

APPLICATION DE L'APPRENTISSAGE PROFOND AU PROVISIONNEMENT AGRÉGÉ NON-VIE

Wendinmanegdé Asaph TAPSOBA

01/02/2024

Mémoire encadré par :



Christian-Yann ROBERT
Tuteur académique



Yanis BENATTIA & Philippe ZOUNGRANA
Tuteurs professionnels

SOMMAIRE

1

Cadre général

2

Théorie de l'intégration des modèles paramétriques aux réseaux de neurones

3

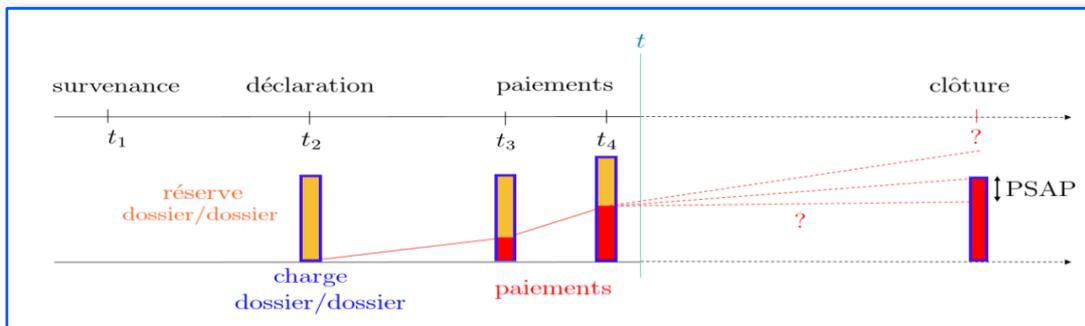
Mise en œuvre des modèles

4

Conclusion

- 1 • Les provisions pour sinistres à payer

Cycle de vie d'un sinistre en Assurance IARD



Pour une année d'accident donnée, deux sources d'incertitudes influent sur la charge des sinistres à l'ultime

PSAP/IBNR

Une incertitude sur les montants des paiements

Les sinistres survenus peuvent donner lieu à des paiements au cours des exercices suivants → incertitudes sur le montant total des paiements à l'ultime

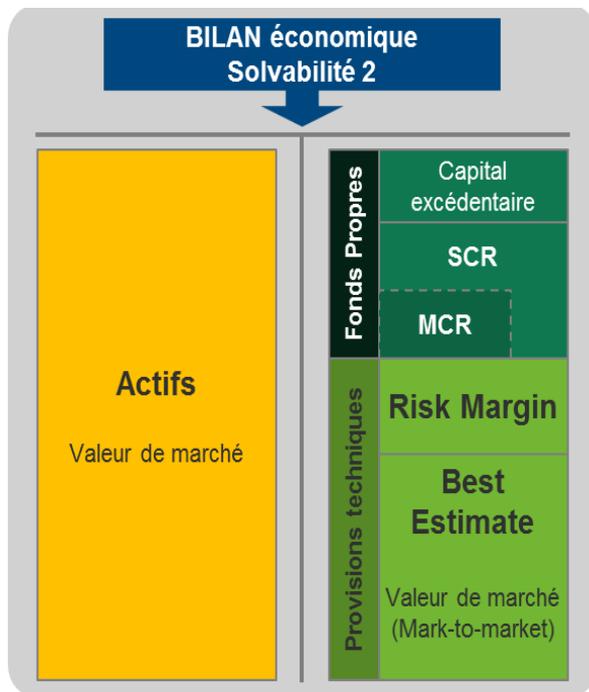
IBNER (*Inced But Not Enough Reported*)

Une incertitude sur le nombre de sinistres survenus au cours de l'année

Etant donné les délais de déclaration, certains sinistres qui ont lieu au cours d'une année donnée peuvent être déclarés au cours des années ultérieures.

IBNYR (*Inced But Not Yet Reported*)

- 1.2 • Enjeux liés à l'estimation des réserves



- Enjeux réglementaires et de communication financière (S2 & IFRS)
 - Provisions évaluées en *Best Estimate*
 - *Seconde opinion des réserves* : disposer des plusieurs méthodes en fonction de l'adéquation des hypothèses avec les données → quantifier et mitiger le risque de modèle
- Bilan
 - Les provisions constituent le post le plus important du passif de l'assureur
 - Une bonne connaissance de la valeur de l'entreprise et de sa solvabilité passe par une bonne évaluation des provisions
- Compte de résultat
 - Les variations des provisions d'une année à l'autre sont enregistrées en charge au résultat
 - Une volatilité des PSAP entraîne donc une volatilité du résultat et impacte donc la communication financière de la boîte

- 1 • Les différentes méthodes de provisionnement

Différentes approches de provisionnement

- ❑ **Approches individuelles** : PSAP calculée par groupes de sinistres homogènes

- ❑ **Approches collectives** :
 - données agrégées en triangle de développement
 - PSAP calculées par année d'accident

Quels modèles pour le calcul de la PSAP ?

- ❑ **Modèles basés sur le machine learning** :
 - **Avantages** : Possibilité de modéliser des fonctions hautement non-linéaires.
 - **Limites** : Implémentation plus complexe, interprétabilité, besoin en données individuelles

- ❑ **Modèles classiques** : CL et variantes, GLM, ...
 - **Avantages** : Simples à implémenter et à interpréter
 - **Limites** : Hypothèses souvent fortes sur la structure des données

Présentation générale

- ❑ **Combien de modèles allons-nous voir?**
 - Le modèle NNODP : permet de prédire une seule variable cible (les paiements incrémentaux)
 - Le modèle NNDODP : extension du modèle NNODP, permet de prédire 2 variables cibles (paiements incrémentaux et les nombres de sinistres déclarés)

- ❑ **Quelle est l'idée générale de ces deux modèles?**
 - Intégrer un modèle classique de régression paramétrique dans une architecture de réseau de neurones
 - Initialisation des paramètres du réseau par les résultats du modèle classique
 - Le réseau explore lors de l'apprentissage une structure qui va au delà de celle du modèle classique

- ❑ **Quelles sont les principaux avantages et inconvénients des modèles?**
 - Avantages : rapidité d'apprentissage, flexibilité, mesure d'incertitude, interprétabilité
 - Inconvénients : contraintes opérationnelles, données, complexité

- 2.2 • Architecture des modèles étudiés

Modèle classique : GLM Poisson avec sur-dispersion

- ❑ Nous supposons qu'il existe un paramètre de dispersion $\phi > 0$ et que les paiements incrémentaux $(X)_{i,j \in \{1, \dots, n\} \times \{0, \dots, n-1\}}$ sont iid
 - $E[X_{i,j}] = \mu_{i,j}$ et $V[X_{i,j}] = \phi \mu_{i,j}$, $\mu_{i,j}$ étant un paramètre positif
 - $X_{i,j} / \phi \sim Poi(\mu_{i,j} / \phi)$
- ❑ En supposant en plus qu'il existe un paramètre d'exposition $\alpha_i \forall i \in \{1, \dots, n\}$ et un paramètre de développement $\beta_j \forall j \in \{0, \dots, n-1\}$ tels que $\ln \mu_{i,j} = \alpha_i + \beta_j$, nous retombons sur les estimateurs du modèle de Chain-Ladder avec :
 - $E(X_{i,j}) = \frac{1}{\phi} V(X_{i,j}) = \mu_{i,j} = \exp\{\alpha_i + \beta_j\}$

Modèle de Chain-Ladder

Le modèle de CL suppose qu'il existe un facteur de développement $f_j \forall j \in \{0, \dots, n-1\}$ tel que $\forall i \in \{1, \dots, n\}$ (l'année d'accident) :

$$C_{i,j+1} = f_j * C_{i,j}$$

$C_{i,j}$ étant le montant cumulé des paiements réalisés au titre de l'année d'accident i après j années de développement.

La méthode de CL suppose donc que l'évolution des paiements cumulés d'une année de développement à l'autre reste identique pour toutes les années de survénance (le facteur de développement ne dépend pas de l'année de survénance)

Le montant des provisions est déterminé par année d'accident. Pour une année d'accident donné $i \in \{1, \dots, n\}$, la PSAP est définie par :

$$R_i = C_{i,n} - C_{i,n+i-1}$$

- $C_{i,n}$ correspond au montant de paiement cumulé à l'ultime pour l'année d'accident i
- $C_{i,n+i-1}$ correspond au montant de paiement cumulé de l'année calendaire en cours pour l'année d'accident i
- On suppose qu'il n'y a pas de paiement au-delà de n années de développement

Nous travaillerons avec des triangles de paiement nets de recours, nous ne considérons pas de réserve dossier/dossier

- 2.3 • Architecture des modèles étudiés

Limites du modèle GLM Poisson avec sur-dispersion

- l'hypothèse d'indépendance des paiements incrémentaux : hypothèse forte en pratique
- la nécessité de disposer de paiements incrémentaux positifs. En pratique, suite à des recours, les triangles de paiements peuvent présenter des incréments négatifs

Limites du modèle Chain Ladder

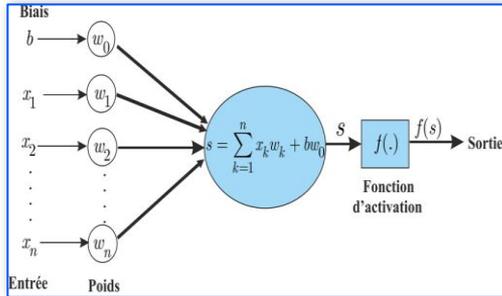
- **Absence d'effet calendaire** : des facteurs comme un changement de législation de jurisprudence, de management au sein de l'entreprise, ..., peuvent avoir une influence non négligeable sur les années de souscription et remettre en cause cette hypothèse.
- **La linéarité entre les années de développement j et $j+1$** : Cette hypothèse ne tient pas compte des évolutions probables au cours du déroulement des sinistres comme par exemple le changement des clauses contractuelles, de composition du portefeuille, de profil de risque...
- Estimation des derniers facteurs de développement se font avec de moins en moins de données en particulier le dernier facteur f_{n-1} qui dépend uniquement des paiement cumulés $C_{i,n}$ et $C_{i,n-1}$

Qu'est ce qu'un réseau de neurones feedforward?

Une composition de transformations linéaires suivies de fonctions d'activation non-linéaires

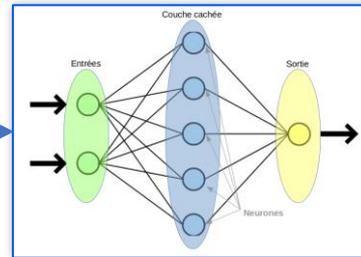
Perceptron : unité élémentaire (neurone)

Sortie := fonction non linéaire de la combinaison linéaire des variables en entrée



Réseau de neurones peu profond

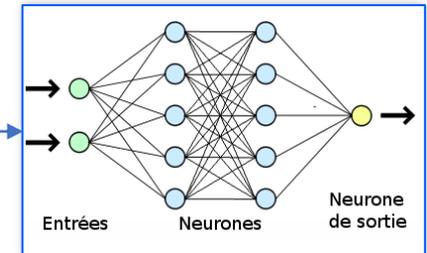
Couche cachée := empilement de K neurones
Couche de sortie := fonction non linéaire de la combinaison linéaire des neurones de la couche cachée



Réseau de neurones peu profond

Le réseau est défini par :

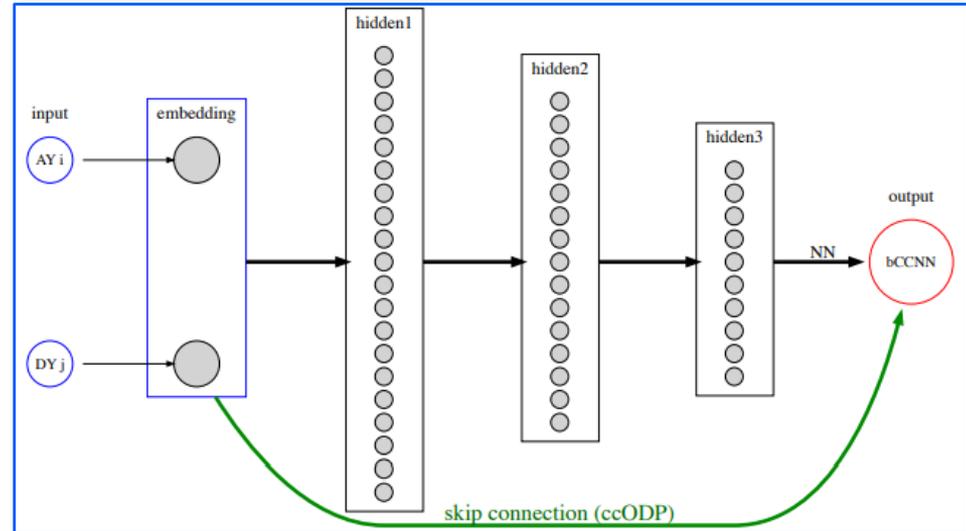
- Une couche d'entrée
- D Couches cachées
- Une couche de sortie : fonction non linéaire de la combinaison linéaire des neurones de la dernière couche cachée D-1



- 2.5 • Architecture des modèles étudiés

Modèle NNODP (Neural Network Overdispersed Poisson Model)

- ❑ **Entrée** : Année d'Accident AY et Année de développement DY
- ❑ **Sortie** : Montant du paiement ou de la charge incrémentale X_{ij}
- ❑ Une couche intégrée en entrée pour le traitement des variables catégorielles AY, DY
- ❑ Un réseau feedforward à D couches
- ❑ Une skip connexion

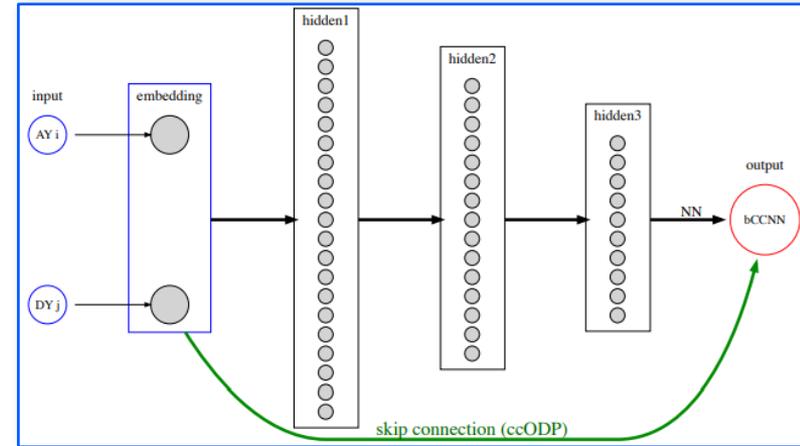


Comment formaliser mathématiquement le modèle NNODP?

Fonction de régression du modèle :

$$\mu_{i,j} = \exp\left\{w(\underbrace{\alpha_i + \beta_j}_{\text{Skip connexion}}) + \underbrace{C_D + B'_D z^{(D-1)}}_{\text{réseau de neurones}}\right\}$$

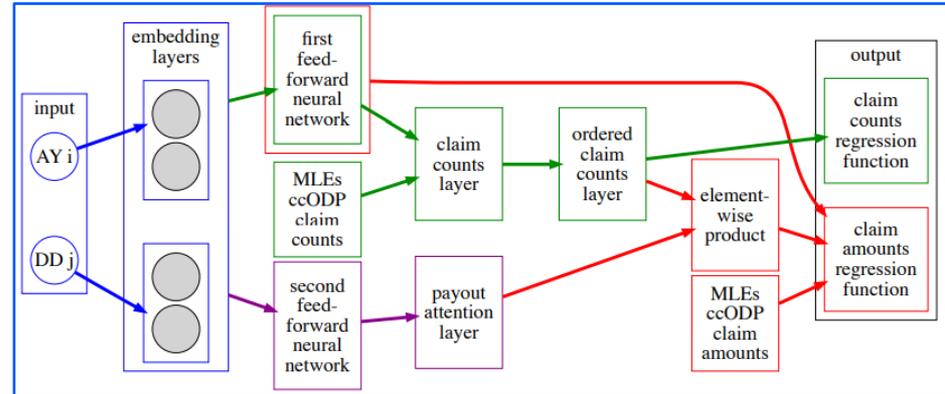
En prenant à $t=0$, $\alpha_i = \widehat{\alpha}_i^{ODP}$, $\beta_j = \widehat{\beta}_j^{ODP}$, $C_D = \widehat{c}^{ODP}$, $B'_D = 0$, $w = 1$
Le modèle commence bien avec l'estimation du modèle de Chain-Ladder



- 2.7 • Architecture des modèles étudiés

Modèle NNDODP (Neural Network Double Overdispersed Poisson Model)

- ❑ Entrée : Année d'Accident AY_i et Année de développement DY_j
- ❑ Sortie : Paiement ou Charge X_{ij} et Nombre N_{ij}
- ❑ Un modèle NNODP sur les paiements et un modèle NNODP sur les nombres de sinistres déclarés sont intégrés dans une architecture unique
- ❑ Interaction entre les nombres et les montants de paiements



- 3.1 • Quelles sont les données utilisées ?

Format des données utilisées

- Données individuelles
 - Regroupées en triangle de développement
 - Variables d'intérêt : année d'accident, année de développement, charges ou paiements incrémentaux, nombre de sinistres déclarés
- Branche MRH
 - Données simulées (Simulateur de Mario Wüthrich)
 - ✓ Historiques complets de sinistralités individuelles
 - ✓ Possibilité de calculer les vraies réserves

	LoB1	LoB2	LoB3	LoB4	LoB5	LoB6
Vraies Réserves	80 806	80 811	18 359	150 061	151 597	32 193
Réserves de Chain Ladder	76 171	75 730	17 540	139 394	145 859	31 411
Biais	-4 635	-5 081	-819	-11 567	-5 938	-782
Paramètre de dispersion	15	21	3	28	22	6

Claims Development Result (CDR)

$$\begin{aligned} \text{CDR}_i^k &= C_{i,n-1}^k - C_{i,n-1}^{k-1} \\ &= R_{i,j}^k - R_{i,j}^{k-1} + (X_{i,j}^* - \widehat{X_{i,j}^{k-1}}) \\ &= \Delta R_i + \text{Av}V_{i,j}^k \end{aligned}$$

- ✓ **CDR** différence entre les prévisions des sinistres ultimes au cours de périodes consécutives
- ✓ **AvV_{i,j}^k** mesure la qualité de la prédiction fournie par le modèle à horizon un an
- ✓ **ΔR_i** mesure la stabilité des estimations des réserves

Remarque :

Il faut noter qu'en fonction des objectifs de la modélisation, il est possible de modifier le CDR afin d'introduire des pondération différentes entre le et l' comme suit :

$$\begin{aligned} \text{CDR}_i^{k,\alpha} &= C_{i,n-1}^k - C_{i,n-1}^{k-1} \\ &= \alpha * \Delta R_i + (1 - \alpha) * \text{Av}V_{i,j}^k \end{aligned}$$

Les scores sont définies comme suit :

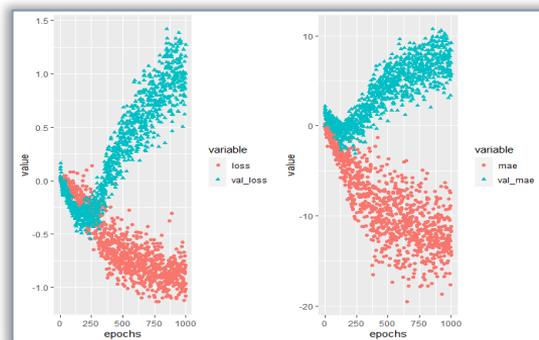
$$\begin{aligned} \text{CDR}_{score} &= \sqrt{\frac{\sum_{i=1}^I |X_{i,j}^*| (\text{CDR}_i^k)^2}{\sum_{i=1}^I |X_{i,j}^*|}} \\ \text{Av}E_{score} &= \sqrt{\frac{\sum_{i=1}^I |X_{i,j}^*| (\text{Av}E_i^k)^2}{\sum_{i=1}^I |X_{i,j}^*|}} \\ \Delta R_{score} &= \sqrt{\frac{\sum_{i=1}^I |X_{i,j}^*| (\Delta R_i^k)^2}{\sum_{i=1}^I |X_{i,j}^*|}} \end{aligned}$$

Ces métriques seront évaluées par année de survenance.

Par soucis de comparaison entre les LoBs, nous allons considérer un score pour chaque métrique qui sera une moyenne pondérée de la valeur de la métrique en question sur toutes les années d'accidents. La pondération se fera en fonction des charges incrémentales de l'année calendaire en cours.

- 3.3 • Calibrage des modèles

- ❑ La première étape de l'apprentissage est la sélection des modèles :
 - Choix des hyperparamètres : nombre de couches cachées n , nombre de neurones par couche
- ❑ Base d'apprentissage/ base de validation/ base de test :
 - portefeuille de sinistres est divisé aléatoirement en 2 parties égales
 - Les sous-portefeuilles sont ensuite agrégés en triangles supérieurs d'apprentissage et de validation et en triangles inférieurs de test.
- ❑ Sur-apprentissage vs sous-apprentissage (tension entre optimisation et généralisation) :
 - Choix du nombre d'epochs
 - Régularisation : ajouter des contraintes sur les informations qu'il est autorisé à stocker
 - Choix des hyperparamètres : modèle parcimonieux
 - Régularisation L1/L2
 - Ajout de couche drop-out



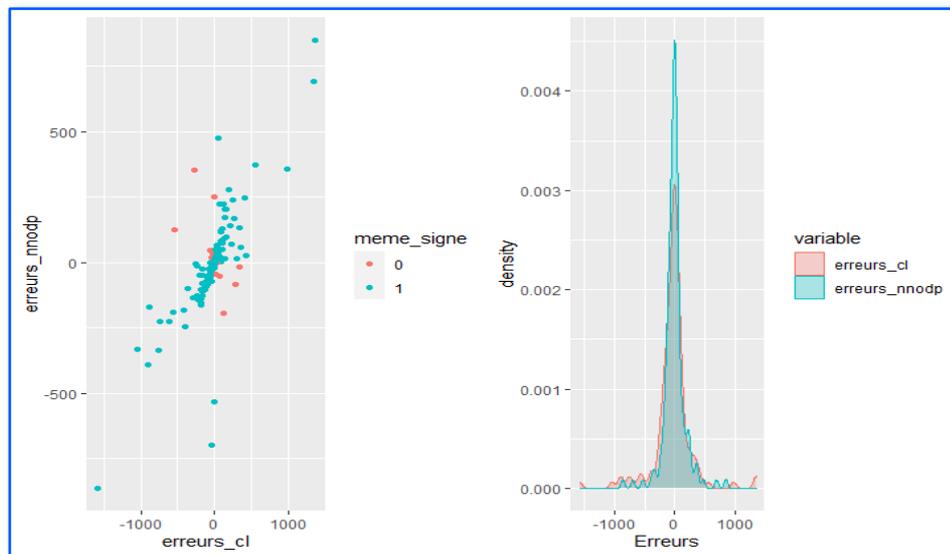
- 3.4 • Résultats sur les données simulées

Fonction de perte	LoB1	LoB2	LoB3	LoB4	LoB5	LoB6
Initiale	-32 029	-32 474	-5 252	-51 570	-52 197	-8 503
Finale	-52 086	-52 957	-8 410	-81 048	-82 094	-13 296
Variation	63%	63%	60%	57%	57%	56%

Biais = Vraie Res – Res prédite	LoB1	LoB2	LoB3	LoB4	LoB5	LoB6
Chain Ladder	6%	6%	4%	8%	4%	2%
NNODP	2%	1%	2%	3%	-1%	0%
NNDODP	2%	-1%	-3%	-1%	1%	-1%

- Diminution sensible de la fonction de perte (la déviance du modèle de poisson) pour les deux modèles
- Diminution sensible du biais pour les deux modèles
- Modèle NNDODP plus conservateur (sous-estime moins souvent les vraies réserves)

- 3.5 • Analyse des erreurs de prédiction



- ❑ Corrélation forte entre les erreurs
- ❑ Amplitude des erreurs faible sur le modèle NNODP et NNDODP
- ❑ Masse plus élevée au voisinage de zéro

- 3.6 • Scores de la branche MRH

Métriques	Chain Ladder	NNODP	NNDODP
AvE	14 348	14 100	13 720
ΔR	5 262	4 296	2 869
CDR	19 824	18 552	15 222

- Les deux modèles affichent des performances meilleures par rapport au modèle de Chain-Ladder
- L'ajout de l'information sur les nombres de sinistres déclarés est d'un apport notable

Mise en pratique des modèles

- 3.7 • Mesure de l'incertitude de prédiction

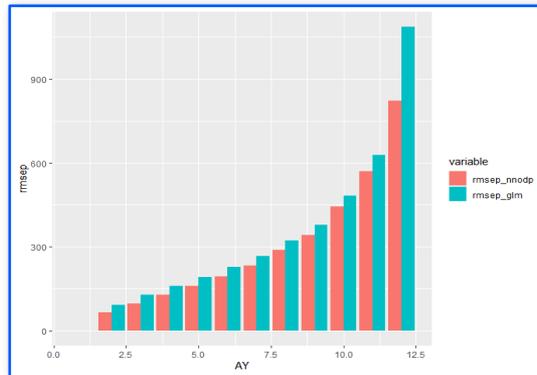
- Erreur quadratique moyenne des prédictions msep

$$msep(R_n|D_n) = V(R_n|D_n) + err(R_n|D_n)$$

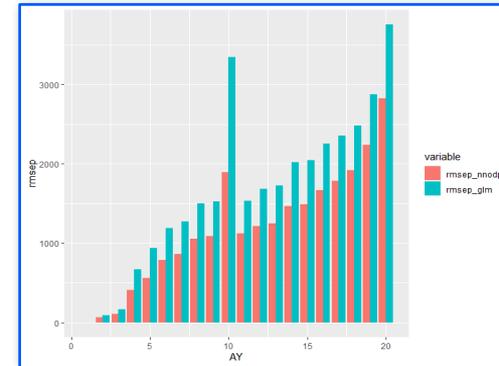
- L'erreur de process est estimé par $\hat{V}(R_n|D_n) = \hat{\phi}R_n$

- L'erreur d'estimation est approchée par une procédure de bootstrap

- Simulation de N triangles supérieurs
- Calibration du modèle et calcul des réserves pour chacun des N triangles
- Estimation de $err(R_n|D_n)$ par la variance empirique de la distribution obtenue



MSEP conditionnelle des réserves prédites par année d'accident pour la LoB1 et le Modèle NNODP



MSEP conditionnelle des réserves prédites par année d'accident pour la branche MRH et le Modèle NNODP

Mise en pratique des modèles

- 3.8 • Les coefficients de passage

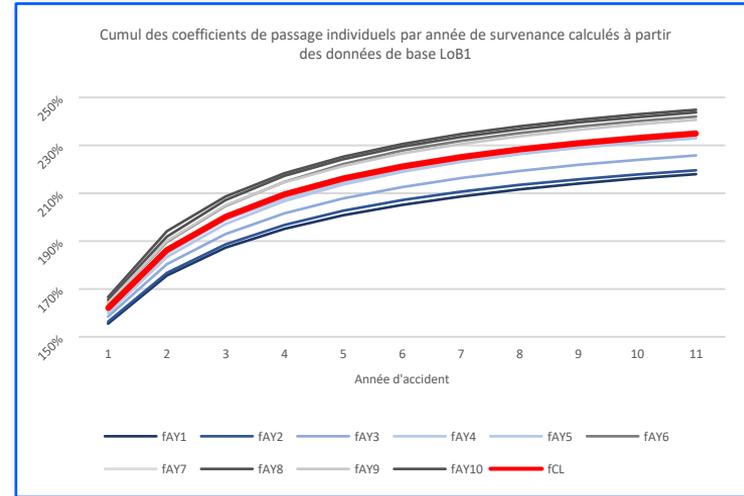
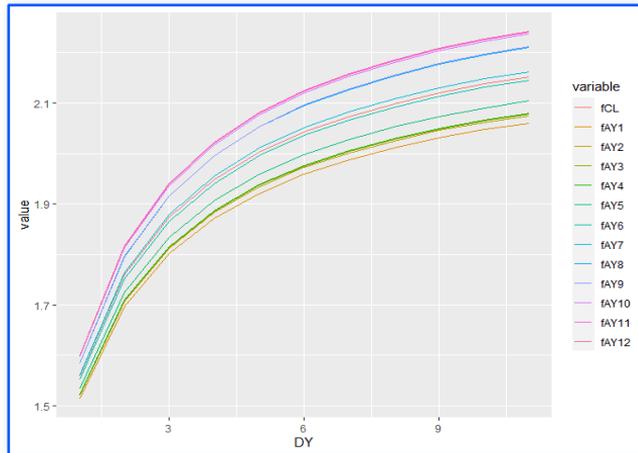
- ✓ Coefficients de passage individuels

$$\widehat{f_{i,j}^{\text{Model}}} = \frac{\sum_{l=0}^j \widehat{\mu_{i,l}^{\text{Model}}}}{\sum_{l=0}^{j-1} \widehat{\mu_{i,l}^{\text{Model}}}} = \frac{C_{i,j}}{C_{i,j-1}}$$

- ✓ Cumul des coefficients de passage individuels

$$fAY_{i,j} = \prod_{l=0}^j f_{i,l}$$

Cumul des coefficients de passage individuels par année de survivance calculés à partir des prédictions du modèle NNODP LoB1



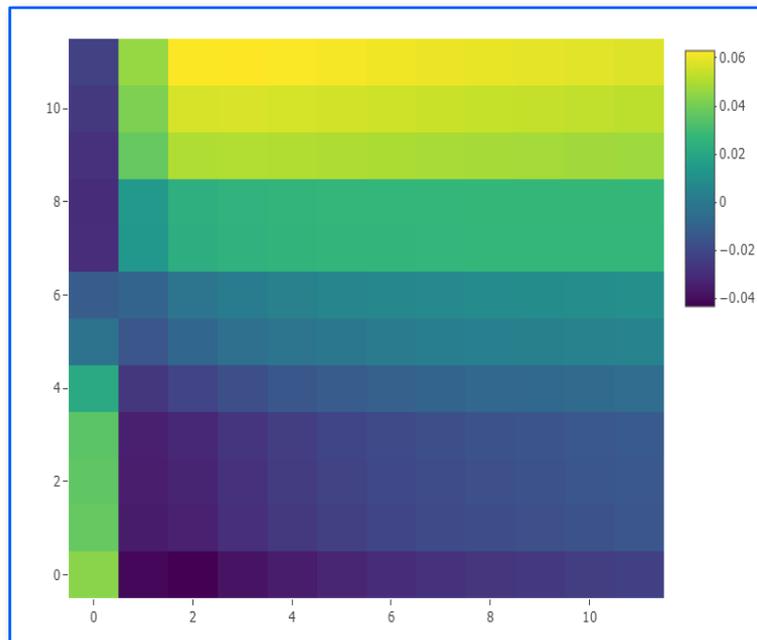
➤ Le modèle NNODP/NNODDP relâchent l'hypothèse de CL selon laquelle le développement ne dépend pas de l'année de survivance

- 3.9 • Ecart relatif de prédiction par rapport au modèle CL

- ✓ Différence relative entre les prédictions des modèles étudiés et les prédictions du modèle de CL

$$r_{i,j;m}^{Model} = \frac{\widehat{\mu}^{Model}(i,j;m)}{\widehat{\mu}^{CL}(i,j;m)} - 1$$

- ✓ Le modèle ajuste à la hausse les prédictions du GLM pour les années d'accident récentes et à la baisse pour les vieilles années d'accident
- ✓ L'ajustement est plus léger pour les années d'accident intermédiaire
- ✓ Pas d'évolution brusque de r , cela laisse penser qu'il n'y a pas de surapprentissage : le modèle a un pouvoir de généralisation



- 3.10 • Sur l'interprétabilité

✓ **Pourquoi faut-il pouvoir expliquer les modèles ?**

- RGPD : l'assuré a le « droit de ne pas faire l'objet de décisions entièrement automatisées »
- ACPR : l'assureur doit être capable de justifier ces décisions, et ne saurait transférer la responsabilité à la machine.
- Validation et communication au sein de l'entité

✓ **Qu'est ce que l'interprétabilité?**

- La capacité d'expliquer ou de présenter en termes compréhensibles à un humain
- Le degré avec lequel un observateur peut comprendre la cause d'une décision
- Une méthode est interprétable si un utilisateur peut prédire correctement et efficacement les résultats de la méthode

✓ **Quelles sont les techniques d'interprétation?**

- Techniques intrinsèques
- Techniques post-hoc

✓ **Quelques remarques sur l'interprétabilité**

- Tous les modèles linéaires ne sont pas interprétables
- Les mécanismes des algorithmes de machine learning sont connus

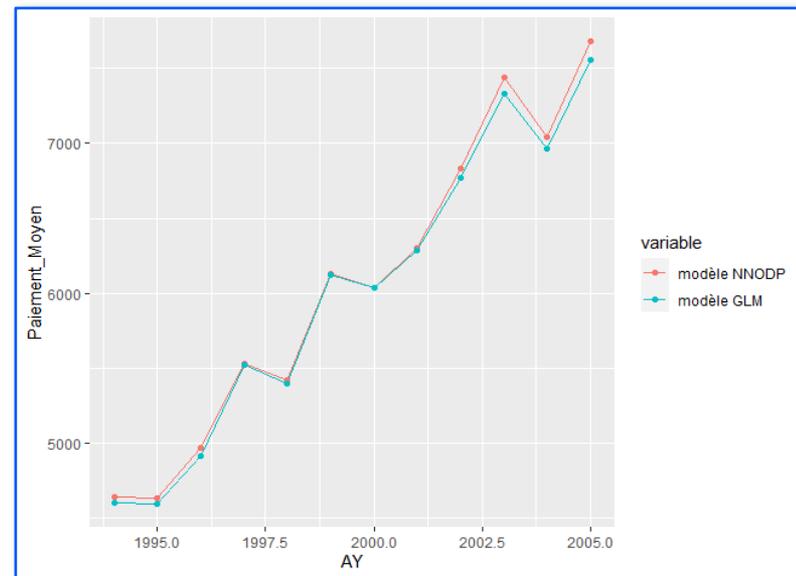
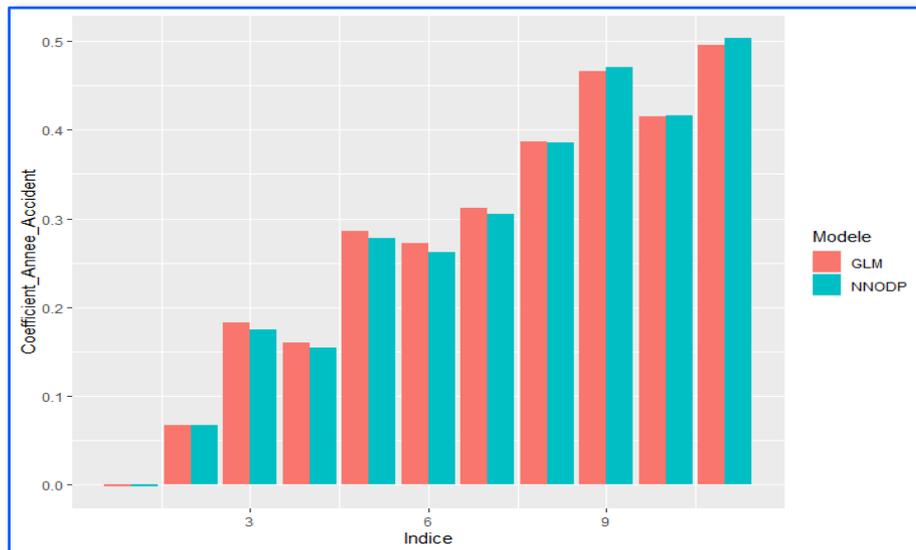
- 1 • Interprétation des modèles : graphe de dépendance partiel

➤ Fonction de dépendance partielle :

$$\checkmark h(z) = E_{X^{-j}}[\hat{f}(x|x^j = z)]$$

➤ Estimateur:

$$\checkmark \hat{h}(z) = \frac{1}{N} \sum_{i=1}^N \hat{f}(x_i | x_i^j = z)$$



Comparaison entre le graphe de dépendance partielle de l'année d'accident pour le modèle NNODP et le GLM

- ❑ Deux modèles , une même logique : intégrer un modèle de régression paramétrique dans un réseau de neurones grâce à plusieurs leviers :
 - L'architecture des réseaux
 - Le choix de la fonction d'activation de la couche de sortie
 - L'initialisation astucieuse des paramètres des réseaux

- ❑ Avantages qui en découlent :
 - Accélération du temps d'apprentissage
 - Possibilité de mettre en œuvre des procédures de bootstrap pour l'estimation des erreurs de prédiction
 - Interprétabilité des résultats

- ❑ Contraintes :
 - Complexité (implémentation x communication)
 - Besoin en données individuelles

- ❑ Des modèles généralisables :
 - Possibilité d'intégrer des variables supplémentaires (au-delà de l'Année d'accident et l'Année de développement)
 - Utilisation possible dans d'autres domaines (tarification par exemple)