

$$\Lambda_x = \sum_{t=1}^{\infty} \frac{1}{(1+i)^t} \mathbf{1}_{]t; \infty[} (T_x)$$

ressources-actuarielles.net



Prédiction de séries temporelles multivariées avec un réseau de neurones RVFL

Version 1.1

Mai 2024

Frédéric PLANCHET
frederic@planchet.net

- Cette présentation vient compléter le cours sur les GSE (voir [ce support](#)) en introduisant une approche non paramétrique basée sur l'utilisation de réseau de neurones ;
- La construction de scénarios économiques avec ce type d'outils a donné lieu à de récents travaux (FLAIG et JUNIKE [2022], RIZZATO et al. [2023]), qui utilisent le plus souvent des [GAN](#).
- Cette présentation a pour objectif de proposer une introduction simple à l'utilisation de réseaux de neurones pour la construction de scénarios économiques.
- Elle est issue de COUSIN et al. [2018], qui propose une modélisation semi-paramétrique de la projection de structures par termes de taux d'intérêt.
- Une approche plus générale est décrite dans la [présentation de L. AUBERT-LASSARADE](#).

- ■ En effet, dans le contexte de l'ORSA, il est nécessaire de disposer d'un modèle de déformation de la structure par termes des taux d'intérêts (cf. par exemple BONNIN et al. [2015]).
- ■ L'utilisation de la représentation proposée par Nelson et Siegel (NELSON et SIEGEL [1987]) fournit un cadre bien adapté pour cela et couramment utilisé, avec des techniques d'ajustement très variables (TURC et al. [2009], BOLDER [2001])
- ■ À partir d'historiques des STTI, on obtient une série temporelle de dimension 3 ou 4 qui doit être extrapolée, ce qui est effectué le plus souvent avec des techniques ARIMA ou VAR.
- ■ Cette étape est délicate du fait de
 - la forte corrélation entre les composantes de la série ;
 - la présence de phénomènes non linéaires.
- ■ On propose ici d'utiliser une classe particulière de réseaux de neurones pour prendre en compte efficacement ces deux aspects.

- On considère le modèle RVFL dans lequel la réponse Y , de dimension n , est expliquée par p prédicteurs $Z^{(j)}$ (chacun de dimension n) de la manière suivante :

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j Z_i^{(j)} + \sum_{l=1}^L \gamma_l g \left(\sum_{j=1}^p W^{(j,l)} Z_i^{(j)} \right) + \epsilon_i$$

- g est la fonction d'activation, L , le nombre de nœuds et $W^{(j,l)}$ des poids ; cette spécification fait donc intervenir une composante linéaire et une composante non linéaire basée sur les prédicteurs transformés

$$\Phi(\mathbf{Z})^{(l)} = g \left(\sum_{j=1}^p W^{(j,l)} Z_i^{(j)} \right)$$

- Les poids $W^{(j,l)}$ ne sont pas estimés mais choisis avec un générateur de nombres uniformes quasi-aléatoire (Sobol), l'idée étant qu'ils soient « bien uniformément répartis ».

- Les prédicteurs doivent être normalisés (centrés et réduits), ce qui peut être effectué de la manière suivante :
 - appliquer une normalisation à X , X_scaled et $scales_x$ (la matrice normalisée et le facteur d'échelle) ;
 - faire $g(X_scaled \% \% W)$ et obtenir la matrice Z de colonnes $Z = [X g(X_scaled \% \% W)]$;
 - appliquer une normalisation à Z , Z_scaled et $scales_z$ (la matrice normalisée et le facteur d'échelle) ;
 - faire la régression de y centrée (conserver la moyenne de y , y_m) sur Z_scaled ;
 - au moment d'obtenir les prédictions, il faut mettre les nouvelles observations X^* , à la même échelle que les observations X en appliquant $scales_x$ et $scales_z$
 - de même la prédiction obtenue doit être augmentée de y_m .
- La normalisation peut se faire de différentes manière (avec l'écart type, l'étendue, etc.)
- Le processus ci-dessus est complexe et très peu analysé dans la littérature (il n'est par exemple pas documenté dans l'article cité en référence et aucun des trois relecteurs n'a fait d'observation !)

- Les paramètres à estimer sont donc les β_j et γ_l , ainsi que le nombre de nœuds L .
- Le principe est de faire une régression linéaire sur les prédicteurs (initiaux et transformés), donc d'utiliser un critère de moindres carrés ; toutefois, pour éviter les problèmes de conditionnement liés à la colinéarité des variables explicatives (d'autant plus forte avec une série temporelle), on impose une pénalisation Ridge :

$$\sum_{j=1}^p \beta_j^2 \leq u \qquad \sum_{l=1}^L \gamma_l^2 \leq v$$

- La résolution des conditions du premier ordre conduit à une expression explicite des paramètres à partir de l'inversion de matrices, pour les coefficients (λ_1, λ_2) issus de u et v dans le Lagrangien donnés.
- On construit des intervalles de confiance de prédiction en utilisant un modèle ARIMA sur les résidus.

- Le nombre de nœuds L et les paramètres de pénalisation sont estimés par validation croisée de la manière suivante :
 - On fixe ces paramètres dans une grille, on estime le modèle et on construit la prédiction depuis l'origine ;
 - On se décale dans le temps d'une unité et on recommence ;
 - On retient le triplet avec la plus petite erreur quadratique (cf. [ce package](#)).

I. Kaastra, M. Boyd / Neurocomputing 10 (1996) 215–236

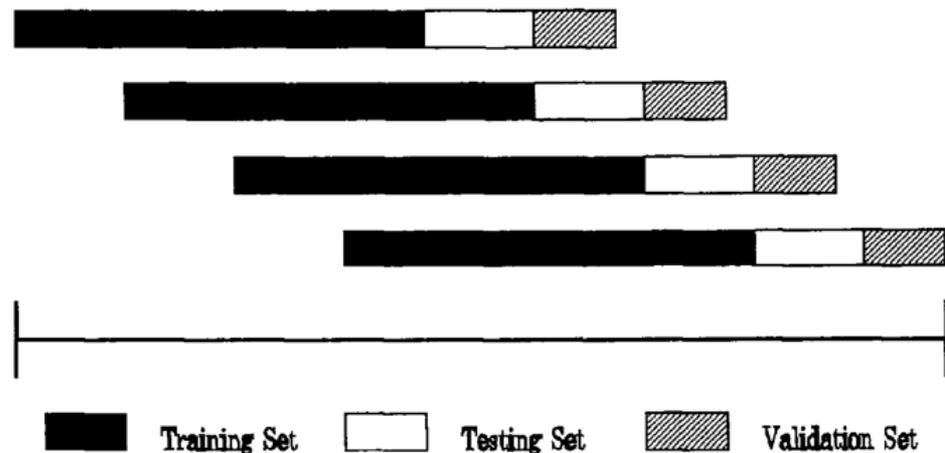


Fig. 3. Walk-forward sliding windows testing routine.

Utilisation avec des séries temporelles

- On dispose de p séries temporelles observées à n dates, on souhaite construire une prédiction de ces p séries à h dates $n+1, \dots, n+h$. On s'appuie sur un historique de $k < n$ valeurs de la série.

- La matrice des prédicteurs est alors constituée $n-k$ lignes de p bloc de k valeurs retardées de la forme

$$\left(X_{n-i}^{(j_0)}, \dots, X_{k+1-i}^{(j_0)} \right)^T$$

- La matrice des prédicteurs transformés est de taille $(n-k) \times (k \times p + L)$

- Avec $p=2$ séries observées à $n=5$ dates, $k=2$ valeurs retardées et $L=3$ nœuds on a :

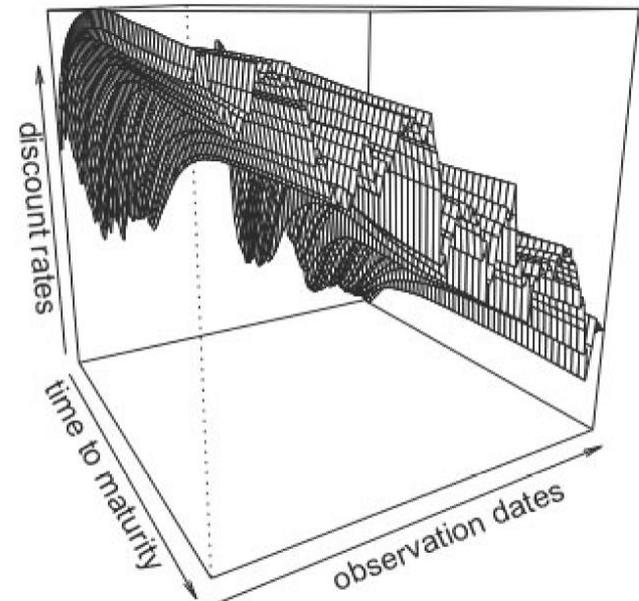
$$Y = \begin{pmatrix} X_{t_5}^{(1)} & X_{t_5}^{(2)} \\ X_{t_4}^{(1)} & X_{t_4}^{(2)} \\ X_{t_3}^{(1)} & X_{t_3}^{(2)} \end{pmatrix} \quad X = \begin{pmatrix} X_{t_4}^{(1)} & X_{t_3}^{(1)} & X_{t_4}^{(2)} & X_{t_3}^{(2)} \\ X_{t_3}^{(1)} & X_{t_2}^{(1)} & X_{t_3}^{(2)} & X_{t_2}^{(2)} \\ X_{t_2}^{(1)} & X_{t_1}^{(1)} & X_{t_2}^{(2)} & X_{t_1}^{(2)} \end{pmatrix} \quad W = \begin{pmatrix} W^{(1,1)} & W^{(1,2)} & W^{(1,3)} \\ W^{(2,1)} & W^{(2,2)} & W^{(2,3)} \\ W^{(3,1)} & W^{(3,2)} & W^{(3,3)} \\ W^{(4,1)} & W^{(4,2)} & W^{(4,3)} \end{pmatrix}$$

Application numérique

- On utilise des données de taux d'intérêt de la Deutsche Bundesbank (http://www.bundesbank.de/Navigation/EN/Statistics/Time_series_databases/time_series_databases.html), observées mensuellement, de 2002 à 2015. On dispose ainsi de 167 courbes avec 50 maturités.
- Un ajustement NS est effectué pour chaque courbe, ce qui conduit à une série temporelle de dimension 3 (le paramètre d'échelle du modèle est estimé par validation croisée).

| Maturity | Min | 1st Qrt | Median | 3rd Qrt | Max |
|----------|--------|---------|--------|---------|-------|
| 1 | -0.116 | 0.858 | 2.045 | 3.072 | 5.356 |
| 5 | 0.170 | 1.327 | 2.863 | 3.807 | 5.146 |
| 15 | 0.711 | 2.616 | 3.954 | 4.702 | 5.758 |
| 30 | 0.805 | 2.594 | 3.962 | 4.814 | 5.784 |
| 50 | 0.749 | 2.647 | 3.630 | 4.590 | 5.467 |

$$R_t(\tau) = \alpha_{1,t} + \alpha_{2,t} \left(\frac{1 - e^{-\tau/\lambda}}{e^{-\tau/\lambda}} \right) + \alpha_{3,t} \left(\frac{1 - e^{-\tau/\lambda}}{e^{-\tau/\lambda}} - e^{-\tau/\lambda} \right)$$



Application numérique

- Les résultats obtenus sur un horizon de projection de 12 mois, comparés avec des approches ARIMA et VAR, sont les suivants :

Table 2. Comparison of 12 months ahead out-of-sample $RMSE$ ($\log(\text{Root Mean Squared Error})$), for the ARIMA(AutoRegressive Integrated Moving Average), RVFL, and VAR (Vector AutoRegressive).

| Method | Min. | 1st Qrt | Median | Mean | 3rd Qrt | Max. |
|--------|---------------|---------------|---------------|---------------|---------------|---------------|
| RVFL | 0.1487 | 0.3092 | 0.4491 | 0.5041 | 0.6414 | 1.1535 |
| ARIMA | 0.2089 | 0.3470 | 0.5187 | 0.6358 | 0.7516 | 5.3798 |
| VAR | 0.1402 | 0.3493 | 0.5549 | 1.9522 | 0.8619 | 122.2214 |

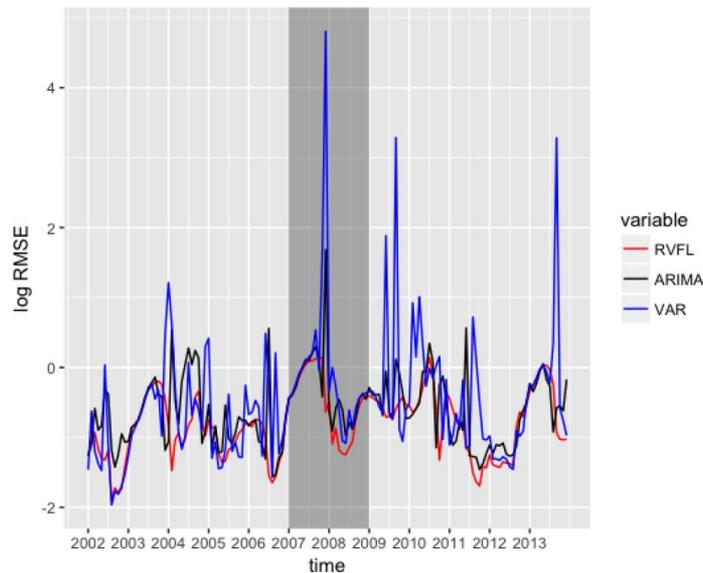
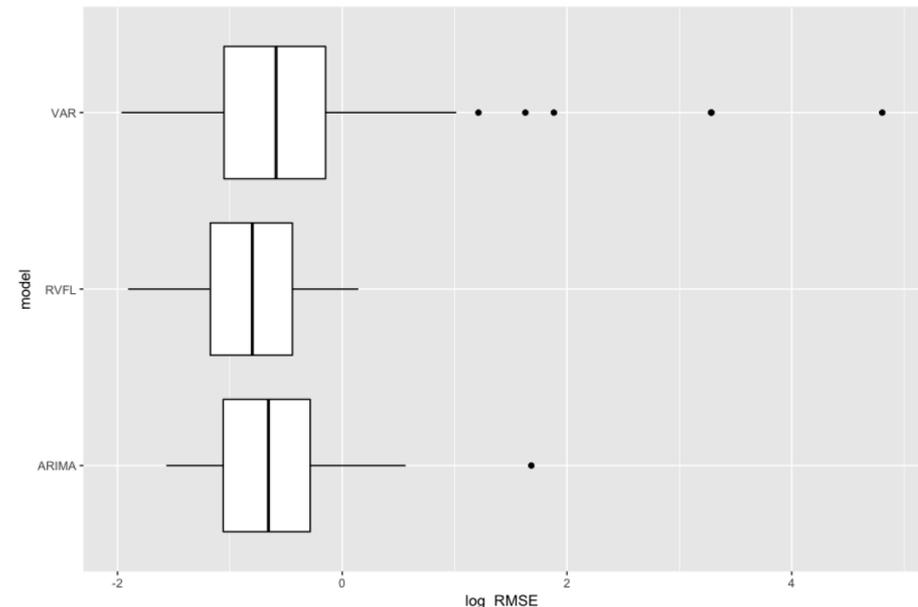


Figure 4. Out-of-sample $\log(RMSE)$, for ARIMA, VAR, and RVFL over time.



(12 mois pour l'estimation, 12 mois pour la projection, L est de l'ordre de 10 ; avec des horizons de projection et d'estimation plus longs, 36 mois, L est de l'ordre de 40)

Conclusion

- Les modèles de type RVFL fournissent un outil performant pour réaliser des projections de séries temporelles multivariées avec une forte dépendance entre les composantes.
- La propriété la plus intéressante de ce modèle est de diminuer la variance de la prédiction sans dégrader le biais.
- Par ailleurs, des variables explicatives supplémentaires peuvent aisément être intégrées et la prise en compte de stress sur les prédicteurs est immédiate.
- Le principe du modèle est simple, mais sa mise en œuvre nécessite un peu d'attention, notamment pour la normalisation des prédicteurs et le choix des méta paramètres choisis par validation croisée.

Références

- BOLDER D.J. [2001]** Affine Term-Structure Models: Theory and Implementation, Banque du Canada, Document de travail 2001-15.
- BONNIN F., COMBES F., PLANCHET F., TAMMAR M. [2015]** Un modèle de projection pour des contrats de retraite dans le cadre de l'ORSA, *Bulletin Français d'Actuariat*, vol. 14, n°28.
- COUSIN A., MOUDIKI T., PLANCHET F. [2018]** Multiple time series forecasting using quasi-randomized functional link neural networks, *Risks*, 2018, 6(1), 22; doi:10.3390/risks6010022.
- FLAIG S.; JUNIKE G. [2022]** Scenario Generation for Market Risk Models Using Generative Neural Networks, *Risks* 10: 199. DOI: <https://doi.org/10.3390/risks10110199>
- NELSON C.R., SIEGEL A.F. [1987]** « Parsimonious modelling of yield curves », *Journal of Business*, 60, 473-489.
- RIZZATO M., WALLART J., GEISSLER C., MORIZET N., BOUMLAIK N. [2023]** Generative Adversarial Networks applied to synthetic financial scenarios generation, *Physica A: Statistical Mechanics and its Applications*, Volume 623, 1, 128899.
- TURC J. , UNGARL S., KURPEL A., HUANG C. [2009]** Filtering the interest rate curve : the MENIR framework, Société Générale, Fixed Income.