenne à un certain niveau et on laisse le réseau apprendre les déviations de cette moyenne. Puisque $\boldsymbol{\lambda}$ est un hyperparamètre, il doit être sélectionné avant d'entraîner l'algorithme. Malheureusement, l'économètre n'a pas toujours connaissance de l'importance que la volatilité aura dans le modèle. Au lieu de choisir aveuglément sa valeur, on peut utiliser le *Mean Squared Error* (MSE) de la prévision *out-of-bag* (OOB) d'un ensemble de réseaux[7] de neurones estimés préalablement comme valeur pour $\boldsymbol{\lambda}$. Le raisonnement est qu'estimer l'erreur OOB pourrait nous renseigner sur l'importance que la variance aura pour le HNN.

Ingrédient 3: Recalibration. Pour des séries macroéconomiques persistantes telles que l'inflation, $\boldsymbol{\lambda}$ sera très petit. Par conséquent, sa volatilité sera réduite de façon significative et nous donnera une couverture nominale sous-évaluée. Pour recalibrer la volatilité, nous exécutons les étapes décrites dans l'Algorithme 1. Celui permet, notamment, d'obtenir une bonne calibration en échantillon et, espérons-le, hors-échantillon.

Ingrédient 4: *Agrégation bootstrap*. Pour inférer de manière adéquate sur l'estimation des moments conditionnels, on peut utiliser la méthode du *bootstrap* (Efron, 1979) pour estimer plusieurs estimateurs individuellement entraînés sur un sous-échantillon du vecteur de régresseurs X_t. Ensuite, on agrège les prévisions de ces estimateurs pour obtenir une prévision finale. Si l'échantillonnage est fait avec remplacement, on appelle cette méthode *Bootstrap Aggregating* (*Bagging*) (Breiman, 1996). Les poids $\theta^{(b)}$ estimés par le réseau b seront fonction de la distribution du sous-échantillon obtenu par le $bootstrap^{(b)}$. D'un point de vue bayésien, cela nous permet d'établir une distribution sur l'ensemble des paramètres estimés. Il est pertinent d'utiliser cette méthode puisque la prévision de densité obtenue par l'agrégation des estimateurs dépend de la distribution de X_t et celle de θ. Cela nous fournit une bonne inférence pour la prévision de la densité.

Dans un contexte de série chronologique, étant donnée la dépendance temporelle entre les observations, l'échantillonnage ne peut pas être fait de la même façon que sur des données qui n'ont pas de dépendance entre elles. Pour cette raison, nous effectuons un échantillonnage en blocs de 2 ans (Goulet Coulombe, 2022). Le *bootstrap* est effectué en échantillonnant un pourcentage $\boldsymbol{\tau}$ de l'échantillon total d'entraînement qui sera utilisé pour l'estimation des paramètres du réseau, et $(1 - \boldsymbol{\tau})$ sera utilisé pour la prévision OOB. Nous effectuons un total de $\boldsymbol{B}$ *bootstraps*, menant à

[7] La prévision OOB et l'utilisation d'un ensemble de réseau seront abordés à l'ingrédient 4.

Algorithme 1 Ajustement de la volatilité
Cet algorithme ajuste la volatilité estimée par le réseau de neurones en utilisant la procédure d'estimation par les moindres carrés pondérés (Wooldridge, 2015). En pratique, l'algorithme est appliqué sur la prévision *out-of-bag* de l'ensemble des estimateurs et ensuite, le paramètre Φ vient multiplié les données hors-échantillon.

Requiert y_t La série d'intérêt.
Requiert $\hat{y}_t$ La moyenne conditionnelle estimée par le modèle.
Requiert $\hat{\sigma}_t$ La variance conditionnelle estimée par le modèle.

On récupère les résidus dénotée $\hat{u}$.

$$\hat{u}_t = y_t - \hat{y}_t$$

On régresse $\log(\hat{u}_t^2)$ sur le log de la volatilité estimée au carré $\log(\hat{\sigma}_t^2)$ pour obtenir la projection $\hat{g}_t$.

$$\log(\hat{u}_t^2) = \alpha + \beta \log(\hat{\sigma}_t^2) + \epsilon_t$$

$$\hat{g}_t = \hat{\alpha} + \hat{\beta} \log(\hat{\sigma}_t^2)$$

On calcule les résidus $\hat{v}_t$ de la régression précédente.

$$\hat{v}_t = \log(\hat{u}_t^2) - \hat{g}_t$$

On ajuste la volatilité

$$\Phi = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \exp(\hat{v}_t)}$$

$$\hat{\sigma}_t = \sqrt{\exp(\hat{g}_t)} \cdot \Phi$$

cette même quantité d'estimateurs avec une architecture identique. Pour bénéficier au maximum du hasard, nous initialisons les paramètres θ de chacun des réseaux indépendamment en utilisant une loi normale :

$$\theta \sim \mathtt{N}(0, \Sigma) \tag{1.12}$$

pour chaque estimateur b, l'équation générale pour calculer la prévision OOB pour l'hémisphère i s'écrit:

1.4 Structure neurale

Les réseaux de neurones artificiels sont des algorithmes d'apprentissage automatique qui s'inspirent des réseaux de neurones biologiques présents dans notre cerveau. Bien que leur fonctionnement

soit différent, leur architecture, basée sur des couches de neurones interconnectés, présente de nombreuses similitudes avec le cortex cérébral. Étonnamment, cette idée de réseau de neurones artificiel remonte à une longue période. En effet, dès 1943, (McCulloch et Pitts, 1943) ont proposé un modèle simplifié de l'activation neuronale, visant à reproduire les interactions observées entre les neurones biologiques.

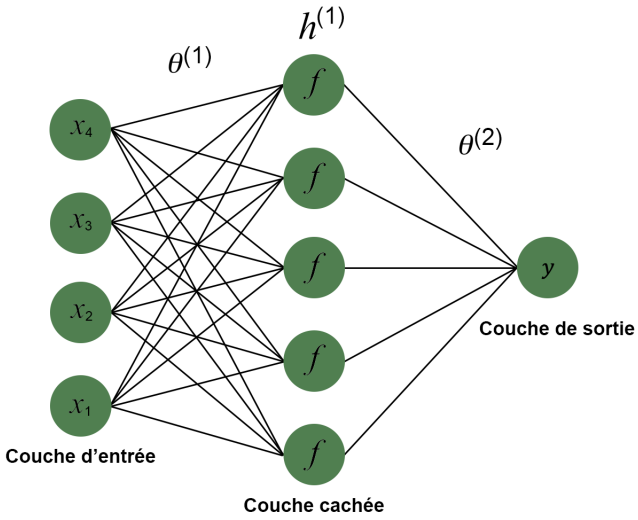

Figure 2: Réseau de neurones
Notes: Les régresseurs $x_{1:4}$ sont transformés dans la couche du milieu en fonction de f. La couche de sortie agrège la couche du milieu en un vecteur qui approxime y.

Entre les années 60 et le début des années 2000, les réseaux de neurones ont perdu de leur attrait face à l'arrivée d'algorithmes d'apprentissage automatique plus simples et facilement interprétables. Cependant, ces dernières années, l'augmentation de la quantité de données disponibles ainsi que la puissance de calcul ont permis aux réseaux de neurones de surpasser les modèles précédents sur des tâches complexes. Les réseaux de neurones sont composés d'une couche d'entrée pour les régresseurs, d'une ou plusieurs couches cachées qui transforment les prédicteurs de façon non-linéaire avec une fonction d'activation, et d'une couche de sortie qui agrège ces transformations en prévision. Les neurones stockent les prédicteurs transformés et sont connectés à tous les neurones de la couche précédente par un paramètre à estimer. Cependant, le nombre de paramètres à estimer augmente rapidement avec le nombre de neurones et de couches, ce qui rend ces algorithmes difficiles à interpréter.

Théorème d'approximation universelle. Le théorème d'approximation universelle Cybenko

(1989); Hornik *et al.* (1990) stipule qu'avec une seule couche cachée et un nombre suffisant de neurones, il est possible d'approximer n'importe quelle fonction continue avec n'importe quel degré de précision quelconque. Pour que ce théorème s'applique, la fonction d'activation doit être non constante, bornée et continue. Bien évidemment, une fonction linéaire ne remplit pas ces exigences, car elle ne possède aucune borne. En pratique, les fonctions d'activation sont souvent non linéaires et constituent un hyperparamètre important pour l'algorithme. Chaque couche cachée ainsi que la couche de sortie possèdent leur propre fonction d'activation, notée f, qui utilise les informations stockées dans leurs neurones, notés x, et les transforme par rapport à cette fonction.

1.4.1 Optimisation

Optimiser les paramètres d'un réseau de neurones n'est pas une tâche facile. Contrairement aux modèles paramétriques, les réseaux de neurones, et en particulier les réseaux profonds, ont une quantité de poids[8] qui peuvent s'étendre de seulement quelques centaines à plusieurs millions. Le fameux robot de clavardage *ChatGPT* contient un total invraisemblable de 175 milliards de paramètres. Le lecteur pourrait se demander comment il est possible d'estimer une telle quantité de paramètres sans rencontrer des problèmes d'ordre computationnel ou même statistique. En effet, leur entraînement a été un sujet de recherche abordé par de nombreux chercheurs depuis leur introduction. C'est en 1986 que la méthode du *backpropagation* a été introduite par Rumelhart *et al.* (1986b). Les auteurs ont démontré qu'en seulement deux étapes, l'algorithme est en mesure de calculer le gradient de l'erreur du réseau par rapport à chacun de ses paramètres. Cela permet de déterminer de combien chaque poids doit être modifié pour pouvoir réduire cette erreur. Cette dernière étape est mieux connue sous le nom de *Gradient Descent*.

Gradient Descent. Le *Gradient Descent* (GD) (Cauchy *et al.*, 1847) est une méthode d'optimisation où les paramètres sont modifiés itérativement dans le but de minimiser une fonction. Supposons une fonction $y = f(x)$ et que $f'(x) = \frac{\partial y}{\partial x}$ évalué en x nous donne la valeur de la pente de la fonction. Si l'objectif est de minimiser $f(x)$, dans ce cas, il suffit de faire un petit pas ϵ dans la direction opposée du signe de la dérivée, tel que $f(x - \epsilon \cdot sign(f'(x)))$, tant que nous savons que sa valeur est inférieure à $f(x)$.

Supposons une fonction de perte $L(\theta)$ avec le vecteur de paramètres à estimer θ. En calculant les

[8] Dans le jargon de l'apprentissage profond, les poids font référence aux paramètres à estimer.

dérivées individuellement, on retrouve le vecteur gradient $\nabla_{L(\theta)}$:

$$\nabla_{L(\theta)} = \begin{bmatrix} \frac{\partial L(\theta)}{\partial \theta_0} \\ \frac{\partial L(\theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial L(\theta)}{\partial \theta_k} \end{bmatrix} \tag{1.13}$$

La modification des paramètres pour chaque itération i devient :

$$\theta^{(i+1)} = \theta^{(i)} - \epsilon \nabla_{L(\theta^{(i)})} \tag{1.14}$$

où ϵ est le paramètre appelé *learning rate*[9]. Puisque le vecteur gradient est calculé pour les T itérations, on obtient la procédure définie dans l'Algorithme 2.

Algorithme 2 Gradient Descent L'algorithme du gradient descent minimise une fonction de perte itérativement.

Requiert Initialiser ϵ à un nombre positif et très petit, où ϵ est le paramètre de *learning rate* (LR).
Initialiser θ aléatoirement.
for $t = 1, ..., T$ **do**
$\theta_t \leftarrow \theta_{t-1} - \epsilon \nabla_{L(\theta_{t-1})}$
end for

Dans cet algorithme, la fonction de perte $L(\theta)$ est minimisée itérativement en ajustant les paramètres du modèle θ en fonction du vecteur gradient $\nabla_{L(\theta)}$. Le paramètre de taux d'apprentissage ϵ contrôle la taille des pas à chaque itération dans la direction opposée du gradient. L'algorithme commence avec une initialisation aléatoire des paramètres θ et se poursuit jusqu'à ce qu'un nombre fixé d'itérations T soit atteint.

Early Stopping. Bien que les réseaux de neurones soient capables de retracer des relations dans des structures de données complexes, le théorème d'approximation universelle les rend particulièrement vulnérables au sur-ajustement. Le sur-ajustement se produit lorsque l'algorithme mémorise les données d'entraînement plutôt que d'apprendre les relations caractérisant les données sous-jacentes. Ainsi, au lieu de choisir un nombre arbitraire d'itérations T pour le GD, on peut forcer l'arrêt de l'entraînement lorsque $L(\theta)$ ne s'améliore pas sur des données

[9] Dans l'apprentissage automatique et les statistiques, le taux d'apprentissage est un paramètre de réglage dans un algorithme d'optimisation qui détermine la taille du pas à chaque itération tout en se déplaçant vers un minimum d'une fonction de perte

qui n'ont pas été vues par l'algorithme[10] durant un nombre d'itérations défini préalablement. En arrêtant l'entraînement prématurément, on empêche le modèle de surapprendre les données d'entraînement en le forçant à apprendre des relations plus généralisables. Cet hyperparamètre est connu sous le nom d'arrêt précoce et est très important dans l'optimisation des réseaux de neurones.

Back Propagation (Forward). Les réseaux de neurones fonctionnent de la manière suivante : les entrées traversent le réseau de gauche à droite et subissent une transformation habituellement non linéaire à chaque couche de neurones cachée. Ces transformations successives sont effectuées par des fonctions d'activation notées f. En d'autres termes, chaque couche de neurones est une fonction de la couche précédente, et chaque neurone d'une couche est directement connecté à tous les neurones de la couche suivante par un ensemble de poids à estimer. Cette mécanique est connue sous le nom de *Feed Forward pass*. Analytiquement, on a $y = f(x;\theta)$ et on cherche à estimer le vecteur de paramètres θ tel que $f = f^*$. En premier lieu, supposons que notre réseau ne contient pas de couches cachées. Si la fonction d'activation de la couche de sortie est une fonction linéaire de la couche d'entrée, on obtiendrait les mêmes résultats qu'une régression linéaire, soit $y = x'\theta$. Supposons maintenant que notre réseau contienne une couche cachée, notée c. On a :

$$c = f^{(1)}\left(x;\theta^{(1)}\right) \tag{1.15}$$

$$y = f^{(2)}\left(c;\theta^{(2)}\right) \tag{1.16}$$

Où $f^{(1)}$ et $f^{(2)}$ sont les fonctions d'activation de la couche cachée et de la couche de sortie respectivement, et $\theta^{(1)}$ et $\theta^{(2)}$ sont leur vecteur de paramètres à estimer. Si l'on cherche à associer x à y en trouvant les paramètres optimaux $\theta^{(1)}$, $\theta^{(2)}$, on obtient l'équation :

$$f\left(x;\theta^{(1)},\theta^{(2)}\right) = f^{(2)}\left(f^{(1)}(x;\theta^{(1)});\theta^{(2)}\right) \tag{1.17}$$

L'équation générale pour un nombre de couches total k est :

$$f(x;\theta) = f^{(k)}\left(f^{(k-1)}\left(...f^{(1)}\left(x;\theta^{(1)}\right);\theta^{(2)}\right);...;\theta^{(k)}\right) \tag{1.18}$$

Back Propagation (Backward). La modification des paramètres d'un réseau de neurones se fait

[10]En pratique, on calcule $L(\theta)$ sur les données de validation qui nous donnent un aperçu de la performance hors échantillon du modèle

dans un ordre prédéfini. Après avoir obtenu une estimation de l'erreur provenant de la fonction de perte, il est nécessaire de calculer la contribution à cette erreur pour chaque paramètre de la couche de sortie. Ensuite, l'algorithme calcule la contribution à cette erreur pour la couche précédente en utilisant la dérivation en chaîne. Cette étape est effectuée pour chaque couche cachée allant dans le sens inverse jusqu'à ce qu'on atteigne la couche d'entrée. Une fois que le gradient de l'erreur pour tous les paramètres du réseau a été obtenu, l'algorithme effectue une étape d'optimisation par *gradient descent* et modifie les paramètres de manière à minimiser notre fonction de perte. Cette étape est connue sous le nom de *backward pass*.

1.4.2 Mieux capturer la dépendance temporelle avec un réseau récurrent

Les *Recurrent Neural Networks* (RNN) (Rumelhart *et al.*, 1986a) sont des types de réseaux de neurones conçus pour traiter des données à caractère séquentiel. Tandis que les réseaux de neurones simples présentés précédemment peuvent également prédire des variables dépendantes du temps, les réseaux récurrents ont l'avantage de pouvoir s'adapter à des séquences beaucoup plus longues. Cela permet notamment de capturer des dépendances non linéaires qui peuvent s'étendre sur de très longues périodes.

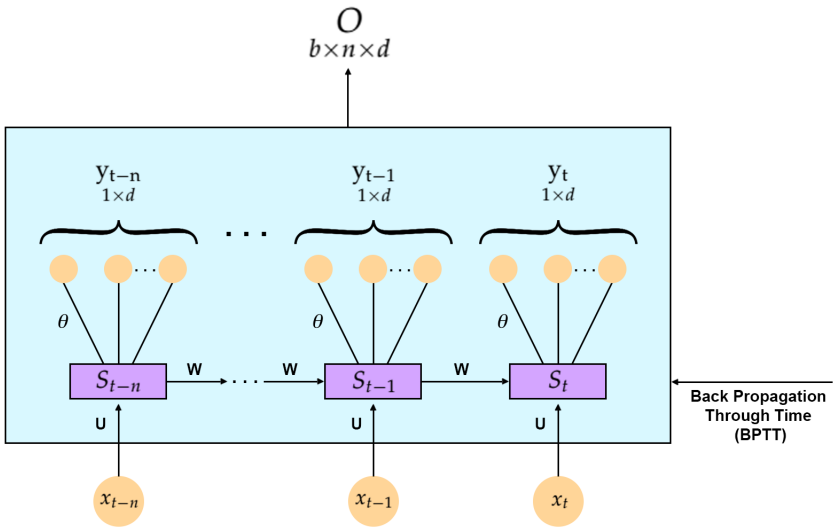

Figure 3: Représentation d'une couche récurrente

Notes: Une couche récurrente reçoit des données en 3 dimensions. Chaque période de la séquence est traitée indépendamment dans un état interne ($\boldsymbol{S}$). La dynamique entre chaque état est ensuite estimée par BPTT. À la sortie d'une couche récurrente, on retrouve une représentation de chaque état en 3 dimensions.

En pratique, l'algorithme apprend la dépendance temporelle de la séquence qui est représentée sous forme de facteurs latents appelés *internal state* noté S. Une des particularités des RNN est qu'ils

reçoivent de l'information provenant de deux sources : un choc externe au système X_t ainsi que leur propre valeur à la période précédente S_{t-1}. Puisque l'algorithme est en mesure d'aller chercher de l'information aux périodes précédentes, on dit qu'ils ont une mémoire interne. Cet état interne est modélisé à l'intérieur de ce qu'on appelle une *memory cell*.

$$S_t = f\left(\boldsymbol{W} S_{t-1} + \boldsymbol{U} X_t\right) \tag{1.19}$$

$$S_{t-1} = f\left(\boldsymbol{W} S_{t-2} + \boldsymbol{U} X_{t-1}\right) \tag{1.20}$$

$$\vdots \tag{1.21}$$

$$S_{t-k} = f\left(\boldsymbol{W} S_{t-k-1} + \boldsymbol{U} X_{t-k}\right) \tag{1.22}$$

où f est la fonction d'activation, k est la longueur de la séquence, X_t est le vecteur de régresseurs à la période t et $\boldsymbol{W}$ et $\boldsymbol{U}$ sont les vecteurs de paramètres qui seront estimés par *Back Propagation Through Time* (BPTT). L'état S_t est directement fonction de l'état à la période précédente S_{t-1}. Cela est intéressant au niveau économique puisque la dynamique est apprise de façon autorégressive par le RNN.

CHAPITRE 2
EXERCICE EMPIRIQUE

2.1 Introduction

Nous avons décidé de prévoir 5 séries macroéconomiques d'intérêt particulier : l'inflation, le PIB, le taux de chômage, la construction de maisons neuves et le S&P 500. Nous faisons un exercice de prévision un trimestre ainsi qu'un an à l'avance pour chaque variable d'intérêt.

Les données. Pour notre exercice empirique, nous utilisons la base de données FRED-QD (McCracken et Ng, 2020)[1] fournies par la *Federal Reserve Economic Data* (FRED). Cette base de données est mise à jour régulièrement et contient des variables provenant de différents domaines économiques tels que l'activité réelle, la consommation, l'emploi et les marchés financiers. Elle comprend 246 séries macroéconomiques. Nous avons appliqué notre exercice empirique sur les données de 1960Q1 à 2022Q4, ce qui donne un total de 249 observations. Les données sont en niveaux et ont été transformées conformément aux recommandations de l'article original, de sorte qu'elles respectent les hypothèses de stationnarité. Hors, les séries qui sont catégorisés par l'article originale comme étant des prix, sont toutes transformées en prenant la première différence du log $\Delta \log(x_t) = \log(x_t) - \log(x_{t-1})$. Certaines séries contiennent des observations manquantes en début de période. Nous avons utilisé l'algorithme d'Expectation Maximization (EM) (Dempster *et al.*, 1977; Stock et Watson, 2002) pour reconstruire les données manquantes à partir des observations disponibles.

Modèles de référence. Le premier modèle de référence que nous utilisons est le modèle GARCH (Bollerslev, 1986) présenté au chapitre 1. Le modèle est défini par les équations (2.1) et (2.2), où ϵ_t est le terme d'erreur, σ_t^2 est la variance conditionnelle, β_k, α_i, et β_j sont des paramètres, et C est une constante.

[1] FRED-MD et FRED-QD sont deux bases de données macroéconomiques américaines mensuelles et trimestrielles conçues à des fins de recherche en données massives.

$$y_{t+1} = \beta_0 + \sum_{k=1}^{K} \beta_k y_{t-k} + \epsilon_{t+1}. \tag{2.1}$$

Avec l'hypothèse standard $\epsilon_t \sim \mathtt{N}(0, \sigma_t^2)$, nous modélisons la variance du processus σ_t^2 avec un modèle GARCH(p,q), où le paramètre p est le nombre de retards du processus σ^2 et q est le nombre de retards du processus du terme d'erreur ϵ^2, défini par l'équation :

$$\sigma_t^2 = C + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j} + U_t. \tag{2.2}$$

Dans l'équation (2.1), nous utilisons un modèle autorégressif d'ordre 2 pour modéliser la moyenne conditionnelle du processus. Dans l'équation (2.2), nous modélisons la variance conditionnelle à l'aide d'un processus GARCH(p, q) avec $p = 1$ et $q = 1$. Pour estimer simultanément la moyenne conditionnelle et la volatilité, nous maximisons la log-vraisemblance $\mathcal{L}$ définie par l'équation :

$$\ln \mathcal{L} = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \ln(\sigma_t^2) - \frac{1}{2} \sum_{t=1}^{T} \frac{\epsilon_t^2}{\sigma_t^2}, \tag{2.3}$$

Au lieu de le faire en deux étapes telles que présentées aux équations (2.1) et (2.2), le deuxième modèle de référence que nous utilisons est le modèle de volatilité stochastique SV, estimé de façon bayésienne par la méthode de MCMC. Supposons le processus stochastique à estimer :

$$y_t = \phi_0 + \phi_1 \mu_t + \sigma_t \eta_t \tag{2.4}$$

$$\log(\sigma_t^2) = \omega + \sum_{k=1}^{K} \beta_k \log(\sigma_{t-k}^2) + \nu_t \tag{2.5}$$

où y_t est l'observation de la série au temps t, μ_t et σ_t sont sa moyenne conditionnelle modélisé comme un processus AR(2) et sa variance conditionelle respectivement. Le terme η_t est une

variable aléatoire suivant une distribution normale. Les paramètres ω et β sont des paramètres du processus autorégressif de la volatilité à estimer. Finalement, ν_t est une variable aléatoire qui suit la distribution du processus de la volatilité. Nous utilisons la librairie *StochVol* du langage de programmation R pour effectuer l'estimation de ce modèle (Kastner et Frühwirth-Schnatter, 2014).

Ensuite, pour évaluer la capacité de notre modèle à prévoir la volatilité, nous le comparons à un modèle GARCH et à un modèle SV sur les résidus du modèle HNN, notés $\epsilon_{t,\text{HNN}}$. Nous dénommons ces modèles HNN-GARCH et HNN-SV, respectivement. De cette façon, nous pouvons effectuer une comparaison des trois modèles de volatilité lorsque le processus de moyenne conditionnelle est le même.

Finalement, nous utilisons un réseau de neurones simple, dénoté DNN, sans couche hémisphérique. Encore une fois, nous appliquons GARCH et SV sur les résidus obtenus, dénotés $\epsilon_{t,\text{DNN}}$. Le tableau présente un résumé des modèles qui seront évalués.

Hyperparamètres. Les hyperparamètres sont des paramètres qui contrôlent le comportement du processus d'entraînement de l'algorithme. Ils affectent notamment la vitesse de convergence de l'algorithme, la précision du modèle, la régularisation, la complexité du modèle et bien d'autres aspects du processus d'apprentissage. Le choix des hyperparamètres doit être fait en amont et doit être effectué avec beaucoup d'attention pour obtenir un modèle efficace et précis. En général, les hyperparamètres sont choisis en utilisant des techniques d'optimisation telles que la recherche par grille, la recherche aléatoire ou l'optimisation bayésienne. Cela implique de tester différentes combinaisons d'hyperparamètres pour trouver celle qui donne les meilleurs résultats en termes de performance ou d'autres critères prédéfinis. Dans le cas des réseaux de neurones, leurs hyperparamètres peuvent être le nombre de neurones, la profondeur du réseau, la fonction d'activation utilisée, le taux d'apprentissage pour la descente de gradient, et bien d'autres encore.

Pour notre modèle, pour rester en ligne avec les réseaux sans hémisphère, nous sélectionnons les

Tableau 1: Modèles de références

Acronyme	Description
HNN	Le modèle proposé en utilisant un réseau de neurones dense.
HNN-Garch	Le modèle proposé avec un réseau de neurones dense et une volatilité GARCH.
HNN-SV	Le modèle proposé avec un réseau de neurones dense et une volatilité SV.
DNN-Garch	Réseau dense sans hémisphère avec volatilité GARCH.
DNN-SV	Réseau dense sans hémisphère avec volatilité SV.
HRNN	Le modèle proposé avec un réseau récurrent
HRNN-Garch	Le modèle proposé avec un réseau récurrent et une volatilité GARCH
HRNN-SV	Le modèle proposé avec un réseau récurrent et une volatilité SV
RNN-Garch	Réseau récurrent avec volatilité GARCH
RNN-SV	Réseau récurrent avec volatilité SV
Garch	Modèle GARCH
SV	Modèle SV
AR	Modèle Autorégressif avec volatilité constante.

Notes: Les modèles de réseau de neurones sont estimés premièrement pour un réseau simple et ensuite pour un réseau récurrent.

hyperparamètres en reproduisant autant que possible ceux du papier (Goulet Coulombe, 2022). Les hyperparamètres sélectionnés sont présentés au Tableau 2.

Il est bon de mentionner que 1000 estimateurs est un grand nombre, du moins, pour la prévision hors-échantillon. Puisque que la recalibration dépend de l'erreur de prévision OOB tel que discuté aux ingrédients 3 et 4 de la Section 1, qui est la moyenne de $(1-0.8)\cdot 1000$ prévisions hors échantillon, il est important que celle-ci soit bien modélisée.

2.1.1 Métrique d'évaluation

Nous évaluons la performance de notre modèle sur les données hors échantillon (*out-of-sample*, OOS) en utilisant plusieurs métriques d'évaluation. Tout d'abord, pour évaluer la qualité de la prédiction

Tableau 2: Hyperparamètres

Hyperparamètres	Valeur
Nombre de couche cachée	2 couches par hémisphères. Profondeur total de 4.
Nombre de neurones	400 par couche cachée
Fonction d'activation	$f(x) = \max\{0, x\}$
Initialization des poids	$\theta \sim \mathcal{N}(0, 0.03)$
Nombre d'*epochs*[1] maximale	100
Arrêt précoce	15
le taux de *dropout*	0.2
Taux d'échantillonnage (τ)	0.8
Taille des blocs	8
Nombre de *bootstrap*	1000

[1] Un *epoch* est une étape d'entraînement où un ensemble complet de données d'entraînement est présenté à l'algorithme d'apprentissage profond.

de la moyenne conditionnelle, nous utilisons le Root Mean Squared Error (RMSE) qui mesure la distance euclidienne entre la valeur réelle de la série et la valeur prédite par le modèle :

$$RMSE = \sqrt{\frac{1}{N_{OOS}} \sum_{t \in OOS} (y_t - \hat{y}_t)^2} \tag{2.6}$$

Ici, N_{OOS} représente le nombre d'observations hors échantillon. Ensuite, puisque notre modèle maximise la vraisemblance de la loi gaussienne, une façon d'évaluer conjointement les processus estimés est de calculer leur fonction de densité de probabilité (PDF). Cette métrique est communément appelée le logarithmic score Good (1952).

$$f(\epsilon_t) = \frac{1}{N_{OOS}} \sum_{t \in OOS} \left(-\frac{1}{2} \left(\frac{\hat{\epsilon}_t}{\hat{\sigma}_t} \right)^2 - \frac{1}{2} \ln(2\pi\hat{\sigma}_t^2) \right) \tag{2.7}$$

La *logarithmic scoring rule* est une métrique avantageuse car elle est facile à interpréter et à

calculer. Elle mesure la log-vraisemblance des probabilités prédites et pénalise les prédictions en fonction de leur écart par rapport aux résultats réels.

En plus de cela, pour une évaluation complète, il est important de vérifier la calibration du modèle en calculant la fréquence à laquelle l'erreur de prévision est supérieure à la volatilité estimée. La volatilité estimée est représentée par l'écart-type d'une loi normale. Par conséquent, 68% des erreurs doivent être couvertes par la volatilité estimée, tandis que 95% des erreurs doivent être à l'intérieur de deux fois celle-ci. Pour faire cela, on calcule la fraction des erreur sortant de l'interval de confiance $\hat{y}_t \pm 1.96 \times \hat{\sigma}_t$.

2.2 Résultats

Nous réestimons le modèle tous les deux ans à partir du premier trimestre 2007. La période d'entraînement sera entre le premier trimestre 1961 et le quatrième trimestre 2006, et la période pseudo hors échantillon, qui sera utilisée pour tester le modèle, sera du premier trimestre 2007 au quatrième trimestre 2022. Les résultats de notre étude démontrent que pour certaines variables, le HNN surpasse les approches plus classiques en économétrie, tant en termes de précision que de calibration. De plus, grâce à sa nature proactive, le HNN est capable de signaler l'incertitude plus rapidement que le modèle SV.

Les résultats relatifs à la prévision d'un trimestre à l'avance sont exposés dans l'Annexe A, dans le Tableau 3 pour le modèle HNN, tandis que les résultats pour le modèle HRNN sont présentés dans le Tableau 6. Tout d'abord, nous présentons les résultats du PIB et du taux de chômage. Nous avons d'abord évalué ces deux variables sur la période allant de 2007Q1 à 2019Q4, puis sur la période allant de 2007Q1 à 2022Q4, en excluant l'année 2020. La motivation derrière ce choix est que nous voulions retirer l'effet de la pandémie sur ces deux variables d'activité réelles. Cette approche nous a permis d'éviter de comparer des modèles sur une période anormale.

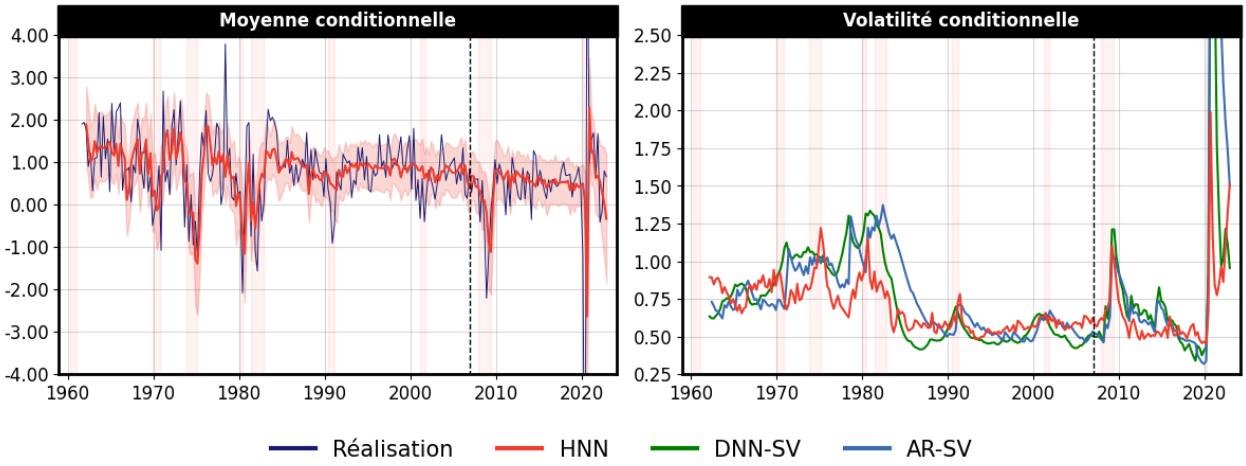

Figure 4: HNN - Produit Intérieur Brut (h=1)

Autant pour le PIB que pour le taux de chômage, le HNN surpasse de manière significative les modèles de référence en termes de RMSE et de Log-Score. On obtient une amélioration de la précision de 17% pour le PIB et de 27% pour le taux de chômage en comparaison avec le modèle AR. Les Figures 4 et 15 mettent en évidence la bonne réactivité de notre modèle face aux impulsions de la crise financière ainsi qu'à celles de la pandémie. Il est à noter que ces deux séries sont caractérisées par une persistance particulièrement marquée.

En termes de qualité générale, le HNN démontre une performance supérieure en termes de Log-Score. Aussi, on peut remarquer que tous les modèles de référence ainsi que le HNN ont des difficultés en termes de couverture nominale à 68% dans le cas du PIB. En effet, elle est trop élevée pour les deux périodes d'évaluation. Cela suggère que l'intervalle de confiance est trop étroit. Puisque le modèle a été entraîné sur la période incluant les années 70-80 et que le PIB a été sujet à un changement structurel durant ces années, il se pourrait que le modèle utilise des relations statistiques qui ne sont plus pertinentes dans le contexte économique de la période évaluée. Cela pourrait être la cause d'une mauvaise couverture nominale. En ce qui concerne le taux de chômage, le HNN obtient une meilleur performance pour la couverture nominale à 68% comparativement aux modèles de référence.

Nous présentons maintenant les résultats pour le taux d'inflation, la mise en chantier de résidences

Tableau 3: HNN - Sommaire des résultats statistiques des prévisions ($h = 1$)

	2007Q1 - 2019Q4						2007Q1 - 2022Q4					
	HNN	DNN_{SV}	DNN_{G}	AR_{SV}	AR_{G}	AR	HNN	DNN_{SV}	DNN_{G}	AR_{SV}	AR_{G}	AR
PIB												
RMSE	**0.83**	0.93	0.93	1.02	1.02	1.00	**0.87**	0.98	0.98	1.03	1.04	1.00
Log Score	**-3.93**	-3.83	-3.78	-3.77	-3.73	-3.66	**-3.87**	-3.70	-3.62	-3.69	-3.60	-3.63
Coverage(68%)	0.83	0.83	**0.79**	0.86	0.83	0.87	**0.82**	**0.82**	**0.82**	0.86	0.85	0.87
Coverage(95%)	0.98	0.96	0.96	0.96	**0.94**	0.98	0.98	0.97	0.97	0.97	**0.95**	0.98
Taux de chômage												
RMSE	**0.73**	0.78	0.78	1.03	0.99	1.00	**0.61**	0.71	0.71	0.76	0.86	1.00
Log Score	**-0.37**	-0.31	-0.33	-0.15	-0.18	-0.05	**-0.24**	-0.10	-0.17	-0.03	-0.02	0.10
Coverage(68%)	**0.69**	0.79	0.75	0.80	0.73	0.77	**0.70**	0.80	0.75	0.83	0.73	0.75
Coverage(95%)	0.96	0.98	0.96	0.96	**0.94**	0.96	0.97	0.97	0.97	0.97	**0.95**	0.97
Taux d'inflation												
RMSE	**0.93**	0.94	0.94	1.09	1.09	1.00	1.12	1.15	1.15	0.99	**0.99**	1.00
Log Score	-3.63	-3.74	**-3.79**	-3.33	-3.19	-3.55	-3.41	-2.67	-3.36	-3.30	-3.15	**-3.45**
Coverage(68%)	**0.67**	0.75	0.71	0.82	0.83	0.87	0.62	0.75	**0.72**	0.79	0.80	0.78
Coverage(95%)	**0.94**	**0.94**	0.92	0.98	0.98	0.98	0.89	0.91	0.91	0.97	**0.95**	0.94
Mise en chantier de résidences privées												
RMSE	**0.99**	1.00	1.00	1.01	1.02	1.00	**0.86**	0.87	0.87	0.99	0.99	1.00
Log Score	**-1.14**	-1.08	-1.08	-1.13	-1.13	-1.12	**-1.07**	-0.97	-0.94	-0.89	-0.90	-0.89
Coverage(68%)	0.63	0.63	0.65	0.70	**0.69**	0.75	0.66	0.66	**0.67**	0.71	0.70	0.75
Coverage(95%)	0.88	0.96	0.96	**0.96**	0.92	0.96	0.89	**0.95**	**0.95**	0.94	0.91	0.94
S&P 500												
RMSE	**0.96**	1.09	1.09	1.02	0.99	1.00	**0.93**	1.05	1.05	1.01	0.99	1.00
Log Score	**-1.55**	-1.22	-1.23	-1.27	-1.41	-1.20	**-1.52**	-1.28	-1.29	-1.27	-1.39	-1.23
Coverage(68%)	**0.73**	0.79	0.77	0.76	0.83	0.77	**0.70**	0.77	0.75	0.74	0.80	0.77
Coverage(95%)	0.98	0.92	0.92	0.90	**0.96**	**0.96**	0.98	**0.94**	**0.94**	0.90	0.97	0.97

Notes: Nous mettons en gras la métrique du modèle le plus performant par période d'évaluation. Les **RMSE** sont divisés par celui du **AR**, d'où l'obtention de 1.0 dans son cas. Les variables d'activité réelles, telles que le PIB et le taux de chômage, sont évaluées en excluant l'année 2020. Le RMSE du S&P 500 est divisé par celui obtenu en utilisant sa moyenne historique sur 40 ans.

privées et le S&P 500. Ces variables seront évaluées, premièrement, sur la période allant de 2007Q1 à 2019Q4 et, ensuite, de 2007Q1 à 2022Q4. Les résultats pour le HNN sont présentés dans le tableau 3. En ce qui concerne l'inflation, nos résultats démontrent une performance légèrement meilleure du HNN pour la période d'évaluation pré-covid. Le modèle produit un RMSE 7% meilleur que le AR durant cette période, mais 12% pire pour la période incluant la covid. Cela est dû au fait que le HNN a prédit une chute de l'inflation largement plus élevée que la réalisation, suivie d'un rebond

important. Naturellement, la calibration s'est retrouvée affectée par ce comportement durant cette même période. On remarque que les couvertures nominales à 68% et 95% sont légèrement en deçà de la cible alors que les modèles de référence sont beaucoup plus proches de celle-ci.

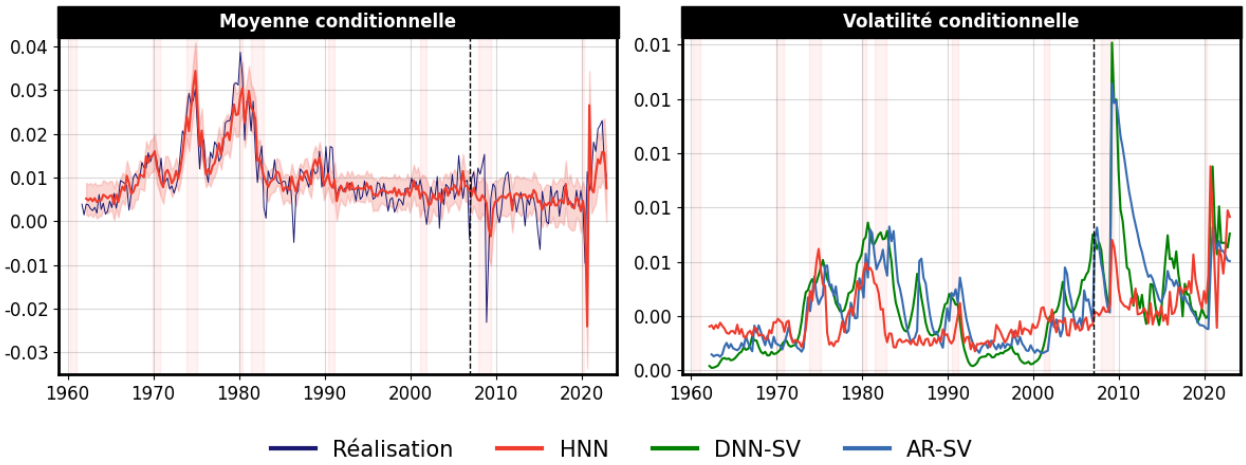

Figure 5: HNN - Inflation (h=1)

Du côté de la mise en chantier de résidences privées, le HNN semble fournir une prévision légèrement supérieure en termes de RMSE pour la période pré-pandémique (1%) et une amélioration significative pour la période incluant la pandémie (12%). Le HNN produit une calibration légèrement en deçà de l'objectif autant pour la couverture nominale à 68% que celle à 95%. Par ailleurs, il est difficile d'expliquer la prévision à la hausse soudaine durant la pandémie telle qu'on peut la voir à la Figure 6. En fait, on pourrait même penser que c'est contre-intuitif étant donné que l'activité réelle a été lourdement affectée par la pandémie. Historiquement, les mises en chantier de logements privés semblent aller à la baisse durant les périodes de récession indiquées par le NBER[1] et à la hausse lors des reprises économiques.

La cinquième et dernière variable d'intérêt est l'indice boursier S&P 500. Un élément qui distingue cette variable des quatre autres est sa faible persistance. La valeur de l'hyperparamètre $\boldsymbol{\lambda}$ déterminée par la méthode proposée est aussi élevée que 0.98. Cela suggère une préférence claire pour mettre l'accent sur la modélisation de la volatilité. Dans la littérature, il est courant d'utiliser la marche aléatoire comme modèle de référence pour les variables persistante. Dans notre cas, le

[1] Le Bureau National de Recherche Économique (NBER) est une organisation dédiée aux sciences économiques et à la recherche empirique associée, notamment en ce qui concerne l'économie américaine.

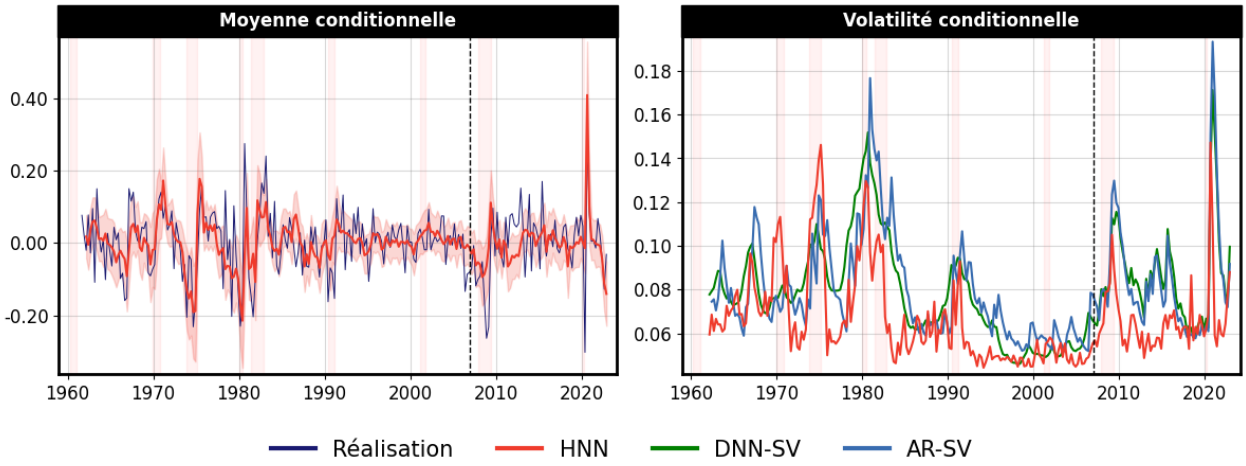

Figure 6: HNN - Mise en chantier de résidences privées (h=1)

modèle AR présenté dans les tableaux sous la variable S&P 500 est obtenu en utilisant la moyenne mobile sur 40 ans du rendement de l'indice. Nos résultats démontrent que le HNN est en mesure de surpasser légèrement celui du AR en termes de RMSE (4% et 7%). En termes de calibration, notre modèle est en mesure de délivrer un Log Score supérieur aux modèles de référence et une couverture nominale concurrentielle. Un phénomène intéressant est la différence de RMSE entre le HNN et le DNN. Cela pourrait être causé par l'impact de notre régularisation mettant l'accent sur la volatilité - qui est pour le S&P 500 très élevée - au lieu de la moyenne conditionnelle.

2.2.1 Prévision un an à l'avance (h=4)

Nous présentons maintenant les résultats pour la prévision 1 an à l'avance (h=4). Les résultats du HNN sont exposés dans l'Annexe A, dans le Tableau 5. L'aspect intéressant de la prévision 4 trimestres à l'avance est qu'on peut bien voir la capacité de notre modèle à signaler l'incertitude plus rapidement que le modèle SV. Ce phénomène est particulièrement observable pour l'inflation telle que présentée à la Figure 7. Durant la crise financière de 2008, le signal de volatilité produit par le HNN arrive bien avant celui des modèles de référence. Le modèle est aussi en mesure d'obtenir un RMSE significativement plus petit sur les deux périodes d'évaluation (12% et 7%). En termes de calibration, le HNN obtient un Log Score légèrement meilleur et une couverture nominale qui semble couvrir excessivement.

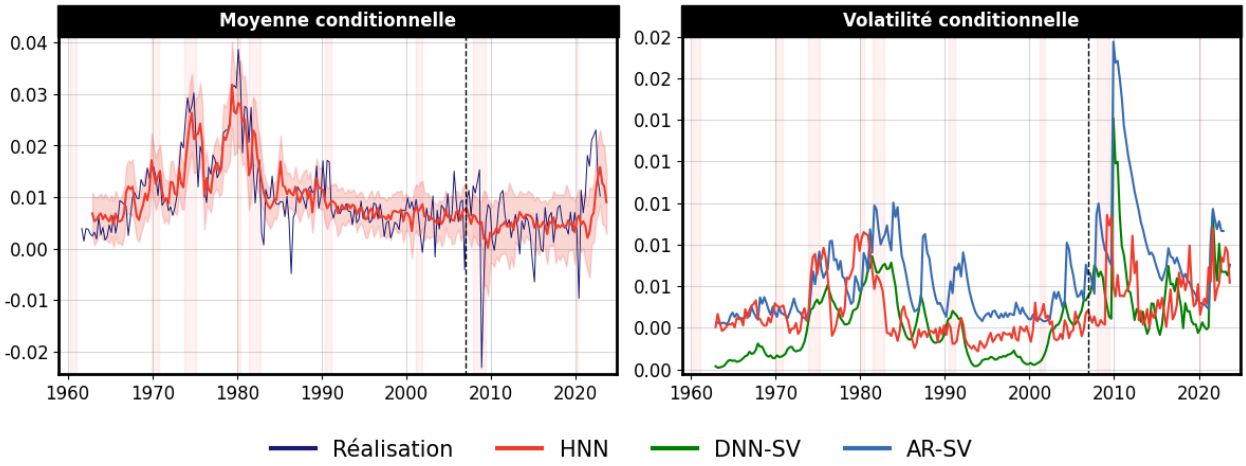

Figure 7: HNN - Inflation (h=4)

En ce qui concerne le S&P 500, il est difficile de prévoir les rendements de l'indice un an à l'avance. Si on regarde la Figure 8, le HNN semble en accord avec la difficulté à utiliser des signaux utiles à sa prévision. C'est pour cette raison que la prévision semble proche de sa moyenne non conditionnelle. De plus, notre modèle semble capturer la volatilité durant la crise de 2008 légèrement plus rapidement que les modèles de référence. Néanmoins, le HNN produit un RMSE légèrement meilleur que les modèles de référence sur les deux périodes d'évaluation (1% et 2%) ainsi qu'une couverture nominale légèrement au-dessus de la cible.

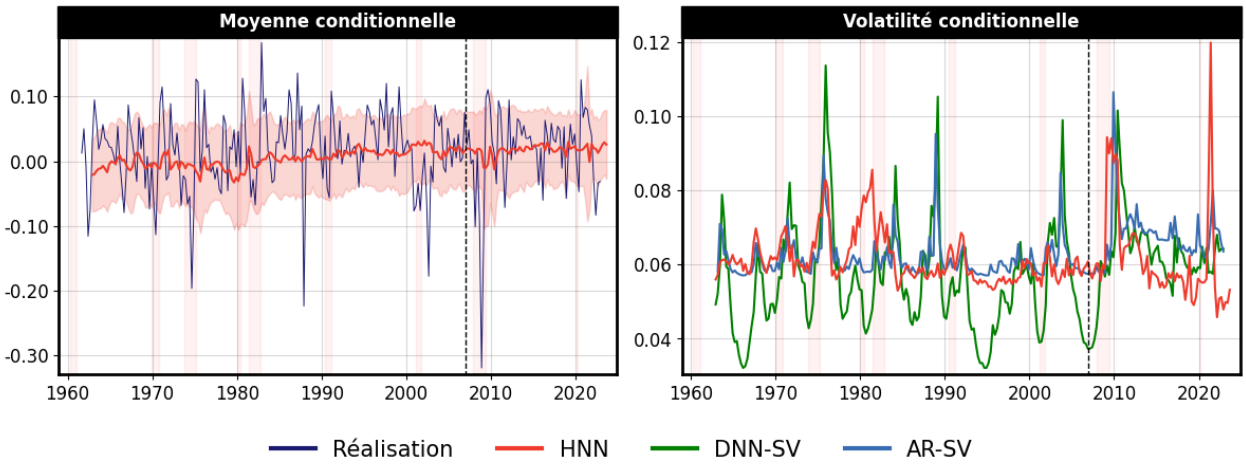

Figure 8: HNN - S&P 500 (h=4)

2.2.2 Est-ce que le HRNN peut améliorer les prévisions ?

Tel que discuté précédemment, les réseaux récurrents conviennent bien aux tâches de prévision grâce à leur architecture autorégressive interne. Cependant, il est bon de mentionner qu'ils sont plus facilement sujets au surapprentissage que les réseaux simples. Cela mène inévitablement à une confiance excessive dans les prévisions, une estimation de l'incertitude plus petite et une sous-couverture nominale. Ce phénomène est visible pour toutes les variables qui ont été analysées. Dans le cas du PIB, cet effet a été bénéfique. Tandis que le HNN et les modèles de référence ont une couverture nominale 68% plus élevée (au-delà de 80%), le HRNN produit une couverture qui est beaucoup plus proche de la cible (70% et 73%). En revanche, la calibration de la mise en chantier de résidences privées ainsi que celle du taux d'inflation sont totalement en dessous de l'objectif. Les résultats pour la prévision d'un trimestre à l'avance sont présentés dans le Tableau 6, tandis que les prévisions 1 an à l'avance sont présentées dans le Tableau 4.

Dans le cas de la prévision 1 an à l'avance, nos résultats montrent que le HRNN peut produire de meilleures prévisions tout en bénéficiant significativement d'une meilleure calibration. De plus, on peut remarquer que pour le PIB et le taux de chômage, présentés aux Figures 9 et 10, notre modèle signale la volatilité plus rapidement que les modèles de référence. Il est bon de mentionner que le HRNN a bien capturé la baisse du PIB durant la crise financière de 2008, ce qui n'est pas le cas pour le HNN, comme on peut le voir à la Figure 17.

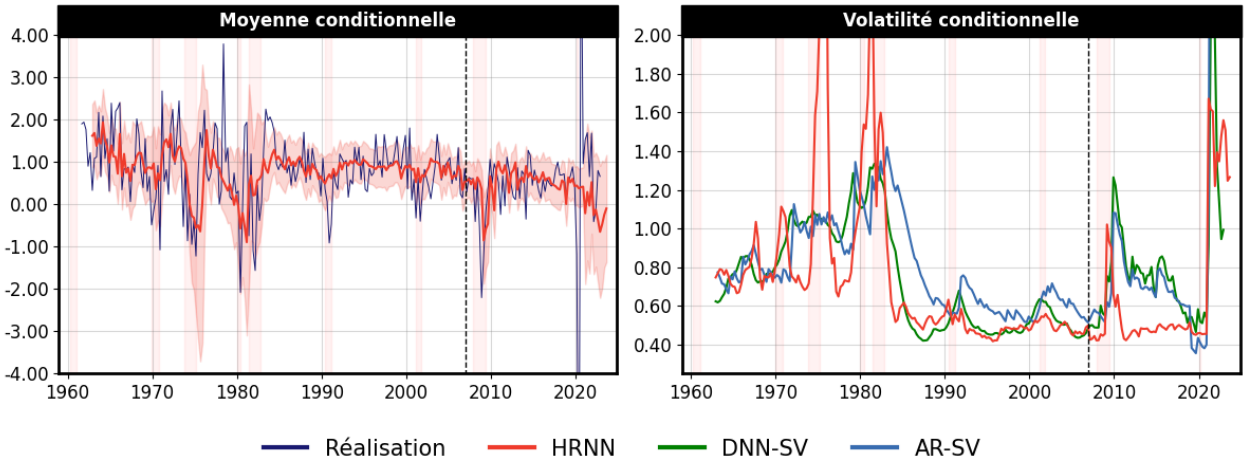

Figure 9: HRNN - Produit Intérieur Brut (h=4)

Un élément d'analyse intéressant est que le HRNN semble prévoir une augmentation du taux

Tableau 4: HRNN - Sommaire des résultats statistiques des prévisions ($h = 4$)

	2007Q1 - 2019Q4						2007Q1 - 2022Q4, Excluant 2020					
	HRNN	RNN_{SV}	RNN_{G}	AR_{SV}	AR_{G}	AR	HRNN	RNN_{SV}	RNN_{G}	AR_{SV}	AR_{G}	AR
PIB												
RMSE	**0.89**	0.91	0.91	1.03	1.04	1.00	**0.99**	1.11	1.11	1.04	1.04	1.00
Log Score	**-3.58**	-3.56	-3.49	-3.41	-3.38	-3.54	**-3.51**	-3.41	-3.42	-3.37	-3.28	-3.51
Coverage(68%)	**0.73**	0.85	0.83	0.83	0.79	0.88	**0.75**	0.82	0.82	0.84	0.82	0.87
Coverage(95%)	0.90	0.98	**0.96**	0.94	0.92	**0.96**	0.92	0.93	0.97	**0.95**	0.93	0.97
Taux de chômage												
RMSE	0.72	**0.71**	**0.71**	1.07	1.01	1.00	**0.80**	1.09	1.09	1.09	1.21	1.00
Log Score	**-0.13**	0.01	0.03	0.59	0.47	0.34	**0.02**	0.16	0.20	0.70	0.57	0.41
Coverage(68%)	**0.71**	0.83	0.83	0.83	0.83	0.79	**0.75**	0.82	0.83	0.84	0.83	0.80
Coverage(95%)	0.90	**0.94**	**0.94**	0.91	0.92	**0.94**	0.92	0.92	0.93	0.93	0.92	**0.95**
Taux d'inflation												
RMSE	**0.91**	1.05	1.05	1.13	1.03	1.00	**0.97**	1.04	1.04	1.08	1.00	1.00
Log Score	-3.25	-3.37	-3.39	-3.39	**-3.51**	-3.49	-2.96	-3.16	-3.16	-3.26	**-3.42**	-3.39
Coverage(68%)	**0.65**	0.73	0.75	0.74	0.90	0.83	0.59	0.66	0.67	**0.68**	0.83	0.75
Coverage(95%)	0.88	0.90	0.90	0.91	**0.96**	0.98	0.84	0.86	0.88	0.88	**0.94**	0.92
La mise en chantier de résidences privées												
RMSE	1.02	1.05	1.05	**0.95**	0.97	1.00	1.02	1.03	1.03	0.98	**0.98**	1.00
Log Score	-0.80	-0.93	-0.99	**-1.06**	-1.05	-1.02	-0.52	-0.63	-0.68	-0.81	-0.80	**-0.90**
Coverage(68%)	0.54	0.71	0.71	0.79	**0.67**	0.79	0.58	0.72	0.70	0.76	**0.67**	0.78
Coverage(95%)	0.85	**0.94**	**0.94**	0.96	0.96	0.96	0.84	0.92	0.92	0.93	**0.94**	**0.94**
S&P 500												
RMSE	1.19	1.18	1.18	1.00	**0.99**	1.00	1.18	1.21	1.21	0.99	**0.98**	1.00
Log Score	-1.05	-0.86	-0.91	-1.15	-1.16	**-1.20**	-1.05	-0.84	-0.90	-1.19	-1.19	**-1.22**
Coverage(68%)	**0.77**	0.81	0.79	0.79	0.79	**0.77**	**0.70**	0.73	0.70	0.78	0.78	0.75
Coverage(95%)	0.92	0.90	0.90	**0.96**	0.94	0.98	0.91	0.89	0.91	0.97	**0.95**	0.98

Notes: Nous mettons en gras la métrique du modèle le plus performant par période d'évaluation. Les **RMSE** sont divisés par celui du **AR**, d'où l'obtention de 1.0 dans son cas. Les variables d'activité réelles, telles que le PIB et le taux de chômage, sont évaluées en excluant l'année 2020. Le RMSE du S&P 500 est divisé par celui obtenu en utilisant sa moyenne historique sur 40 ans.

de chômage aux États-Unis pour l'année 2023, ainsi qu'un recul du PIB. La hausse des taux directeurs dans l'objectif de réduire l'inflation pourrait être le facteur principal d'une telle prévision du modèle. Néanmoins, le modèle semble exhiber beaucoup d'incertitude pour 2024. Il est donc important que les prévisions soient analysées avec nuance.

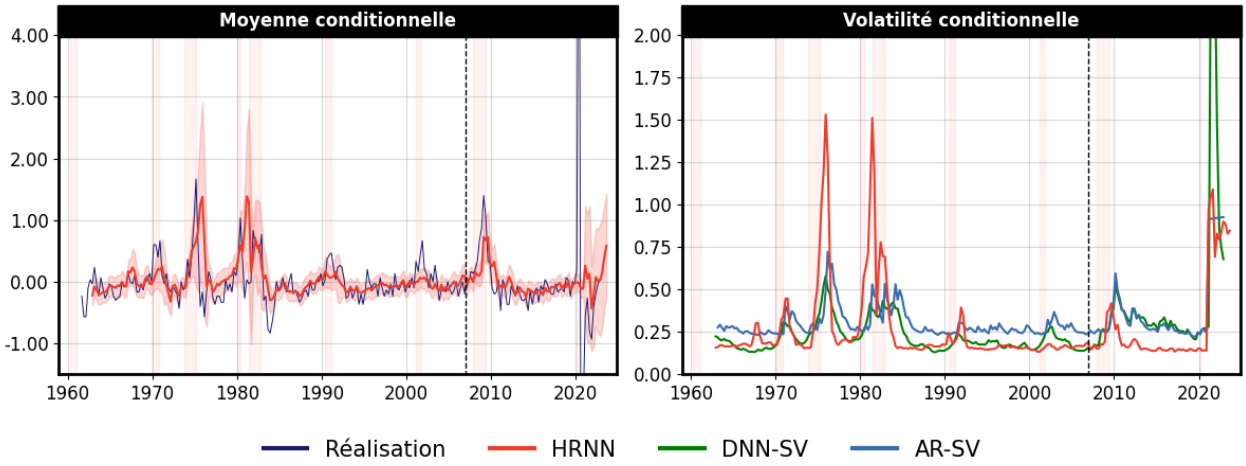

Figure 10: HRNN - Taux de chômage (h=4)

2.3 Importance des variables

En apprentissage automatique supervisé, mesurer la contribution des variables à l'exercice de prévision est une pratique qui est très importante pour tout scientifique soucieux de comprendre les phénomènes sous-jacents. Pour une régression linéaire, mesurer la significativité des coefficients est suffisant pour déterminer si une variable est importante (en supposant qu'il n'y a pas de variables omises). Dans le modèle Lasso, les variables sont sélectionnées par un processus de réduction de la taille des coefficients. Dans les algorithmes à base d'arbre, l'importance des variables peut être mesurée par la réduction du "gini impurity" lors de la séparation d'un nœud. Malheureusement, pour un réseau de neurones, la tâche n'est pas aussi simple. Non seulement la quantité de coefficients impliqués dans l'exercice de prévision rend l'analyse de ceux-ci très difficile, mais la transformation non linéaire successive les rend intraçables. Pour cette raison, nous mesurons l'importance des variables IV en calculant la distance entre la prévision originale de l'hémisphère j, dénotée $mathttH_j(X_t)$, et celle obtenue lorsque les retards de la variable k_i sont perturbés, soit $mathttH_j(\tilde{X}_t)$ (Goulet Coulombe, 2022).

$$IV_k^j = \left(\frac{\frac{1}{T}\sum_{t=1}^{T} \left(\mathtt{H}_j(X_t) - \mathtt{H}_j(\tilde{X}_t) \right)^2}{Var(\mathtt{H}_j(X_t))} \right) \times 100 \tag{2.8}$$

En répétant le processus 5 fois par variable et en prenant la moyenne comme résultat final, on s'assure de fiabilité de ceux-ci. Nous effectuons l'exercice pour la période allant de 2007Q1 à

2019Q4. La raison pour laquelle nous laissons de côté la période de pandémie est que certains régresseurs tels que certains agrégats monétaires ou certaines industries telles que les loisirs présentent une très grande variance durant cette période. Par conséquent, ils sont majoritairement les conducteurs des prévisions durant la pandémie pour chacune de nos variables.

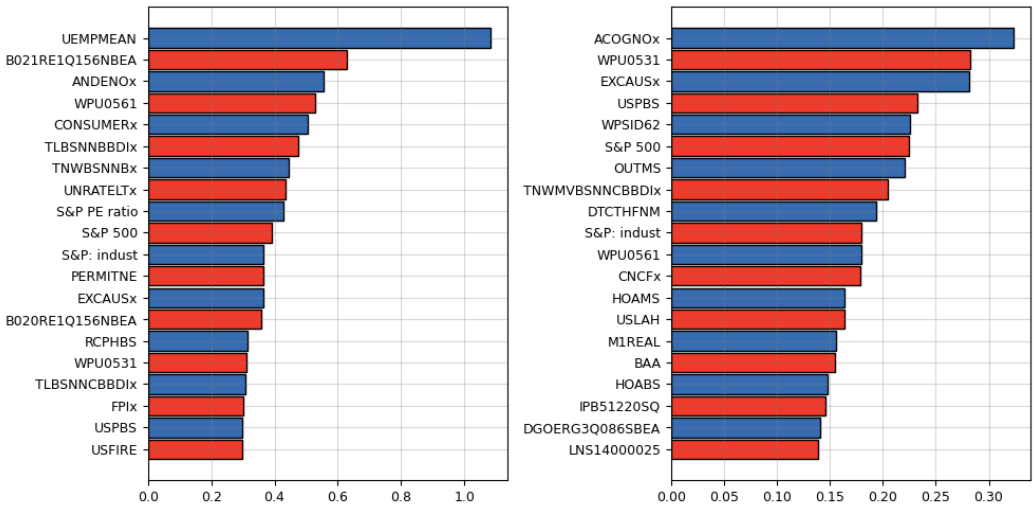

Figure 11: Importance des variables pour le PIB. Dans le panneau de gauche, on présente l'importance des variables pour l'hémisphère de la moyenne conditionnelle, et dans le panneau de droite, celle de la variance conditionnelle.

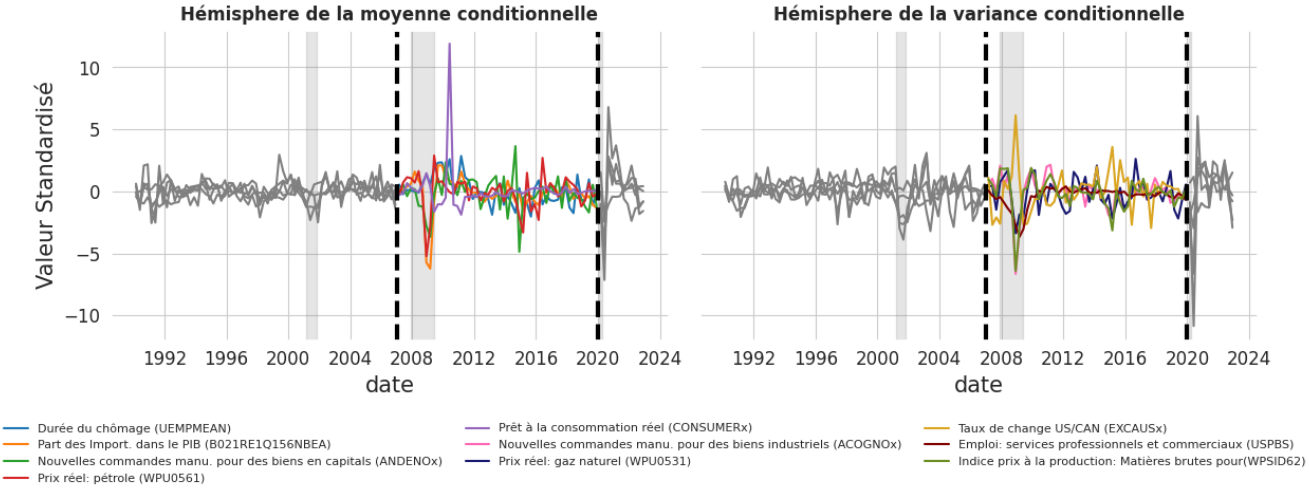

Figure 12: Graphique des variables importantes par hémisphère pour le PIB. Dans le panneau de gauche, on trouve les variables qui ont le plus d'influence sur le taux d'inflation durant la période allant de 2007Q1 à 2019Q4 pour l'hémisphère de la moyenne conditionnelle. À droite, celles de l'hémisphère de la variance conditionnelle.

Du côté du PIB, nos résultats sont présentés à la Figure 11 et complétés à la Figure 12. Ceux-ci démontrent que l'emploi, le taux de change, les nouvelles commandes manufacturières et le prix

des matières premières sont les principaux facteurs qui influencent la variation du PIB. Ces résultats sont en ligne avec la théorie économique puisque le marché du travail joue un rôle crucial en affectant la production économique globale. Plus précisément, la création d'emplois peut stimuler la consommation et, par conséquent, augmenter la demande agrégée. Le taux de change, quant à lui, influence la compétitivité des entreprises et le coût des importations, ce qui peut impacter la croissance économique. Les nouvelles commandes manufacturières sont également un facteur important car elles sont un précurseur de l'activité économique future et de la demande pour les produits manufacturés. Enfin, le prix des matières premières est un élément déterminant de la production et de la croissance économique, car il peut affecter les coûts de production et les bénéfices des entreprises. On peut également observer une hausse des prêts à la consommation réelle aux États-Unis au milieu de 2010. Puisque la consommation est un déterminant important dans l'équation du PIB, la hausse considérable de sa variance durant cette période pourrait expliquer pourquoi elle est un déterminant de sa moyenne conditionnelle.

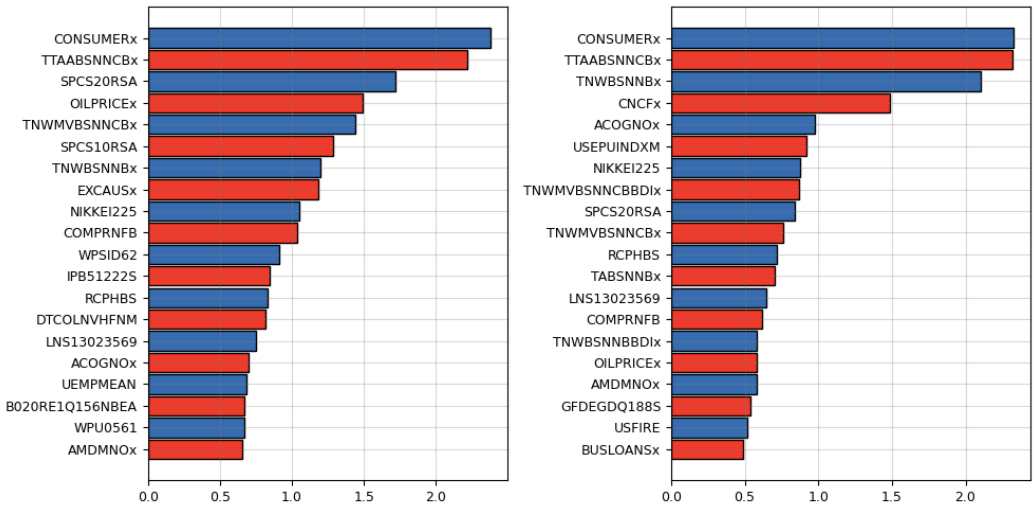

Figure 13: Importance des variables pour l'inflation. Dans le panneau de gauche, on présente l'importance des variables pour l'hémisphère de la moyenne conditionnelle, et dans le panneau de droite, celle de la variance conditionnelle.

En ce qui concerne le taux d'inflation, les résultats sont présentés à la Figure 13 et complétés à la Figure 14. Ceux-ci indiquent que la valeur des actifs des entreprises non financières ainsi que les prêts à la consommation réelle sont des facteurs déterminants qui influencent l'inflation dans les deux hémisphères. Intuitivement, une baisse des prix peut justifier une réduction des revenus et donc de l'encaisse des entreprises. Un autre facteur d'intérêt est l'indice des prix des maisons, qui

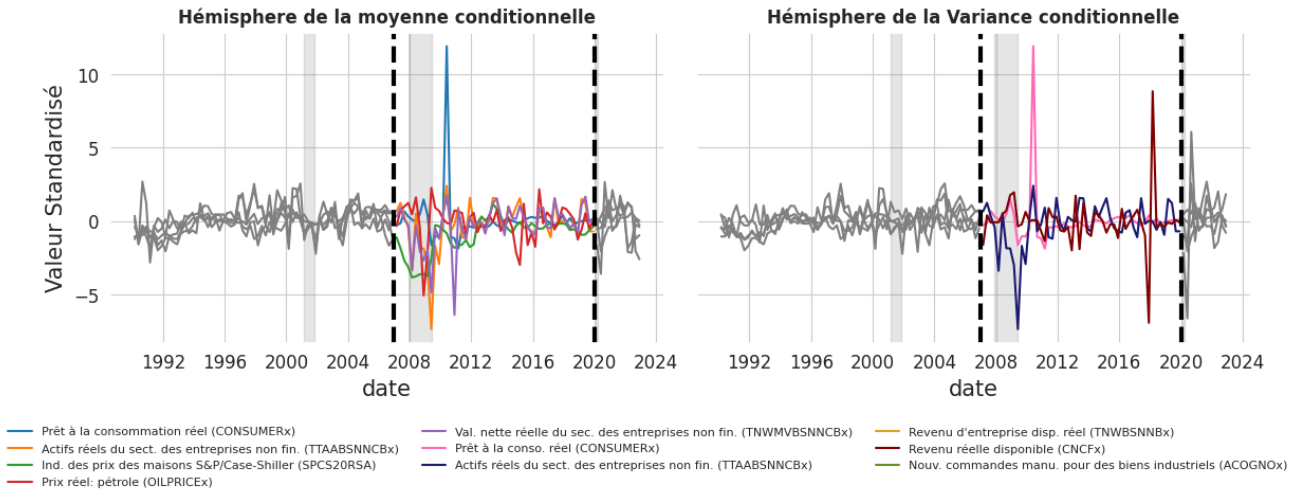

Figure 14: Graphique des variables importantes par hémisphère pour l'inflation. Dans le panneau de gauche, on trouve les variables qui ont le plus d'influence sur le taux de croissance du PIB durant la période allant de 2007Q1 à 2019Q4 pour l'hémisphère de la moyenne conditionnelle. À droite, celles de l'hémisphère de la variance conditionnelle.

est le principal moteur de la crise financière de 2008. Cet indice commence à diminuer dès 2007 et atteint son point le plus bas peu après que l'inflation ait atteint le sien. Sans surprise, le prix du pétrole est également un facteur influençant le taux d'inflation au cours de cette période. Le choc pétrolier des années 1970 a été le principal déclencheur de la période de stagflation, et il semble que notre algorithme ait été en mesure de réutiliser cette relation pour modéliser la moyenne conditionnelle pendant la période étudiée. Enfin, le revenu des entreprises semble également jouer un rôle dans la volatilité de l'inflation. En effet, la réduction des impôts mise en place par l'ancien président américain Donald Trump en décembre 2017 a eu un impact sur le revenu des entreprises. L'augmentation de la variance des revenus peut expliquer la hausse de la volatilité de l'inflation au cours de la période d'évaluation, comme on peut le voir sur la Figure 5.

En ce qui concerne la construction de logements privés, les résultats de l'algorithme suggèrent que la performance financière des entreprises ainsi que les prix des matières premières, telles que les métaux et le gaz naturel, sont des facteurs importants. En ce qui concerne le taux de chômage, bien que le changement en pourcentage du MSE soit relativement faible, l'emploi et le nombre d'heures travaillées dans certains secteurs sont les principaux déterminants. Enfin, les variables qui ont un impact sur l'indice S&P 500 durant cette période sont la valeur des actifs des ménages et des entreprises. Intuitivement, la hausse de la valeur réelle des actifs des ménages pourrait être

attribuée à la hausse de la valeur des actions de l'indice. Le lecteur pourrait se demander pourquoi aucun taux d'intérêt ne figure comme facteur important pour l'indice boursier le plus populaire aux États-Unis. Cela peut s'expliquer par le fait que les taux d'intérêt étaient particulièrement bas et stables entre 2007Q1 et 2019Q4. Par conséquent, leur variance peut ne pas avoir été suffisamment importante pour influencer les processus conditionnels.

2.4 Conclusion

Ce mémoire avait pour objectif de développer un modèle de réseau de neurones qui est en mesure d'effectuer une prévision ainsi que l'incertitude autour de cette prévision. Cela a été accompli en utilisant un *Hemisphere Neural Network*, tel que proposé par (Goulet Coulombe, 2022), contenant deux hémisphères. Nous avons montré qu'il est possible d'utiliser la sortie des hémisphères pour maximiser la log-vraisemblance de la distribution conditionnelle et ainsi pouvoir effectuer la prévision de densité d'une série macroéconomique. De plus, nous avons montré qu'en imposant une contrainte à la sortie de l'hémisphère de la volatilité, on peut spécifier l'emphase qui sera mise sur le processus de la variance conditionnelle. En procédant ainsi, l'algorithme est contraint de mettre l'accent sur le moment statistique qui convient le mieux à la distribution de la variable qu'on tente de prévoir. Nous avons effectué un exercice empirique sur des données macroéconomiques pour prévoir le PIB, le taux de chômage, le taux d'inflation, la mise en chantier de résidences privées et le S&P 500. Nos résultats démontrent que notre modèle est en mesure d'améliorer la précision des prévisions tout en fournissant une estimation de l'incertitude plus proactive que les modèles standards tels que GARCH et SV. Nous avons aussi effectué un exercice d'analyse d'importance des variables. L'architecture du modèle nous a permis d'effectuer cette analyse par hémisphères. Ainsi, nous avons été en mesure de définir les déterminants de la prévision ainsi que ceux de sa volatilité. Les résultats démontrent que l'algorithme est en mesure d'identifier des relations qui sont en ligne avec la théorie macroéconomique.

Même si le modèle semble prometteur, plusieurs autres avenues doivent être explorées pour témoigner de sa robustesse. En particulier, il pourrait être intéressant d'évaluer son application sur des données à une fréquence plus élevée. En effet, une analyse utilisant la base de données Fred-MD pourrait être un élément d'analyse très intéressant. On pourrait également l'utiliser pour effectuer des prévisions

dans des pays autres que les États-Unis. En effet, il serait intéressant de voir si la relation entre les variables est similaire à celle en Amérique du Nord. Un autre élément d'analyse qui pourrait être intéressant serait de maximiser la log-vraisemblance avec une distribution qui fournirait davantage d'informations. Par exemple, on pourrait maximiser la log-vraisemblance avec une loi normale asymétrique en ajoutant une troisième variable qui serait forcée d'estimer le paramètre d'asymétrie dans la distribution conditionnelle. Il est à noter que l'estimation du sens du risque est un sujet très populaire en macroéconomie. Une autre catégorie d'avenues possibles serait le cas multivarié. En ayant plus d'un prédicteur, le réseau est forcé de réorganiser l'information apprise. Cela mène à une forme de régularisation en empêchant le surapprentissage et à une représentation plus généralisable de l'information provenant des régresseurs.

CHAPITRE 3
ANNEXES

A Sommaire des résultats statistiques

Tableau 5: HNN - Sommaire des résultats statistiques des prévisions ($h = 4$)

	2007Q1 - 2019Q4						2007Q1 - 2022Q4, Excluant 2020					
	HNN	DNN_{SV}	DNN_{G}	AR_{SV}	AR_{G}	AR	HNN	DNN_{SV}	DNN_{G}	AR_{SV}	AR_{G}	AR
PIB												
RMSE	0.91	**0.89**	**0.89**	1.03	1.04	1.00	**0.94**	1.61	1.61	1.04	1.04	1.00
Log Score	**-3.70**	-3.59	-3.58	-3.41	-3.38	-3.54	**-3.61**	-3.36	-3.44	-3.37	-3.28	-3.51
Coverage(68%)	0.83	0.87	0.85	0.83	**0.79**	0.88	0.85	0.85	0.85	0.84	**0.82**	0.87
Coverage(95%)	**0.94**	0.98	0.98	0.94	0.92	0.96	**0.95**	0.97	0.97	**0.95**	0.93	0.97
Taux de chômage												
RMSE	0.77	**0.72**	**0.72**	1.07	1.01	1.00	**0.89**	2.80	2.80	1.09	1.21	1.00
Log Score	**-0.17**	0.02	0.05	0.59	0.47	0.34	**0.03**	0.59	0.83	0.70	0.57	0.41
Coverage(68%)	0.81	0.85	0.85	0.83	0.83	**0.79**	0.82	0.82	0.85	0.84	0.83	0.80
Coverage(95%)	**0.94**	**0.94**	**0.94**	0.91	0.92	**0.94**	**0.95**	0.90	0.93	0.93	0.92	**0.95**
Taux d'inflation												
RMSE	**0.88**	0.90	0.90	1.13	1.03	1.00	**0.93**	0.96	0.96	1.08	1.00	1.00
Log Score	-3.48	-3.61	**-3.64**	-3.39	-3.51	-3.49	-3.30	-3.38	-3.32	-3.26	**-3.42**	-3.39
Coverage(68%)	0.77	0.81	0.77	**0.74**	0.90	0.83	0.70	0.73	0.69	**0.68**	0.83	0.75
Coverage(95%)	0.96	**0.94**	0.92	0.91	0.96	0.98	0.91	0.89	0.88	0.88	**0.94**	0.92
Mise en chantier de résidences privées												
RMSE	1.02	1.07	1.07	**0.95**	0.97	1.00	1.01	1.12	1.12	0.98	**0.98**	1.00
Log Score	-0.88	-0.94	-0.99	**-1.06**	-1.05	-1.02	-0.67	-0.56	-0.68	-0.81	-0.80	**-0.90**
Coverage(68%)	0.60	0.71	**0.69**	0.79	0.67	0.79	0.61	**0.69**	**0.69**	0.76	0.67	0.78
Coverage(95%)	0.90	0.94	0.96	**0.96**	0.96	0.96	0.89	0.89	**0.94**	0.93	**0.94**	**0.94**
S&P 500												
RMSE	**0.99**	1.15	1.15	1.00	0.99	1.00	**0.98**	1.20	1.20	0.99	0.98	1.00
Log Score	**-1.28**	-1.03	-1.05	-1.15	-1.16	-1.20	**-1.28**	-0.99	-1.02	-1.19	-1.19	-1.22
Coverage(68%)	**0.77**	0.83	0.79	0.79	0.79	**0.77**	**0.73**	0.80	0.75	0.78	0.78	0.75
Coverage(95%)	0.98	0.94	0.94	**0.96**	0.94	0.98	0.97	0.94	0.92	0.97	**0.95**	0.98

Notes: Nous mettons en gras la métrique du modèle le plus performant par période d'évaluation. Les **RMSE** sont divisés par celui du **AR**, d'où l'obtention de 1.0 dans son cas. Les variables d'activité réelles, telles que le PIB et le taux de chômage, sont évaluées en excluant l'année 2020. Le RMSE du S&P 500 est divisé par celui obtenu en utilisant sa moyenne historique sur 40 ans.

Tableau 6: HRNN - Sommaire des résultats statistiques des prévisions ($h = 1$)

	2007Q1 - 2019Q4						2007Q1 - 2022Q4, Excluant 2020					
	HRNN	RNN_{SV}	RNN_{G}	AR_{SV}	AR_{G}	AR	HRNN	RNN_{SV}	RNN_{G}	AR_{SV}	AR_{G}	AR
PIB												
RMSE	**0.88**	0.95	0.95	1.02	1.02	1.00	**0.90**	1.00	1.00	1.03	1.04	1.00
Log Score	**-3.95**	-3.82	-3.82	-3.77	-3.73	-3.66	**-3.90**	-3.71	-3.68	-3.69	-3.60	-3.63
Coverage(68%)	**0.73**	0.79	0.81	0.86	0.83	0.87	**0.70**	0.80	0.82	0.86	0.85	0.87
Coverage(95%)	0.92	0.96	0.96	0.96	**0.94**	0.98	0.93	0.97	0.97	0.97	**0.95**	0.98
Taux de chômage												
RMSE	0.77	**0.72**	**0.72**	1.03	0.99	1.00	**0.63**	0.65	0.65	0.76	0.86	1.00
Log Score	**-0.43**	-0.34	-0.36	-0.15	-0.18	-0.05	**-0.29**	-0.16	-0.22	-0.03	-0.02	0.10
Coverage(68%)	**0.63**	0.81	0.79	0.80	0.73	0.77	**0.63**	0.80	0.78	0.83	0.73	0.75
Coverage(95%)	0.96	**0.94**	**0.94**	0.96	**0.94**	0.96	0.93	0.93	**0.95**	0.97	**0.95**	0.97
Taux d'inflation												
RMSE	**0.98**	1.07	1.07	1.09	1.09	1.00	1.15	1.16	1.16	0.99	**0.99**	1.00
Log Score	-3.49	-3.69	**-3.83**	-3.33	-3.19	-3.55	-3.30	-3.21	**-3.51**	-3.30	-3.15	-3.45
Coverage(68%)	0.54	**0.71**	**0.71**	0.82	0.83	0.87	0.48	**0.66**	**0.66**	0.79	0.80	0.78
Coverage(95%)	0.85	0.88	0.90	**0.98**	0.98	0.98	0.81	0.86	0.89	0.97	**0.95**	0.94
Mise en chantier de résidences privées												
RMSE	**0.97**	1.10	1.10	1.01	1.02	1.00	**0.85**	0.97	0.97	0.99	0.99	1.00
Log Score	-1.07	-0.89	-1.02	-1.13	**-1.13**	-1.12	**-1.00**	-0.81	-0.84	-0.89	-0.90	-0.89
Coverage(68%)	0.54	0.62	0.62	0.70	**0.69**	0.75	0.58	0.64	0.64	0.71	**0.70**	0.75
Coverage(95%)	0.83	0.96	0.96	0.96	0.92	0.96	0.84	**0.94**	**0.94**	0.94	0.91	**0.94**
S&P 500												
RMSE	1.25	1.31	1.31	1.02	**0.99**	1.00	1.22	1.24	1.24	1.01	**0.99**	1.00
Log Score	**-1.47**	-1.04	-1.21	-1.27	-1.41	-1.20	**-1.41**	-1.13	-1.27	-1.27	-1.39	-1.23
Coverage(68%)	**0.73**	0.75	0.75	0.76	0.83	0.77	**0.70**	0.77	0.73	0.74	0.80	0.77
Coverage(95%)	**0.96**	0.90	0.92	0.90	**0.96**	**0.96**	**0.95**	0.92	0.94	0.90	0.97	0.97

Notes: Nous mettons en gras la métrique du modèle le plus performant par période d'évaluation. Les **RMSE** sont divisés par celui du **AR**, d'où l'obtention de 1.0 dans son cas. Les variables d'activité réelles, telles que le PIB et le taux de chômage, sont évaluées en excluant l'année 2020. Le RMSE du S&P 500 est divisé par celui obtenu en utilisant sa moyenne historique sur 40 ans.

B Graphiques de prévision

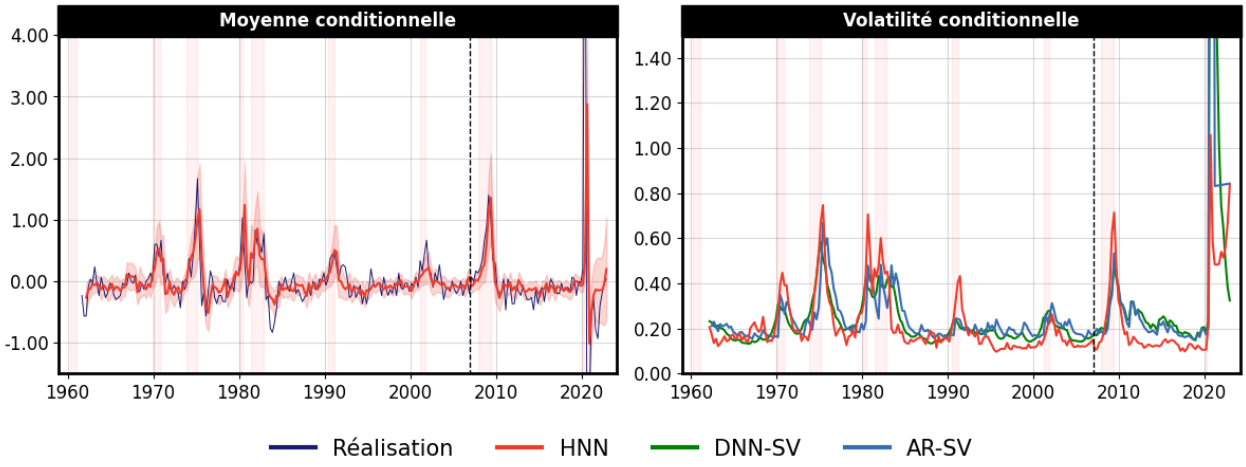

Figure 15: HNN - Taux de chômage (h=1)

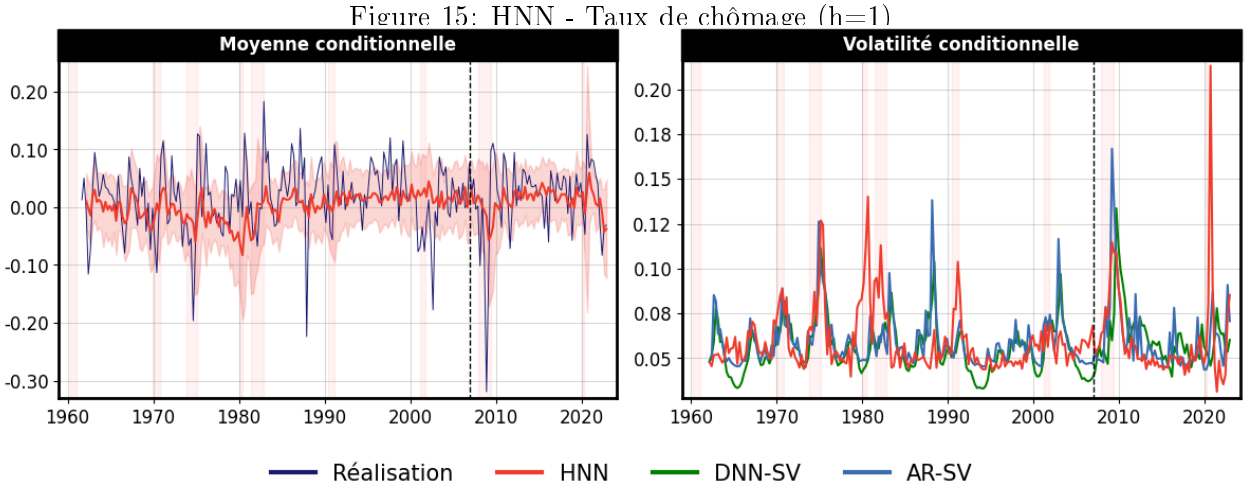

Figure 16: HNN - S&P 500 (h=1)

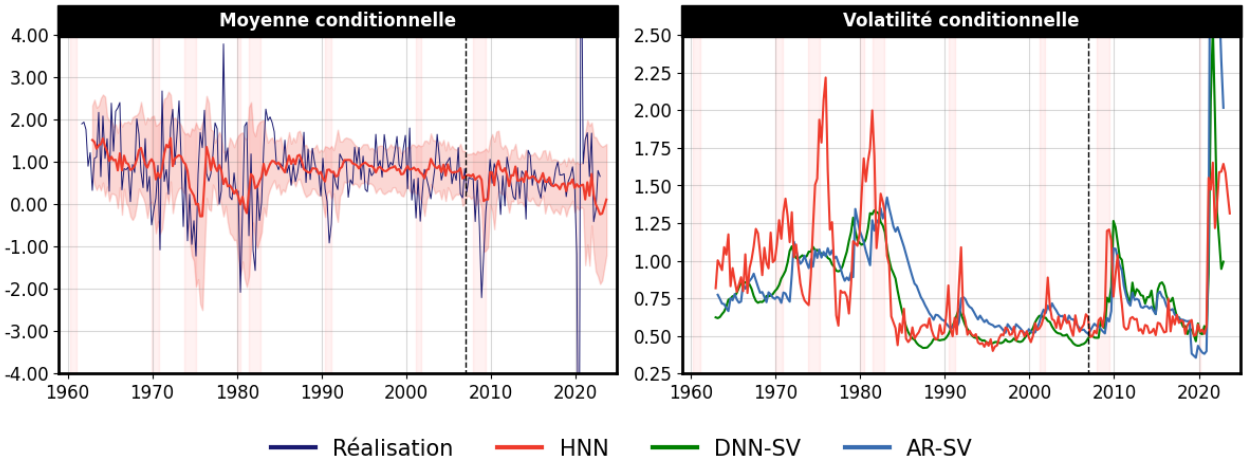

Figure 17: HNN - Produit Intérieur Brut (h=4)

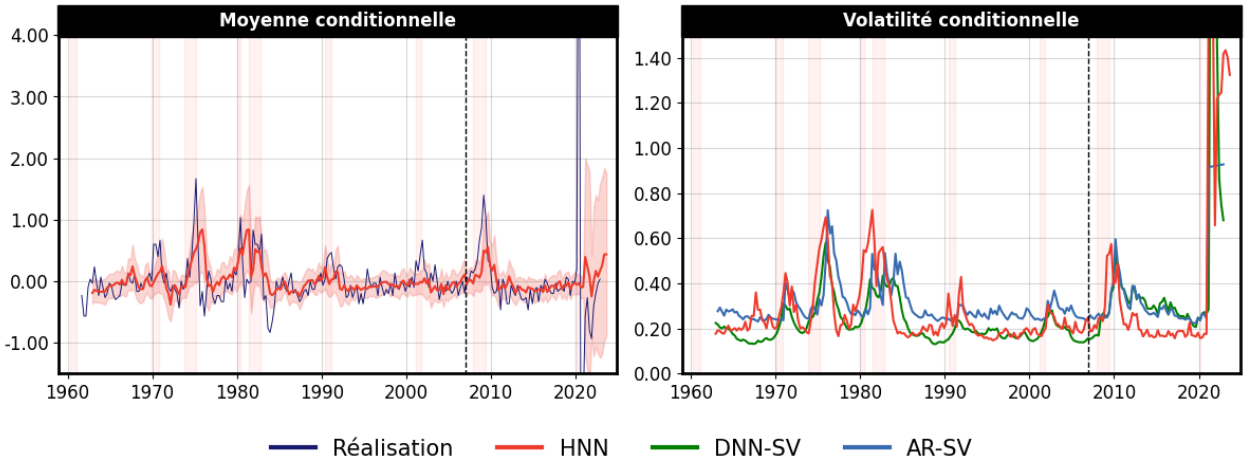

Figure 18: HNN - Taux de chômage (h=4)

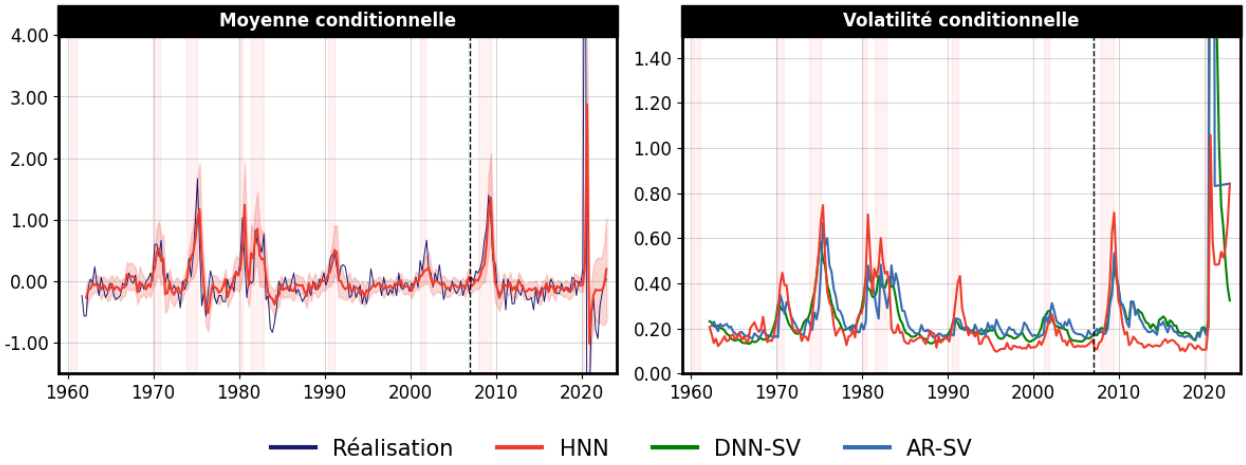

Figure 19: HRNN - Taux de chômage (h=1)

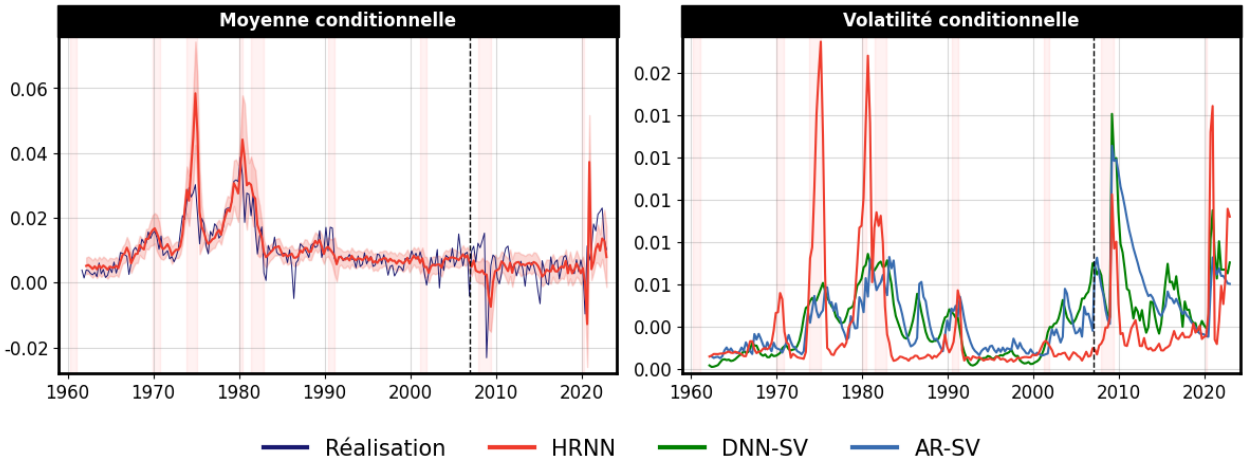

Figure 20: HRNN - Inflation (h=1)

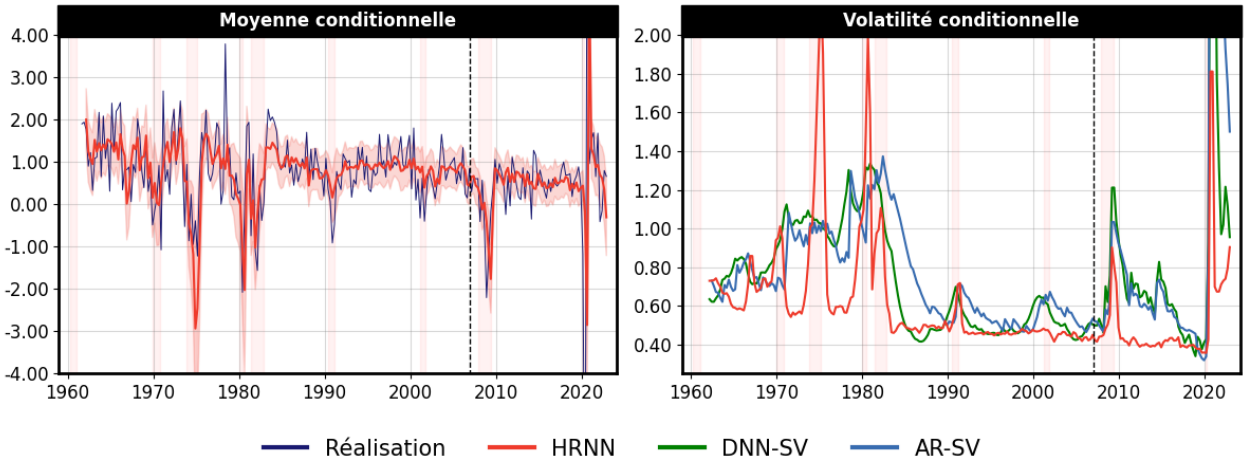

Figure 21: HRNN - Produit Intérieur Brut (h=1)

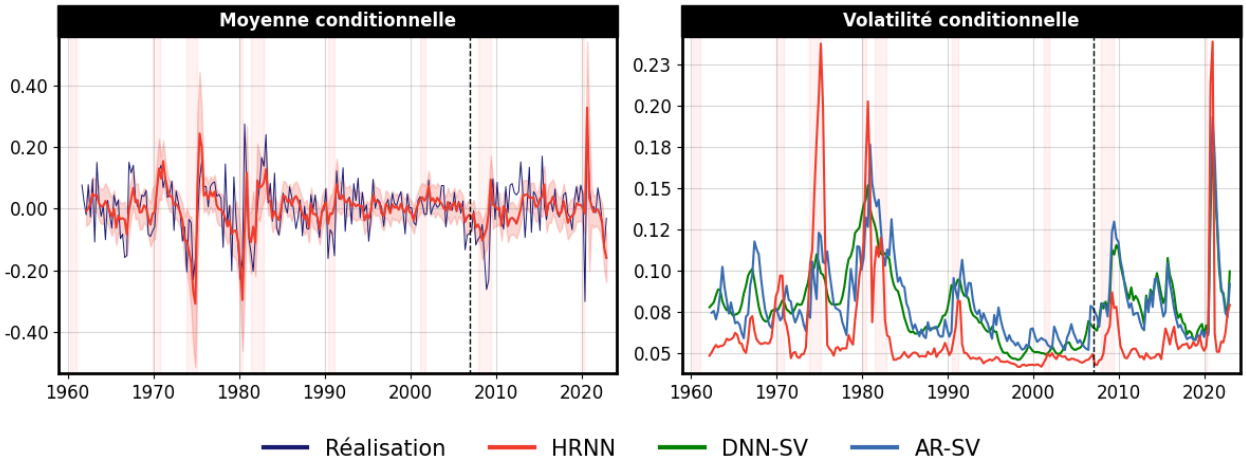

Figure 22: HRNN - Mise en chantier de résidence privé (h=1)

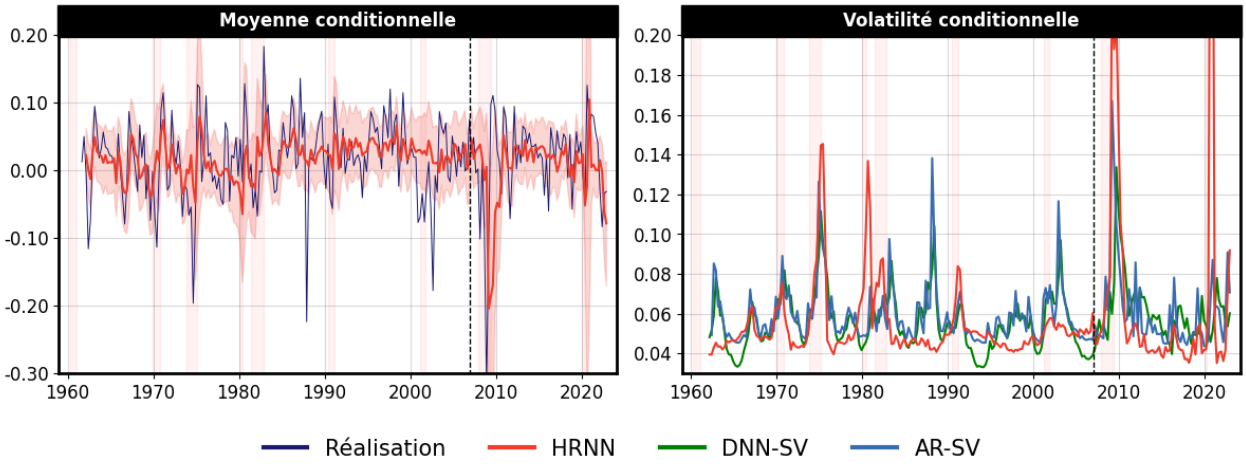

Figure 23: HRNN - S&P 500 (h=1)

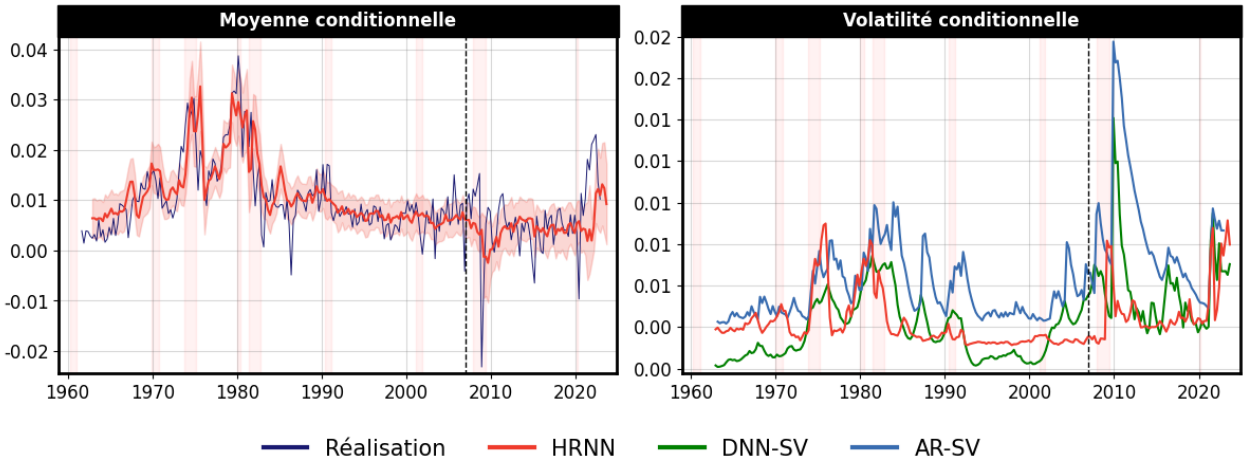

Figure 24: HRNN - Inflation (h=4)

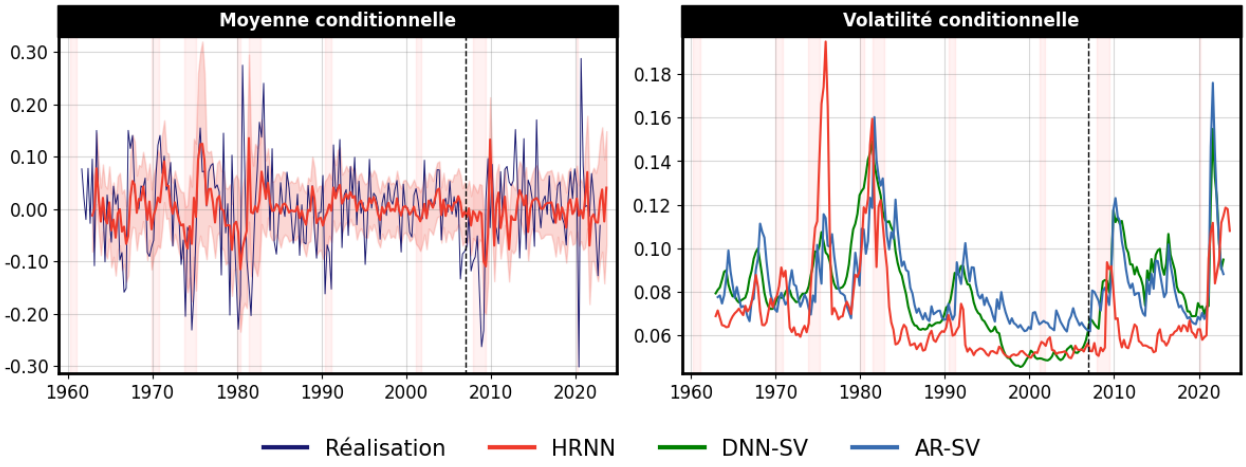

Figure 25: HRNN - Mise en chantier de résidence privé (h=4)

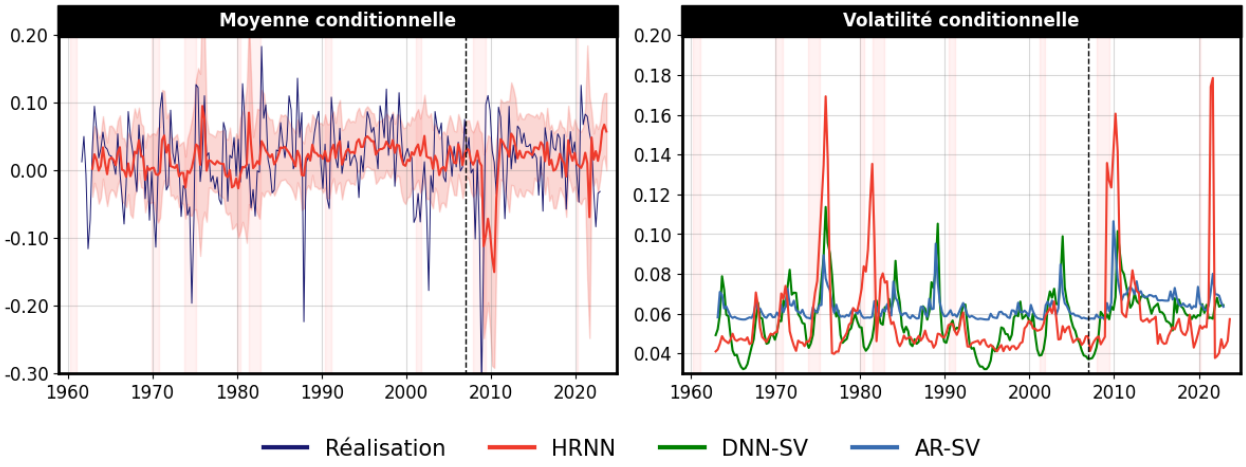

Figure 26: HRNN - S&P 500 (h=4)

C Importance de variable

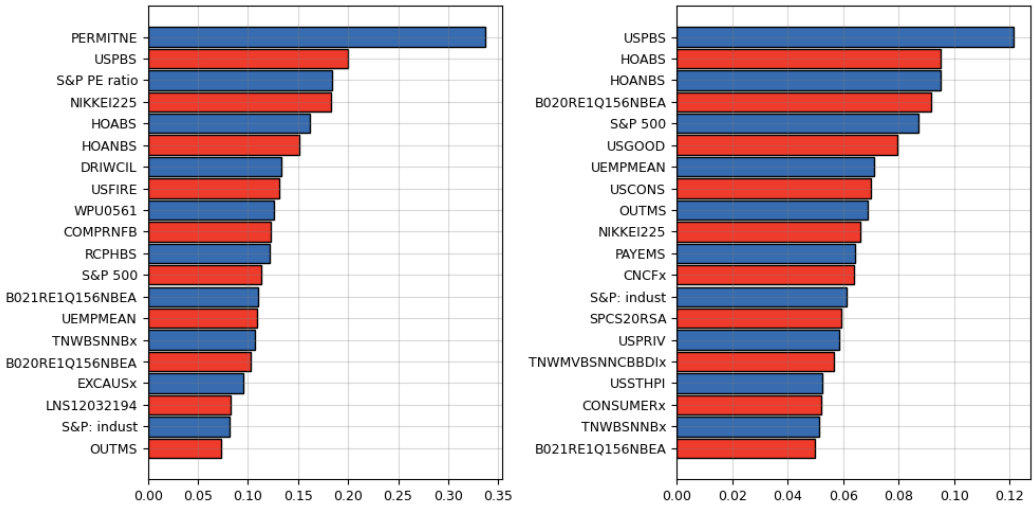

Figure 27: Importance des variables pour le taux de chômage. Dans le panneau de gauche, on présente l'importance des variables pour l'hémisphère de la moyenne conditionnelle, et dans le panneau de droite, celle de la variance conditionnelle.

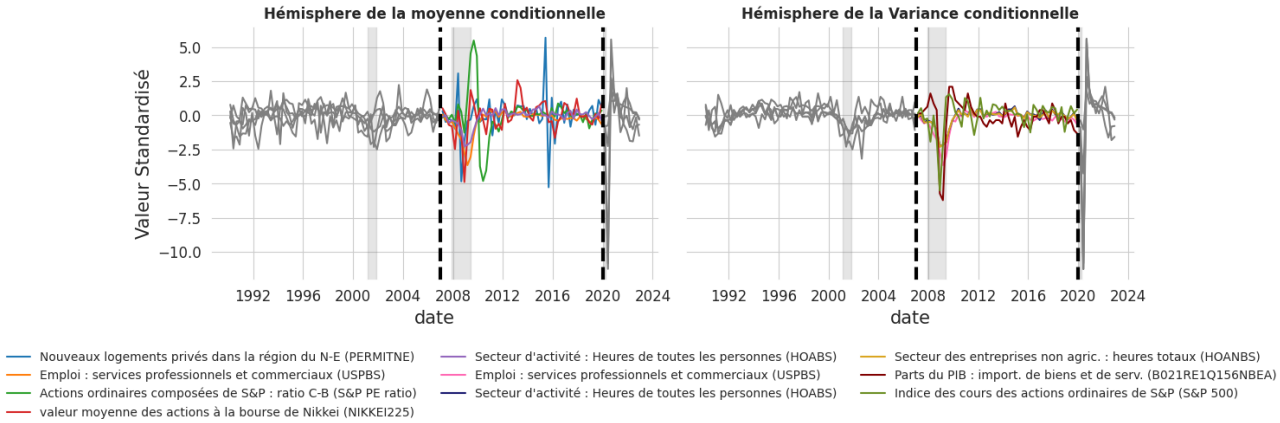

Figure 28: Graphique des variables importantes par hémisphère pour le taux de chômage. Dans le panneau de gauche, on trouve les variables qui ont le plus d'influence sur le taux de chômage durant la période allant de 2007Q1 à 2019Q4 pour l'hémisphère de la moyenne conditionnelle. À droite, celles de l'hémisphère de la variance conditionnelle.

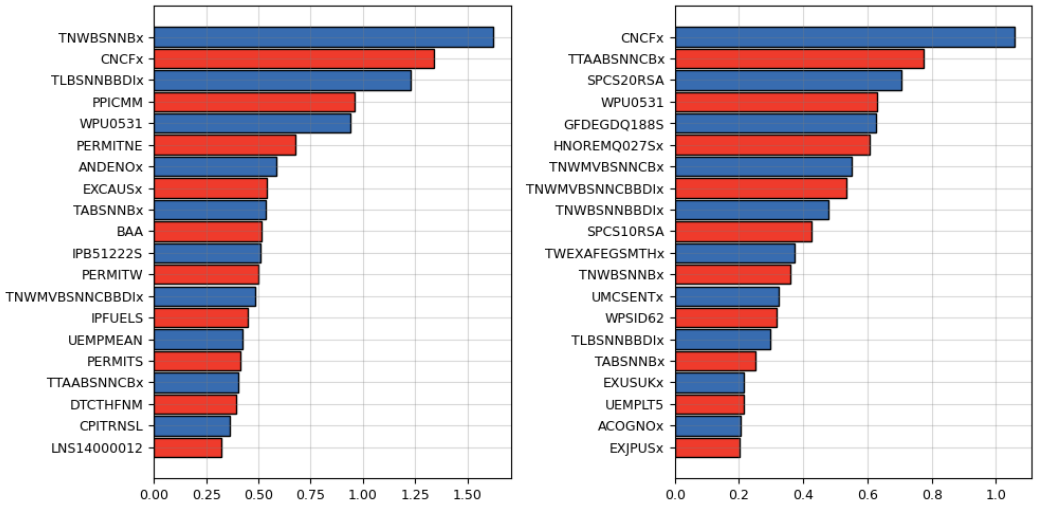

Figure 29: Importance des variables pour la mise en chantier de résidences privées. Dans le panneau de gauche, on présente l'importance des variables pour l'hémisphère de la moyenne conditionnelle, et dans le panneau de droite, celle de la variance conditionnelle.

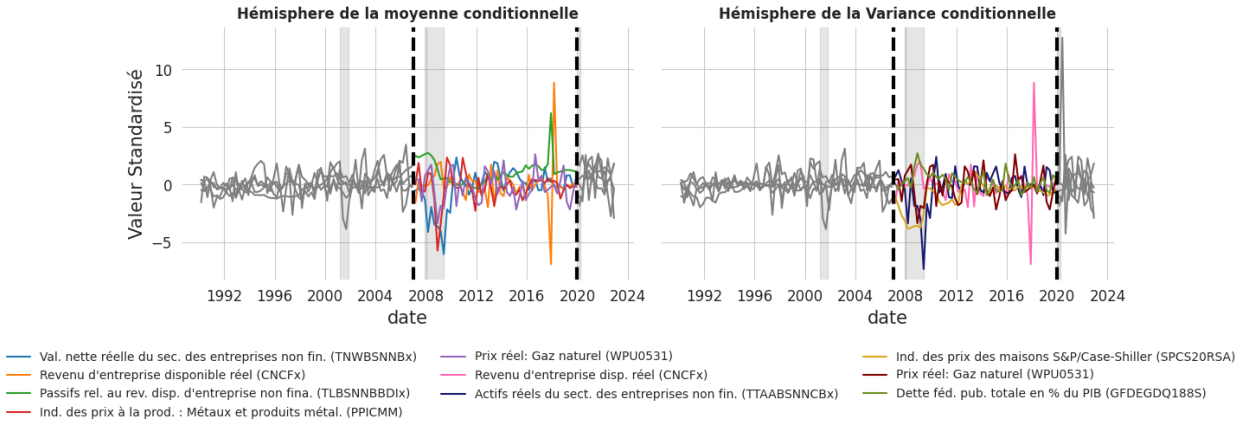

Figure 30: Graphique des variables importantes par hémisphère pour la mise en chantier de résidences privées. Dans le panneau de gauche, on trouve les variables qui ont le plus d'influence sur la mise en chantier de résidences privées durant la période allant de 2007Q1 à 2019Q4 pour l'hémisphère de la moyenne conditionnelle. À droite, celles de l'hémisphère de la variance conditionnelle.

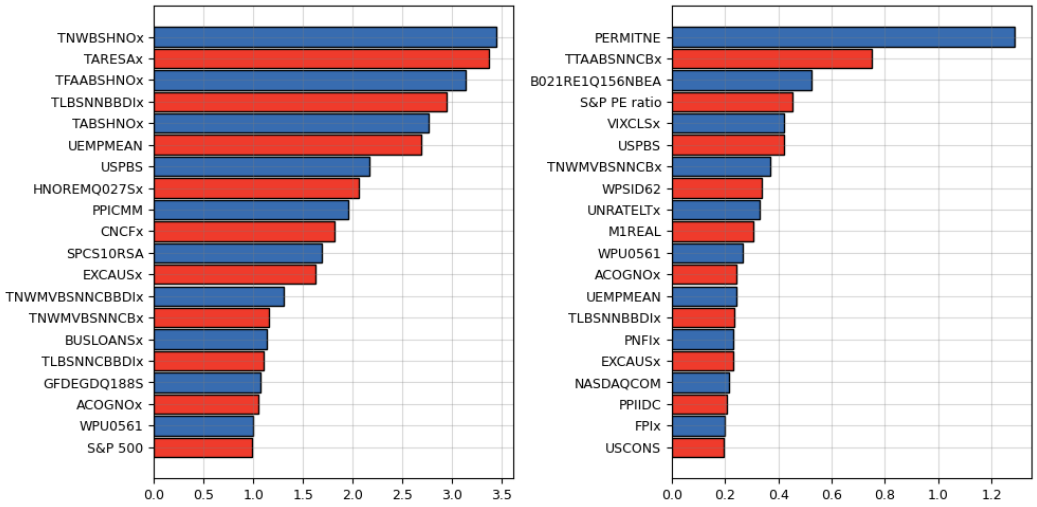

Figure 31: Importance des variables pour le S&P 500. Dans le panneau de gauche, on présente l'importance des variables pour l'hémisphère de la moyenne conditionnelle, et dans le panneau de droite, celle de la variance conditionnelle.

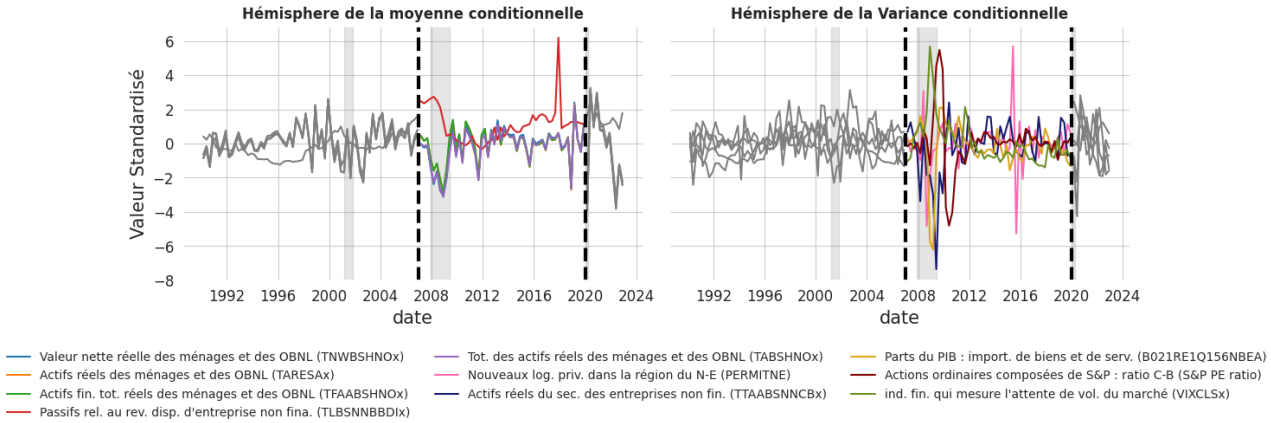

Figure 32: Graphique des variables importantes par hémisphère pour le S&P 500. Dans le panneau de gauche, on trouve les variables qui ont le plus d'influence sur le S&P 500 durant la période allant de 2007Q1 à 2019Q4 pour l'hémisphère de la moyenne conditionnelle. À droite, celles de l'hémisphère de la variance conditionnelle.

BIBLIOGRAPHIE

Adams, P. A., Adrian, T., Boyarchenko, N. et Giannone, D. (2021). Forecasting macroeconomic risks. *International Journal of Forecasting*, *37*(3), 1173–1191.

Adrian, T., Boyarchenko, N. et Giannone, D. (2019). Vulnerable growth. *American Economic Review*, *109*(4), 1263–89.

Almosova, A. et Andresen, N. (2023). Nonlinear inflation forecasting with recurrent neural networks. *Journal of Forecasting*, *42*(2), 240–259.

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J. et Mané, D. (2016). Concrete problems in ai safety. *arXiv preprint arXiv:1606.06565*.

Bao, Y., Lee, T.-H. et Saltoğlu, B. (2007). Comparing density forecast models. *Journal of Forecasting*, *26*(3), 203–225.

Bassetti, F., Casarin, R. et Ravazzolo, F. (2020). Density forecasting. *Macroeconomic Forecasting in the Era of Big Data: Theory and Practice*, 465–494.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, *19*(4), 465–474.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*(3), 307–327.

Box, G. et Jenkins, G. (1970). *Time Series Analysis, Forecasting and Control*. Holden-Day.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*(2), 123–140.

Caldara, D., Cascaldi-Garcia, D., Cuba-Borda, P. et Loria, F. (2021). Understanding growth-at-risk: A markov switching approach. *Available at SSRN 3992793*.

Cauchy, A. *et al.* (1847). Méthode générale pour la résolution des systemes d'équations simultanées. *Comp. Rend. Sci. Paris*, *25*(1847), 536–538.

Chatfield, C. (1993). Calculating interval forecasts. *Journal of Business & Economic Statistics*, *11*(2), 121–135.

Chipman, H. A., George, E. I. et McCulloch, R. E. (2010). Bart: Bayesian additive regression trees.

Clark, T. E., Huber, F., Koop, G., Marcellino, M. et Pfarrhofer, M. (2022). Tail forecasting with multivariate bayesian additive regression trees. *International Economic Review*.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, *2*(4), 303–314.

Delle Monache, D., De Polis, A. et Petrella, I. (2020). Modeling and forecasting macroeconomic downside risk.

Dempster, A. P., Laird, N. M. et Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)*, *39*(1), 1–22.

Diebold, F. X., Gunther, T. A. et Tay, A. (1997). Evaluating density forecasts.

Diebold, F. X. et Lopez, J. A. (1996). 8 forecast evaluation and combination. *Handbook of statistics*, *14*, 241–268.

Diebold, F. X., Shin, M. et Zhang, B. (2022). On the aggregation of probability assessments: Regularized mixtures of predictive densities for eurozone inflation and real interest rates. *Journal of Econometrics*.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, 987–1007.

Farrell, M. H., Liang, T. et Misra, S. (2021). Deep neural networks for estimation and inference. *Econometrica*, *89*(1), 181–213.

Gawlikowski, J., Tassi, C. R. N., Ali, M., Lee, J., Humt, M., Feng, J., Kruspe, A., Triebel, R., Jung, P., Roscher, R. *et al.* (2021). A survey of uncertainty in deep neural networks. *arXiv preprint arXiv:2107.03342*.

Gelman, A., Carlin, J. B., Stern, H. S. et Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC.

Gneiting, T., Balabdaoui, F. et Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *69*(2), 243–268.

Gneiting, T. et Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, *102*(477), 359–378.

Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society: Series B (Methodological)*, *14*(1), 107–114.

Goodfellow, I., Bengio, Y. et Courville, A. (2016). *Deep learning*. MIT press.

Goodfellow, I. J., Shlens, J. et Szegedy, C. (2014). Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*.

Goulet Coulombe, P. (2022). A neural phillips curve and a deep output gap. *arXiv preprint arXiv:2202.04146*.

Goulet Coulombe, P., Leroux, M., Stevanovic, D. et Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*, *37*(5), 920–964.

Gouriéroux, C. (1997). *ARCH models and financial applications*. Springer Science & Business Media.

Granger, C. W. et Pesaran, M. H. (2000). Economic and statistical measures of forecast accuracy. *Journal of Forecasting*, *19*(7), 537–560.

Guidolin, M., La Cara, D. et Marcellino, M. G. (2021). Boosting the forecasting power of conditional heteroskedasticity models to account for covid-19 outbreaks. *BAFFI CAREFIN Centre Research Paper*, (2021-169).

Guo, C., Pleiss, G., Sun, Y. et Weinberger, K. Q. (2017). On calibration of modern neural networks. Dans *International conference on machine learning*, 1321–1330. PMLR.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the econometric society*, 357–384.

Hinton, G. E., Osindero, S. et Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, *18*(7), 1527–1554.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I. et Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*.

Hoerl, A. E. et Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.

Hornik, K., Stinchcombe, M. et White, H. (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural networks*, *3*(5), 551–560.

Hull, J. et White, A. (1987). The pricing of options on assets with stochastic volatilities. *The journal of finance*, *42*(2), 281–300.

Ioffe, S. et Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. Dans *International conference on machine learning*, 448–456. PMLR.

Jacquier, E., Polson, N. G. et Rossi, P. E. (2002). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, *20*(1), 69–87.

Kastner, G. et Frühwirth-Schnatter, S. (2014). Ancillarity-sufficiency interweaving strategy (asis) for boosting mcmc estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, *76*, 408–423.

Kroese, D. P. et Rubinstein, R. Y. (2012). Monte carlo methods. *Wiley Interdisciplinary Reviews: Computational Statistics*, *4*(1), 48–58.

Lakshminarayanan, B., Pritzel, A. et Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, *30*.

LeCun, Y., Bengio, Y. et Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436–444.

Luo, R., Zhang, W., Xu, X. et Wang, J. (2018). A neural stochastic volatility model. Dans *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

McCracken, M. et Ng, S. (2020). *FRED-QD: A quarterly database for macroeconomic research*. Rapport technique, National Bureau of Economic Research.

McCulloch, W. S. et Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, *5*(4), 115–133.

Medeiros, M. C., Teräsvirta, T. et Rech, G. (2006). Building neural network models for time series: a statistical approach. *Journal of Forecasting*, *25*(1), 49–75.

Montavon, G., Samek, W. et Müller, K.-R. (2018). Methods for interpreting and understanding deep neural networks. *Digital signal processing*, *73*, 1–15.

Moshiri, S. et Cameron, N. (2000). Neural network versus econometric models in forecasting inflation. *Journal of forecasting*, *19*(3), 201–217.

Nakamura, E. (2005). Inflation forecasting using a neural network. *Economics Letters*, *86*(3), 373–378.

Paranhos, L. (2021). Predicting inflation with neural networks. *arXiv preprint arXiv:2104.03757*.

Rumelhart, D. E., Hinton, G. E., McClelland, J. L. *et al.* (1986a). A general framework for parallel distributed processing. *Parallel distributed processing: Explorations in the microstructure of cognition*, *1*(45-76), 26.

Rumelhart, D. E., Hinton, G. E. et Williams, R. J. (1986b). Learning representations by back-propagating errors. *nature*, *323*(6088), 533–536.

Salinas, D., Flunkert, V., Gasthaus, J. et Januschowski, T. (2020). Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, *36*(3), 1181–1191.

Smalter Hall, A. et Cook, T. R. (2017). Macroeconomic indicator forecasting with deep neural networks. *Federal Reserve Bank of Kansas City Working Paper*, (17-11).

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. et Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, *15*(1), 1929–1958.

Stock, J. H. et Watson, M. W. (1998). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series.

Stock, J. H. et Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–162.

Swanson, N. R. et White, H. (1997). A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *Review of Economics and Statistics*, *79*(4), 540–550.

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes-a study of the daily sugar prices 1961-75. *Time series analysis: theory and practice*, *1*, 203–226.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Tibshirani, R. J. et Efron, B. (1993). An introduction to the bootstrap. *Monographs on statistics and applied probability*, *57*(1).

White, H. (1989). Some asymptotic results for learning in single hidden-layer feedforward network models. *Journal of the American Statistical association*, *84*(408), 1003–1013.

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.

Yin, Z. et Barucca, P. (2022). Neural generalised autoregressive conditional heteroskedasticity. *arXiv preprint arXiv:2202.11285*.