



Mémoire présenté devant le jury de l'EURIA en vue de l'obtention du
Diplôme d'Actuaire EURIA
et de l'admission à l'Institut des Actuaires

le 18 Septembre 2025

Par : Benoît PARPAY

Titre : **Impacts d'une modélisation segmentée des actifs immobiliers sur la tarification d'un portefeuille de Prêts Viagers Hypothécaires**

Confidentialité : Non

*Membres présents du jury de l'Institut
des Actuaires :*

Romain LAÏLY

Signature :

Anthony NAHELOU

Signature :

Membre présent du jury de l'EURIA :
Rainer BUCKDAHN

Entreprise :

Deloitte France

Signature :

*Directeur de mémoire en
entreprise :*

David CRESPIE

Signature :

*Autorisation de publication et de mise en ligne sur un site de diffusion
de documents actuariels*

(après expiration de l'éventuel délai de confidentialité)

Signature du responsable entreprise :

Signature du candidat :

Résumé

Le vieillissement de la population, conjugué à la diminution du niveau de vie des retraités, accentue la pression sur le système de retraite français et sur la situation économique des seniors. Le prêt viager hypothécaire - un crédit viager garanti par un bien immobilier dont le remboursement est plafonné par la valeur du bien - apparaît comme une réponse possible à cet enjeu. Ce produit offre aux seniors propriétaires la possibilité de mobiliser une partie du capital immobilisé dans leur logement, sans avoir à le vendre ni à le quitter. Ce type de prêt est exposé à plusieurs risques majeurs, notamment l'incertitude liée à la longévité et à l'évolution de la valeur du bien immobilier. Ce mémoire analyse l'impact d'une modélisation segmentée des trajectoires de prix immobiliers sur la tarification du prêt viager hypothécaire, du point de vue du prêteur.

Pour ce faire, un ensemble de scénarios économiques est simulé et alimente un moteur de tarification qui détermine, pour chaque individu, le montant maximal pouvant être prêté en fonction de ses caractéristiques et de celles de son bien, tout en respectant les contraintes de rentabilité et de gestion du risque fixées.

Concrètement, une fois l'environnement central établi, reposant sur la simulation des rendements immobiliers nationaux via un générateur de scénarios économiques de type Ahlgrim simplifié, deux méthodes de modélisation segmentée sont construites. Ce type d'approche requiert la construction d'indices de prix à une maille précise, réalisés ici à partir d'un modèle hybride combinant une régression hédonique et un modèle à ventes répétées. Ces indices sont ensuite utilisés pour modéliser les rendements immobiliers segmentés par un ensemble de GSEs de type Ahlgrim simplifiés. La seconde approche explorée s'appuie sur un réseau de neurones récurrent de type LSTM (*Long Short-Term Memory*), permettant de capturer les dépendances temporelles complexes dans l'évolution des prix immobiliers.

Mots clefs: Prêt viager hypothécaire, rendement immobilier, générateur de scénarios économiques, Ahlgrim, régression hédonique, ventes répétées, réseau de neurones récurrent, *Long Short-Term Memory*, rentabilité, gestion de risque

Abstract

The ageing of the population, combined with the decline in retirees' standard of living, increases the pressure on the French pension system and on the economic situation of seniors. The reverse mortgage – a lifetime loan secured by a property, whose repayment is capped at the value of the asset – appears as a possible response to this challenge. This product allows senior homeowners to mobilise part of the capital tied up in their housing, without having to sell or leave it. Such a loan, however, is exposed to several major risks, notably the uncertainty related to longevity and to the evolution of property values. This thesis analyses the impact of segmented modelling of real estate price trajectories on the pricing of reverse mortgages, from the lender's perspective.

To this end, a set of economic scenarios is simulated and feeds into a pricing engine that determines, for each individual, the maximum amount that can be lent depending on their characteristics and those of their property, while respecting predefined profitability and risk management constraints.

Concretely, once the central environment is established – based on the simulation of national real estate returns using a simplified Ahlgrim-type economic scenario generator – two segmented modelling approaches are developed. This type of approach requires the construction of fine-grained price indices, built here from a hybrid model combining a hedonic regression and a repeat-sales model. These indices are then used to model segmented real estate returns through a set of simplified Ahlgrim-type ESGs. The second approach explored relies on a recurrent neural network of the LSTM type (*Long Short-Term Memory*), designed to capture complex temporal dependencies in the evolution of real estate prices.

Mots clefs: Equity Release Mortgage, real estate return, economic scenario generator, Ahlgrim, hedonic regression, repeat sales, recurrent neural network, Long Short-Term Memory, profitability, risk management

Remerciements

Je remercie sincèrement David CRESPIE, mon tuteur de mémoire, pour ses conseils avisés et le temps qu'il m'a consacré tout au long de ce travail. Son accompagnement a été essentiel à la bonne réalisation de ce mémoire. Je remercie également Claude-Florence CHASSAIN, Baptiste BRECHOT et Cyril CHALIN pour la confiance qu'ils m'ont accordée tout au long de mon alternance.

Je souhaite également remercier Fabrice HAMON, pour son encadrement académique ainsi qu'Arnaud MONNIN pour son temps et sa vision marché.

Je remercie chaleureusement ma famille et amis proches pour leur soutien sans faille tout au long de la rédaction de ce mémoire.

Enfin, je remercie toute l'équipe Deloitte Actuariat Assurance pour leur bonne humeur et leur soutien, et particulièrement Erwan-Henri BURLISSON pour ses remarques et relectures.

Note de synthèse

Le vieillissement démographique en France, conjugué à la diminution du nombre d'actifs par rapport aux retraités, génère une tension de plus en plus forte sur le système de retraite en France, basé sur la solidarité intergénérationnelle. Cette tension provoque une diminution progressive des revenus des seniors lors de leur passage de la vie active à la retraite. En palliant cette baisse de revenus sans compromettre leur qualité de vie, le prêt viager hypothécaire (PVH) apparaît comme une solution de plus en plus à même de couvrir les besoins financiers des seniors. En effet, ce type de prêt leur permet de mobiliser une partie de la valeur de leur bien immobilier, tout en conservant le droit d'y résider.

Le PVH présente cependant plusieurs risques majeurs pour le prêteur, en particulier le risque de baisse de valeur du bien en hypothèque. Ce risque est généralement modélisé à l'aide de processus stochastiques, par exemple via un modèle de type Black-Scholes pour les prix ou Vasicek pour les rendements. Ces approches sont toutefois souvent calibrées sur un unique indice national et appliquées uniformément à l'ensemble du marché. Or, dans un cadre de tarification ou de gestion du risque, une telle démarche tend à effacer les disparités locales et peut ainsi éloigner la modélisation du risque de son profil réel. En réalité, l'évolution du prix d'un bien immobilier dépend fortement de sa typologie et de sa localisation.

L'objectif principal de ce mémoire est de mesurer les impacts d'une modélisation segmentée des trajectoires de prix immobiliers sur la tarification du PVH, du point de vue du prêteur. Pour ce faire, après avoir constitué un portefeuille de prêts, en s'appuyant sur les statistiques nationales INSEE puis en les retraitant afin qu'elles correspondent au mieux aux caractéristiques spécifiques des PVH, un outil de tarification ainsi que trois modèles de projection immobilière ont été construits. En fin de modélisation, plusieurs indicateurs sont comparés entre la méthode fondée sur des rendements agrégés et celles basées sur des rendements segmentés, tels que les niveaux de rentabilité ou le ratio *Loan-To-Value* (LTV), c'est-à-dire la valeur prêtée exprimée en pourcentage de la valeur du bien.

Outil de tarification

Le modèle de tarification développé permet de calculer une LTV optimale sous contraintes d'une rentabilité définie et d'un niveau de risque donné. Ce modèle simule N scénarios de mortalité et utilise en entrée N scénarios économiques (taux court, taux long, inflation et rendement immobilier), à partir desquels il calcule plusieurs indicateurs de rendement et de gestion du risque. La LTV optimale calculée traduit le montant de prêt maximal qui respecte les contraintes de performance et de gestion du risque définies.

À partir de N trajectoires jointes de mortalité et de variables économiques, la LTV optimale LTV^* est définie comme la solution du schéma de maximisation sous contraintes suivant :

$$\left\{ \begin{array}{l} \text{(i) } \mathbb{E}[R_{\text{ann}}] \geq R_{\text{cible}}, \\ \text{(ii) } \text{VaR}_{\alpha}(R_{\text{ann}}) \geq r_{\text{min}}, \\ \text{(iii) } \mathbb{P}(\text{CRD}_T > V_T) \leq \beta \quad (\iff \text{VaR}_{\beta}(V_T - \text{CRD}_T) > 0), \\ \text{(iv) } \mathbb{E}\left[\frac{\text{NNEG}}{C_0}\right] \leq \gamma. \end{array} \right.$$

avec $LTV \in \{5\%, 10\%, \dots, 90\%, 95\%\}$, où :

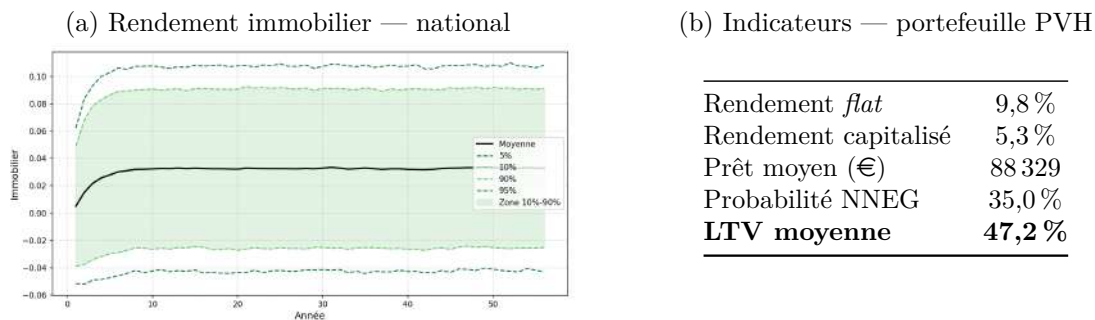
- R_{ann} représente le rendement annuel *flat*, R_{cible} le rendement minimal cible, fixé à 6,5%,

- r_{\min} le rendement minimal cible pour les $\alpha=5\%$ scénarios les plus adverses, fixé à 2,5%,
- CRD_T le capital restant dû à la date de clôture du prêt (sans garantie de non-dépassement), V_T la valeur du bien à T et β la probabilité maximale de déclenchement de la garantie de non-dépassement (*Non Negative Equity Guarantee*), fixée à 45%,
- NNEG la valeur de la garantie de non-dépassement, C_0 le capital prêté en t_0 et γ la valeur maximale de NNEG acceptée, exprimée en % du capital prêté, fixée à 8,5%.

Chaque composante est évaluée scénario par scénario puis agrégée, permettant de rendre compte de la performance attendue et des situations extrêmes au travers de moyennes et quantiles. Les LTV optimales présentées par la suite sont très sensibles aux types de contraintes et aux seuils définis ici.

Modélisation classique des rendements immobiliers

Le modèle central repose sur un processus de Vasicek calibré sur les log-rendements immobiliers nationaux, calculés à partir des données de l'INSEE, conformément au cadre du GSE d'Ahlgrim en monde réel. Les scénarios économiques ainsi générés sont ensuite intégrés dans l'outil de tarification. Les scénarios obtenus sont validés par une étude des moments, une comparaison avec l'historique utilisé pour le paramétrage des modèles et une étude de cohérence.



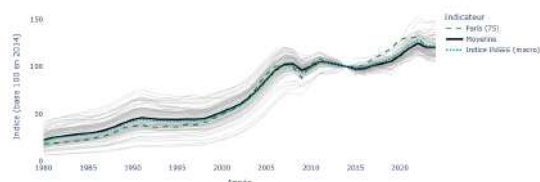
Dans ce cadre, la LTV moyenne s'établit à 47,2% ; les disparités individuelles proviennent des caractéristiques influençant la longévité des emprunteurs.

Modélisation segmentée des rendements immobiliers

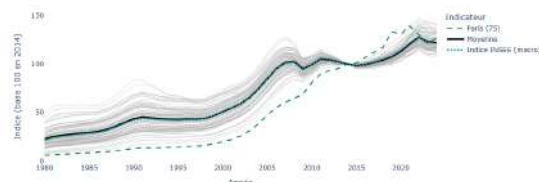
Pour tenir compte des hétérogénéités spatiales et typologiques masquées par l'agrégation nationale, les projections de rendements sont menées à la maille département \times type de bien, soit 182 segments. Deux méthodes sont utilisées : la Méthode 1, fondée sur des GSEs segmentés, et la Méthode 2, basée sur un réseau de neurones récurrent de type LSTM, avec encodage des segments.

Pour mettre en œuvre ces deux méthodes, des données immobilières détaillées sont nécessaires. Cependant, les rares indices immobiliers publics segmentés sont peu profonds et ne couvrent pas l'ensemble du territoire français, tandis que les données publiques disponibles à ce niveau de granularité restent limitées, avec un historique d'une dizaine d'années. Un modèle hybride combinant régression hédonique et ventes répétées a été mis en œuvre et permet de construire des indices à la maille souhaitée, à partir de la base de données DVF (2014-2024). Ces indices ont ensuite été prolongés vers le passé afin d'étendre la profondeur d'historique disponible. Cette projection rétroactive est facilitée par la disponibilité des variables macroéconomiques observées, contrairement à la prévision future qui repose sur des scénarios simulés. Un retraitement a été appliqué pour garantir la cohérence agrégée : la moyenne des

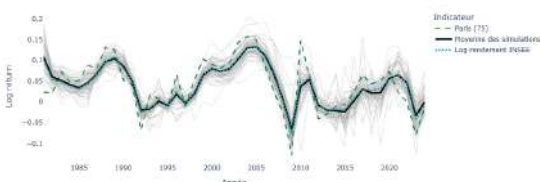
indices reconstruits a été ajustée pour correspondre à l'indice immobilier global publié par l'INSEE, et les valeurs extrêmes ont été corrigées. L'application de cette méthode constitue l'une des limites de l'étude, dans la mesure où elle introduit un biais. Un regard critique est donc nécessaire quant à l'interprétation des résultats obtenus. Les indices et rendements segmentés, alignés avec l'indice INSEE national, présentent les dynamiques suivantes :



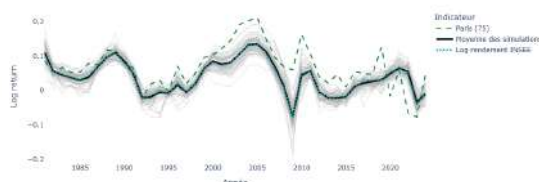
(a) Indices - Maisons



(b) Indices - Appartements



(c) Log-rendements - Maisons

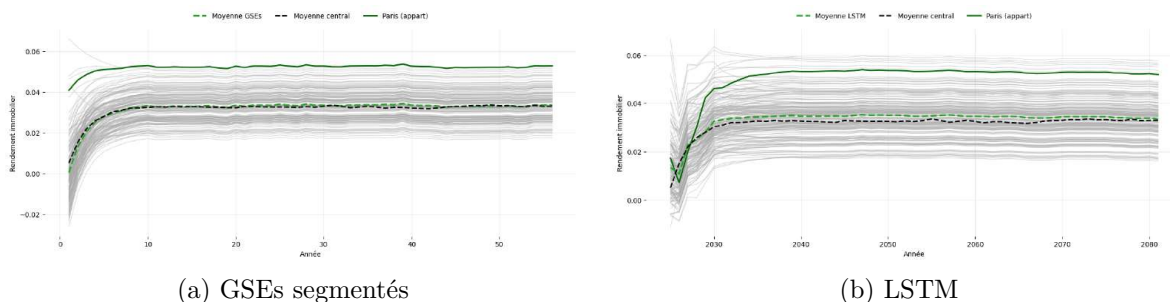


(d) Log-rendements - Appartements

Ces rendements ont ensuite été utilisés pour les deux méthodes décrites ci-après :

Méthode 1 - Générateurs de Scénarios Economiques segmentés : Pour assurer une modélisation segmentée stable et robuste, plusieurs générateurs de scénarios distincts ont été développés, chacun dédié à un segment et intégrant le même ensemble de quatre variables clés : taux court, taux long, inflation et rendement immobilier segmenté. Afin de préserver la cohérence inter-segments, les résidus aléatoires associés aux variables économiques hors immobilier sont partagés entre tous les générateurs. Une même réalisation de l'environnement économique s'applique donc simultanément à l'ensemble des segments, ce qui introduit une corrélation implicite et garantit que les écarts observés dans les trajectoires simulées reflètent uniquement les spécificités immobilières de chaque segment. Cette structure permet de capter l'hétérogénéité des dynamiques immobilières tout en considérant un environnement économique commun au niveau national. Ainsi, cette méthode est très proche de la méthode agrégée, dite centrale, puisqu'elle repose sur les mêmes principes théoriques mais sur un paramétrage plus fin.

Méthode 2 - Réseau de neurones récurrent : Pour capturer les dépendances temporelles complexes et non linéaires dans l'évolution des rendements immobiliers départementaux, un réseau de neurones récurrent de type *Long Short-Term Memory* (LSTM) a été mis en œuvre et utilisé de manière récursive. Ce modèle intègre en entrée les variables économiques issues du GSE central (taux court, taux long, inflation, rendement immobilier national) ainsi que des vecteurs d'*embedding* représentant les segments géographiques et typologiques. Les prédictions générées à chaque période sont réinjectées dans le modèle pour alimenter la période suivante, ce qui permet de projeter des trajectoires sur un horizon long. Cette architecture assure la cohérence avec les tendances observées dans l'historique tout en capturant l'hétérogénéité entre départements et types de biens, offrant ainsi un cadre flexible pour simuler l'impact des scénarios économiques sur les dynamiques locales. Les scénarios de taux et d'inflation générés en central sont utilisés en parallèle des scénarios de rendements immobiliers produits par LSTM.



Les résultats obtenus avec les GSEs segmentés et ceux issus du LSTM sont comparables, en moyenne, entre eux ainsi qu'avec le rendement moyen en environnement central mais présentent des écarts intra-segment. En effet, la combinaison d'une variabilité moindre en début de projection (dûe à la structure même du LSTM) et une variabilité supérieure une fois le régime établi (autour de 20 ans) aboutit à une hausse de la valeur NNEG espérée en LSTM *vs.* scénarios fondés sur les GSEs. Les simulations produites sont intégrées dans l'outil de tarification, permettant d'évaluer précisément leur impact dans le cadre du PVH. Sous l'effet combiné des intérêts composés, de la pondération inégale des départements dans le portefeuille et de l'influence de la longévité individuelle sur les indicateurs étudiés à l'échelle du portefeuille, ces différences de dispersion et de trajectoires impactent la tarification des prêts.

Résultats et analyses

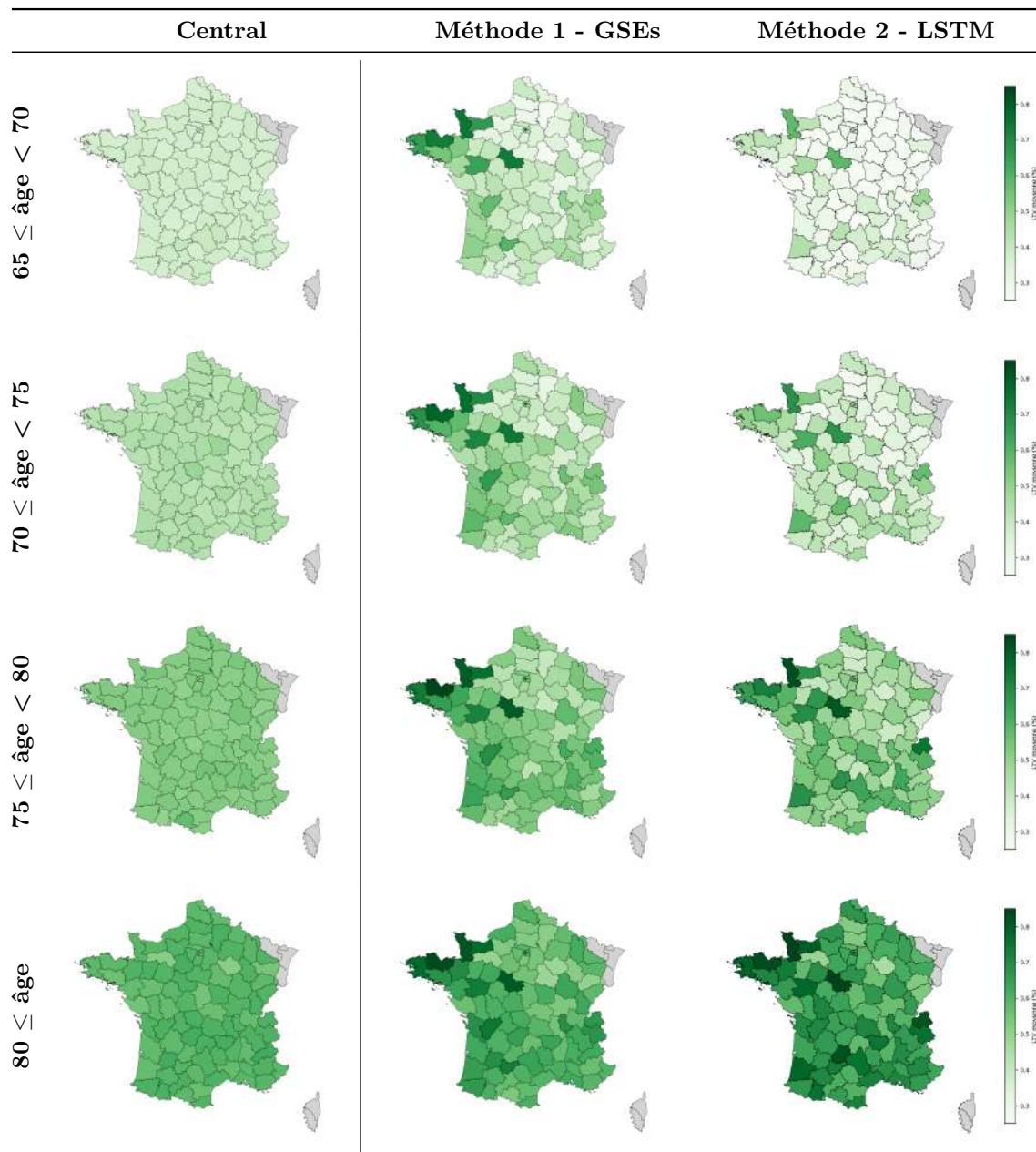
Les trois approches sont basées sur le même modèle de tarification - seuls les scénarios de rendements utilisés lors de la projection des flux diffèrent - et présentent des métriques finales relativement proches à l'échelle du portefeuille. Cependant, les contraintes effectivement bloquantes ((i)-(iv) ci-dessous) ne sont pas les mêmes selon les profils de risque, et leur hiérarchie varie fortement d'une méthode à l'autre :

Indicateurs	Central	GSEs segmentés	LSTM
Rendement <i>flat</i>	9,8 %	9,9 %	9,8 %
Rendement capitalisé	5,3 %	5,3 %	5,4 %
Prêt moyen (€)	88 329	97 223	87 171
Probabilité NNEG	35,0 %	38,0 %	28,4 %
LTV moyenne	47,2 %	52,0 %	46,6 %
(i) Proportion du portefeuille contrainte par le rdt. <i>flat</i>	13,2 %	15,2 %	13,9 %
(ii) Prop. du ptf. contrainte par le rdt. <i>flat</i> au 5 ^{ème} percentile	6,6 %	1,3 %	–
(iii) Prop. du ptf. contrainte par la proba. NNEG	46,3 %	80,2 %	5,9 %
(iv) Prop. du ptf. contrainte par la valeur NNEG	33,5 %	2,9 %	79,8 %
Prop. du ptf. refusée	0,4 %	0,4 %	0,4 %
Proportion du portefeuille - Total	100 %	100 %	100 %

Ainsi, la LTV moyenne évolue peu ; en revanche, des écarts significatifs apparaissent lorsque la maille d'étude s'affine. La cartographie des LTV par département et âge confirme le poids de ces deux facteurs dans le calcul de LTV optimale. Toutefois, certains niveaux de LTV paraissent élevés, notamment en Bretagne et à Paris, et font l'objet de réserves et d'analyses complémentaires.

L'intégration de scénarios immobiliers segmentés dans la tarification modifie nettement la distribution départementale des LTV, soulignant l'intérêt d'une modélisation segmentée pour une tarification cohérente sur cet axe. Concrètement, au-delà des divergences départementales, une tendance à accentuer la différenciation déjà présente en central se dégage : baisse des LTV déjà faibles chez les plus jeunes et hausse des LTV déjà élevées chez les plus

âgés. Ce « tirage des extrémités » est d'autant plus marqué pour la méthode par LSTM et s'explique par le caractère asymétrique du *payoff* du PVH, capté par la contrainte de valeur NNEG au sein du modèle de tarification développé. Cette sensibilité de la LTV aux hypothèses de rendement et de longévité des cohortes est déjà documentée dans la littérature [A-W. Shao et al., 2014].



Tarification prudente

Au delà des tarifications des modèles seuls, une tarification prudente est mise en place en combinant les LTV obtenues par les différents modèles segmentés. Deux types d'ajustements ont été étudiés, avant calibrage puis en sortie de modèle, pour réduire des LTV considérées trop élevées, en particulier sur Paris et la Bretagne. Ces ajustements visent à tempérer des rendements projetés élevés, eux-mêmes hérités d'un historique caractérisé par de fortes hausses de prix. Sous cette approche prudente, la LTV retenue correspond systématiquement

à la valeur la plus faible parmi celles estimées par les différents modèles. Ce choix, fondé sur une analyse critique des premiers résultats, vise à garantir une cohérence durable entre les montants prêtés et les dynamiques immobilières locales, en réduisant le risque d'exposition excessive.

Cette approche conduit à une LTV moyenne de 43%, en cohérence avec les conditions de marché, mais sensiblement plus basse que celle du modèle central. Les rendements et les indicateurs de risque demeurent globalement stables malgré ce gain de prudence ; seul le volume de prêt est impacté. Bien que cette approche présente une marge de prudence, un cinquième des emprunteurs voient le montant de prêt qui leur est proposé augmenter par rapport au central (pouvant atteindre +20%). Ce résultat illustre que la modélisation de scénarios de rendements immobiliers segmentés offre une représentation plus fidèle du risque réel et accroît l'attractivité pour certains profils, en comparaison avec une approche homogène sur l'ensemble du marché. Enfin, une analyse cartographique par département complétée par des indicateurs agrégés, confirme la persistance de disparités territoriales, qui constitue précisément l'apport majeur de cette approche segmentée. Malgré la prudence adoptée, la segmentation de la modélisation immobilière demeure manifeste et constitue un levier essentiel pour une gestion plus fine du risque.

Limites et perspectives

L'ambition initiale de cette étude était de développer une modélisation individuelle des projections immobilières, directement liée aux caractéristiques fines du bien (localisation précise, qualité du logement, état, etc) et du profil de l'emprunteur. Une telle approche aurait permis d'obtenir des projections personnalisées et de rapprocher au plus près la dynamique immobilière du risque réel. Toutefois, la disponibilité limitée des données a conduit à recentrer cet objectif. Le choix a été fait de privilégier une modélisation segmentée, fondée sur des regroupements par département et type de bien. Cette approche, plus réaliste dans le cadre d'un mémoire basé sur des données *open-source*, permet de conserver une capacité explicative et prospective robuste, tout en ouvrant la voie à des développements ultérieurs plus fins, basés sur des données individuelles plus riches.

En effet, les principales limites de l'étude tiennent à la disponibilité et à la granularité des données immobilières, qui conditionnent la qualité des indices construits. Des travaux futurs pourraient s'appuyer sur un historique plus profond ou sur des informations plus détaillées relatives aux biens sous-jacents, afin de renforcer la pertinence de l'approche. Il serait également intéressant d'adopter le point de vue de l'emprunteur, afin de mieux comprendre ses attentes, d'identifier les contraintes auxquelles il fait face et, in fine, de proposer des solutions plus attractives.

Ce mémoire met en évidence qu'une modélisation segmentée des rendements immobiliers constitue un levier pour affiner la tarification des PVH. Toutefois, d'autres axes d'amélioration pourraient également renforcer la solidité et l'attractivité du produit, comme le développement de nouvelles modalités de tirage ou l'élargissement des types de prêts considérés, qui influencent eux aussi directement la tarification.

Synthesis

The demographic ageing in France, combined with the decrease in the number of working people relative to retirees, is generating increasing pressure on the French pension system, which is based on intergenerational solidarity. This pressure leads to a gradual reduction in seniors' income when they move from working life to retirement. By compensating for this loss of income without compromising their quality of life, the reverse mortgage (PVH) appears as an increasingly relevant solution to cover seniors' financial needs. Indeed, this type of loan allows them to mobilize part of the value of their real estate while retaining the right to live in it.

However, the reverse mortgage carries several major risks, among which the risk of real estate depreciation is predominant. This risk is generally modeled using stochastic processes, for instance through a Black-Scholes model for prices or a Vasicek model for returns. These approaches are often calibrated on a single national index and uniformly applied to the whole market. Yet, in the context of pricing or risk management, such a procedure tends to smooth out local disparities and may thus move the risk modeling away from its actual profile. In reality, the evolution of a property's price depends strongly on its type and its location.

After building a loan portfolio, based on INSEE data and adjusted to best match the specific characteristics of reverse mortgages, a pricing tool and three real estate projection models were developed. The main objective of this thesis is to assess the impacts of a finely segmented modeling of real estate returns on the pricing of reverse mortgages, from the lender's perspective. To do so, several indicators are compared between the method based on aggregated returns and those based on segmented returns, such as profitability levels or the Loan-To-Value (LTV) ratio, i.e. the loan amount expressed as a percentage of the property value.

Pricing tool

The pricing model developed makes it possible to compute an optimal Loan-To-Value (LTV) ratio under a given profitability target and a specified level of risk. This model simulates N mortality scenarios and takes as input N economic scenarios (short rate, long rate, inflation, and real estate return), from which it computes several indicators of profitability and risk management. The optimal LTV therefore reflects the maximum loan amount accepted by the lender while satisfying the predefined profitability and risk management constraints. Based on N joint trajectories of mortality and economic environment, the optimal LTV LTV^* is defined as the solution to the following maximization scheme under constraints :

$$\left\{ \begin{array}{l} \text{(i) } \mathbb{E}[R_{\text{ann}}] \geq R_{\text{target}}, \\ \text{(ii) } \text{VaR}_{\alpha}(R_{\text{ann}}) \geq r_{\text{min}}, \\ \text{(iii) } \mathbb{P}(\text{CRD}_T > V_T) \leq \beta \quad (\iff \text{VaR}_{\beta}(V_T - \text{CRD}_T) > 0), \\ \text{(iv) } \mathbb{E}\left[\frac{\text{NNEG}}{C_0}\right] \leq \gamma. \end{array} \right.$$

with $LTV \in \{5\%, 10\%, \dots, 90\%, 95\%\}$, where :

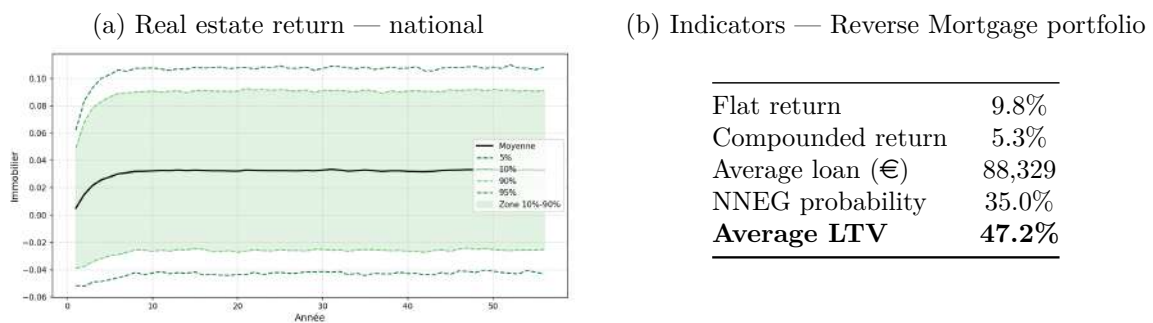
- R_{ann} denotes the annual flat return, R_{target} the minimum target return, set at 6.5%,
- r_{min} the minimum return required for the $\alpha = 5\%$ most adverse scenarios, set at 2.5%,

- CRD_T the outstanding loan balance at maturity (before applying the non-negative equity guarantee), V_T the property value at T , and β the maximum probability of triggering the Non-Negative Equity Guarantee (NNEG), set at 45%,
- NNEG the value of the non-negative equity guarantee, C_0 the loan principal at t_0 , and γ the maximum acceptable NNEG value, expressed as a percentage of the loan amount, set at 8.5%.

Each constraint is assessed scenario by scenario and then aggregated, which allows capturing both expected performance and extreme situations through averages and quantiles.

Classical modeling of real estate returns

The central model is based on a Vasicek process, calibrated on the INSEE national real estate index, in line with the real-world Ahlgrim ESG framework. The economic scenarios thus generated are then integrated into the pricing tool. The obtained scenarios are validated through moment analysis, comparison with the historical data used for calibration, and a real-world consistency check.

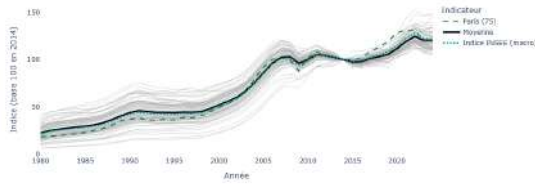


In this framework, the average LTV stands at 47.2%; individual disparities stem from the characteristics influencing borrowers' longevity.

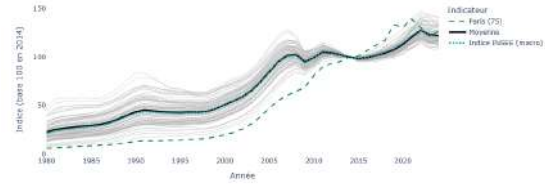
Granular modeling of real estate returns

To account for the spatial and typological heterogeneities hidden by national aggregation, return projections are carried out at a fine level (department \times property type, i.e. 182 segments). Two methods are used : Method 1, based on segmented ESGs, and Method 2, relying on a recurrent neural network of the LSTM type, with segment encoding.

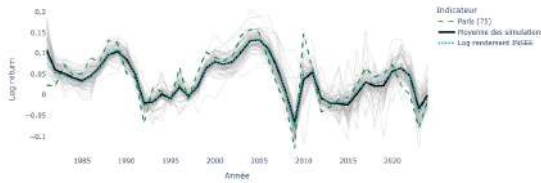
To implement these two methods, detailed real estate data are required. However, the few publicly available segmented real estate indices are shallow and do not cover the entire French territory, while public data at this level of granularity remain limited, with a history of barely ten years. A hybrid model combining hedonic regression and repeat sales was implemented, allowing the construction of indices at the desired level from the DVF database (2014–2024). These indices were then extended backwards in time to increase historical depth. This backcasting is facilitated by the availability of observed macroeconomic variables, unlike forward-looking projections which rely on simulated scenarios. A post-processing adjustment was applied to ensure aggregate consistency : the average of the reconstructed indices was aligned with the global real estate index published by INSEE, and extreme values were corrected to limit excessive deviations from this average, thereby ensuring a realistic continuity in the reconstructed time series. The application of this method constitutes one of the study's limitations, as it introduces a bias. A critical perspective is therefore necessary when interpreting the results obtained. The segmented indices and returns, aligned with the INSEE index, show the following dynamics :



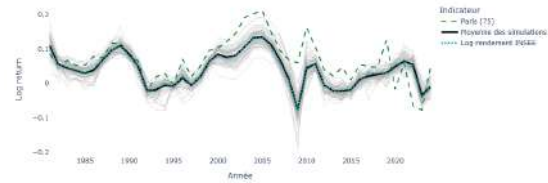
(a) Indices – Houses



(b) Indices – Apartments



(c) Log return – Houses



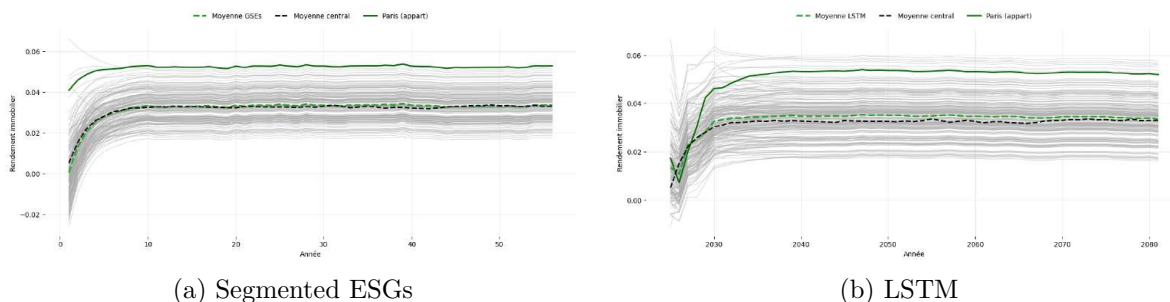
(d) Log return – Apartments

These returns were then used for the two methods described below :

Method 1 – Segmented Economic Scenario Generators : To ensure a stable and robust segmented modeling, several distinct scenario generators were developed, each dedicated to a segment and incorporating the same set of four key variables : short rate, long rate, inflation, and segmented real estate return. This structure captures the specific dynamics of each segment while relying on a common economic framework. Thus, this method is very close to the aggregated, so-called central, method since it is based on the same theoretical principles but with more granular parameterization. To preserve inter-segment consistency, the random residuals associated with the non-real-estate economic variables are shared across all generators. A single realization of the economic environment therefore applies simultaneously to all segments, which introduces an implicit correlation and ensures that the differences observed in the simulated paths only reflect the real-estate specificities of each segment.

The results obtained with the segmented ESGs and those from the LSTM are, on average, comparable to each other as well as to the returns under the central environment, but they exhibit differences in intra-segment and intra-scenario variance :

Method 2 – Recurrent Neural Network : To capture the complex and non-linear temporal dependencies in the evolution of departmental real estate returns, a recurrent neural network of the *Long Short-Term Memory* (LSTM) type was implemented and used recursively. This model takes as input the economic variables from the central ESG (short rate, long rate, inflation, global real estate return), together with *embedding* vectors representing the geographical and typological segments. The predictions generated at each period are fed back into the model to serve as input for the following period, allowing trajectories to be projected over a long horizon. This architecture ensures consistency with historical trends while capturing heterogeneity across departments and property types, thus providing a flexible framework to simulate the impact of economic scenarios on local dynamics.



The simulations produced are integrated into the pricing tool, allowing for a precise assessment of their impact in the context of reverse mortgages. While the averages of projected returns remain close, differences in variance nevertheless emerge within the scenarios of a given segment. Through the combined effect of compound interest, the unequal weighting of departments within the portfolio, and the influence of individual longevity on portfolio-level indicators, these differences in dispersion and trajectories affect loan pricing.

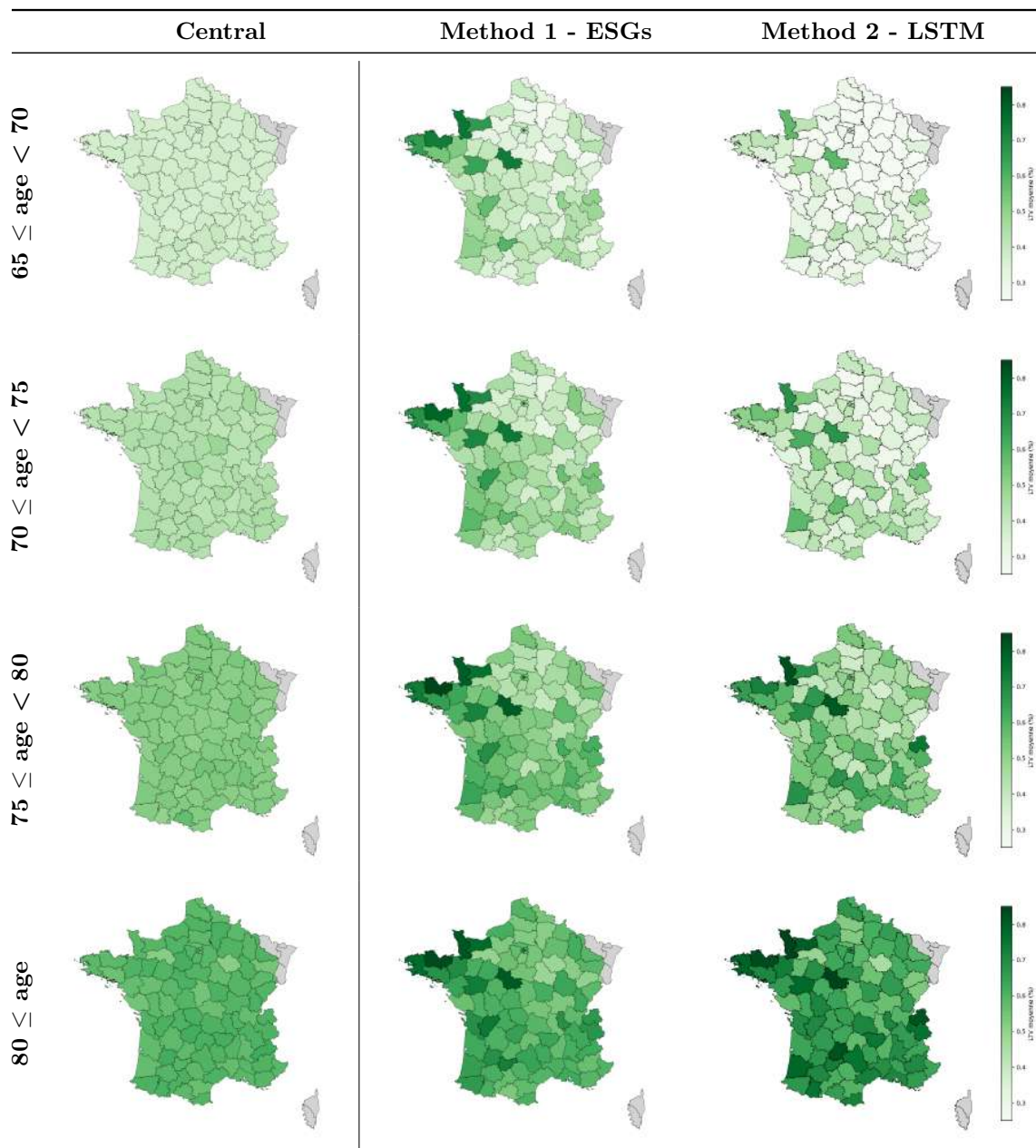
Results and analysis

Both approaches significantly alter the pricing of the portfolio. In other words, the actually binding constraints ((i)–(iv) below) differ depending on risk profiles, and their hierarchy varies considerably from one method to another :

Indicators	Central	Segmented ESGs	LSTM
Flat return	9.8 %	9.9 %	9.8 %
Compounded return	5.3 %	5.3 %	5.4 %
Average loan (€)	88,329	97,223	87,171
NNEG probability	35.0 %	38.0 %	28.4 %
Average LTV	47.2 %	52.0 %	46.6 %
(i) Share of portfolio constrained by flat return	13.2 %	15.2 %	13.9 %
(ii) Share of portfolio constrained by flat return p5	6.6 %	1.3 %	–
(iii) Share of portfolio constrained by NNEG probability	46.3 %	80.2 %	5.9 %
(iv) Share of portfolio constrained by NNEG value	33.5 %	2.9 %	79.8 %
Share of portfolio rejected	0.4 %	0.4 %	0.4 %
Share of portfolio — Total	100 %	100 %	100 %

Thus, the average LTV evolves only slightly; however, significant discrepancies emerge when the level of analysis is refined. The mapping of LTVs by department and age confirms the weight of these two factors in the calculation of the optimal LTV. Nevertheless, certain LTV levels appear high, particularly in Brittany and in Paris, and are subject to reservations and further analyses.

The integration of segmented real estate scenarios into pricing significantly modifies the departmental distribution of LTVs, underlining the relevance of segmented modeling for coherent pricing along this dimension. Concretely, beyond departmental divergences, a trend emerges toward accentuating the differentiation already present in the central model : a decrease in LTVs that are already low among the youngest and an increase in LTVs that are already high among the oldest. This "stretching of the extremes" is even more pronounced under the LSTM method and can be explained by the asymmetric nature of the PVH payoff, captured by the NNEG value constraint within the developed pricing model. This sensitivity of LTV to return and longevity assumptions of cohorts is already documented in the literature [A-W. Shao et al., 2014].



Prudent Pricing

Beyond the individual model-based pricings, a prudent pricing approach is implemented by combining the LTVs obtained from the different segmented models. Two types of adjustments have been explored, either before calibration or at the model output stage, to reduce LTVs considered too high, particularly in Paris and Brittany. These adjustments aim to temper projected high returns, themselves inherited from a historical context marked by strong price increases. Under this prudent approach, the retained LTV systematically corresponds to the lowest value among those estimated by the different models. This choice, based on a critical analysis of the initial results, seeks to ensure long-term consistency between loan amounts and local real estate dynamics, while reducing the risk of excessive exposure.

This approach leads to an average LTV of 43%, consistent with market conditions but noticeably lower than that of the central model. Returns and risk indicators remain broadly

stable despite this additional prudence; only the loan volume is impacted. Although this approach introduces a margin of prudence, one fifth of borrowers see the loan amount offered increase compared to the central model (up to +20%). This result illustrates that modeling segmented real estate return scenarios provides a more accurate representation of actual risk and increases attractiveness for certain profiles, compared to a homogeneous approach across the entire market. Finally, a cartographic analysis by department, complemented by aggregated indicators, confirms the persistence of territorial disparities, which constitutes the key contribution of this segmented approach. Despite the prudence adopted, real estate segmentation remains evident and represents a crucial lever for finer risk management.

Limitations and Perspectives

The initial ambition was to develop an individual-level real estate projection model, directly linked to detailed property characteristics (precise location, housing quality, condition, etc.) and the borrower's profile. Such an approach would have allowed for personalized projections, closely aligning housing market dynamics with the actual risk borne by the lender. However, limited data availability, the technical complexity of such a framework, and the time constraints of this thesis led to a refocusing of the work. The choice was therefore made to favor a segmented modeling approach, based on groupings by department and property type. This approach, more realistic in the context of a thesis relying on *open-source* data, maintains robust explanatory and forecasting capacity while paving the way for future, more granular developments using richer individual-level data.

The main limitations of the study stem from the availability and granularity of real estate data, which condition the quality of the constructed indices. Future work could build on a deeper historical dataset or more detailed information on the underlying properties, thereby strengthening the robustness of the approach. It would also be relevant to adopt the borrower's perspective, in order to better understand expectations, identify the constraints faced, and ultimately propose more attractive solutions.

This thesis highlights that a refined modeling of real estate returns is a powerful lever to improve the pricing of ERMs. However, other enhancements could also strengthen both the robustness and attractiveness of the product, such as the development of new drawdown modalities or the expansion of the range of loan types considered, both of which directly influence pricing.

Sommaire

Introduction	1
1 Contexte	2
1.1 Cadre général	2
1.1.1 Les seniors : situation économique et projection	2
1.1.2 L'immobilier	3
1.2 Le prêt viager hypothécaire	6
1.2.1 Fonctionnement du prêt viager hypothécaire	6
1.2.2 Marché du prêt viager hypothécaire	10
1.2.3 Risques liés au prêt viager hypothécaire	11
1.3 Objectifs du mémoire	13
2 Construction du portefeuille	17
2.1 Généralités	17
2.2 Démarche suivie	17
2.2.1 Caractéristiques des seniors	18
2.2.2 Caractéristiques du patrimoine immobilier	20
2.2.3 Intéraction senior - patrimoine immobilier	23
2.2.4 Retraitements liés aux spécificités du prêt viager hypothécaire	24
2.3 Analyses et regard critique	25
2.3.1 Analyses globales	25
2.3.2 Cartes	25
2.3.3 Statistiques	27
2.3.4 Limites	27
3 Outil de tarification en environnement central	32
3.1 Méthodologies de tarification	32
3.2 Généralités	32
3.2.1 Valeur Actuelle Nette	33
3.2.2 Taux Effectif Global	33
3.2.3 <i>Value at Risk</i>	34
3.3 Scénarios de mortalités	34
3.3.1 Tables de mortalité et probabilités sur une tête	34
3.3.2 Probabilités sur deux têtes	35
3.3.3 Simulations	36
3.4 Scénarios économiques	36
3.4.1 Modèle d'Ahlgrim	36
3.4.2 Modélisation des rendements immobiliers	37
3.4.3 Modélisation de l'inflation	39
3.4.4 Modélisation conjointe des taux court et long terme	40
3.4.5 Taux zéro-coupon et construction des taux d'actualisation	42

3.4.6	Simulations	44
3.5	Contrainte d'optimalité de la <i>Loan-To-Value</i>	47
3.5.1	Rentabilité	48
3.5.2	Gestion de risque	49
3.6	Analyses et regard critique	50
3.6.1	Analyses à la maille portefeuille	51
3.6.2	Limites	57
4	Modélisation segmentée du risque immobilier	63
4.1	Étape préliminaire : Construction d'indices segmentés	63
4.1.1	Principes théoriques	64
4.1.2	Construction des indices département \times type de bien	67
4.1.3	Projection passée des indices	69
4.2	Modélisation via GSEs segmentés	73
4.2.1	Projection des indices, branchement aux GSEs	73
4.3	Modélisation via réseau de neurones récurrent	75
4.3.1	Concepts généraux et application à l'immobilier	76
4.3.2	Fondements théoriques	77
4.3.3	Mise en application	83
4.4	Résultats et analyses	87
4.4.1	Analyses graphiques	87
4.4.2	Analyses de l'explication de la variance de la <i>Loan-To-Value</i>	92
4.4.3	Comparaison des scénarios économiques projetés	93
4.4.4	Analyses d'indicateurs quantitatifs	95
4.5	Regard critique	96
4.5.1	Rationnel sur les <i>Loan-To-Value</i> jugées trop élevées	96
4.5.2	Rationnel de modification des perspectives immobilières	97
4.5.3	Analyses après retraitements	98
4.6	Tarification prudente	100
5	Conclusion et limites	105
5.1	Conclusion	105
5.2	Limites	106
	Bibliographie	111
	Liste des figures	111
	Liste des tableaux	114
	Annexes	116
.1	Outils utilisés	116
.2	Base de données DVF	117
.3	Portefeuille construit	118
.4	Sensibilités	119
.5	Calibrage GSE	121
.6	Construction d'un indice	123
.7	Projection passée	125
.8	Réseaux de neurones	129
.9	Analyses - Modélisation segmentée	131

Introduction

Dans un contexte de vieillissement démographique marqué et de baisse des revenus à la retraite, les seniors se retrouvent souvent confrontés à une insuffisance de liquidités alors même qu'ils disposent d'un patrimoine immobilier non négligeable (60% du patrimoine total selon l'INSEE). Face à ce déséquilibre, le prêt viager hypothécaire (PVH) apparaît comme un outil pertinent, permettant aux emprunteurs âgés de monétiser partiellement leur logement tout en continuant à y vivre.

Pourtant, le PVH reste peu développé en France. Plusieurs freins persistent : du point de vue de l'emprunteur, ce produit apparaît comme coûteux et complexe. Pour le prêteur, il s'agit d'un produit soumis à des risques multiples - dont le risque de longévité et de dépréciation immobilière sont les principaux - et difficile à structurer de manière rentable et prudente. Dès lors, une question se pose : Comment, du point de vue du prêteur, améliorer la gestion de ces risques pour proposer un produit plus rentable et potentiellement plus attractif à terme ?

Une meilleure mesure des risques permettrait d'ajuster finement la tarification selon les profils emprunteurs, et in fine, d'optimiser la composition et la gestion du portefeuille de prêt. Ce mémoire se concentre sur le risque de dépréciation immobilière. L'étude stratégique d'un portefeuille de PVH repose sur la projection des flux associés aux contrats, simulés sous différents scénarios économiques et de mortalité de manière à capter l'asymétrie du *payoff* du PVH. Or, pour estimer ces flux, la valorisation future des actifs immobiliers joue un rôle central. Cette projection repose généralement sur un processus stochastique unique, tel qu'un modèle de Vasicek pour les rendements ou un Black-Scholes pour les prix, appliqué de manière identique à l'ensemble du portefeuille. Une telle approche implique que tous les biens partagent le même profil d'évolution, indépendamment de leurs caractéristiques propres. Il en résulte une *Loan-To-Value* déterminée principalement en fonction des caractéristiques de l'emprunteur, sans prise en compte de la nature du bien, de sa localisation ou de sa dynamique de valorisation spécifique.

Dans ce mémoire, une modélisation plus fine de l'évolution des prix immobiliers est proposée afin de mieux prendre en compte l'hétérogénéité du marché immobilier et ainsi améliorer la mesure des risques du prêteur. Cette approche vise à dépasser les modèles appliqués uniformément sur l'ensemble du marché en intégrant des dynamiques différenciées selon les caractéristiques du bien (localisation et type). L'impact de cette modélisation est évalué à l'aide d'un outil de tarification permettant de calculer la *Loan-To-Value* maximale sous contraintes de rentabilité et de gestion du risque. L'objectif est de mesurer les impacts d'une projection immobilière segmentée sur la tarification du PVH. Une telle approche pourrait rendre la tarification plus juste, équitable et robuste, en tenant compte à la fois des enjeux économiques et prudentiels du produit.

Chapitre 1

Contexte

1.1 Cadre général

L'analyse des données démographiques et financières récentes permet de mieux cerner les dynamiques actuelles du système de retraite et les défis auxquels il est confronté. Le contexte actuel est marqué par un vieillissement accéléré de la population et une tension croissante sur le ratio entre actifs et retraités.

1.1.1 Les seniors : situation économique et projection

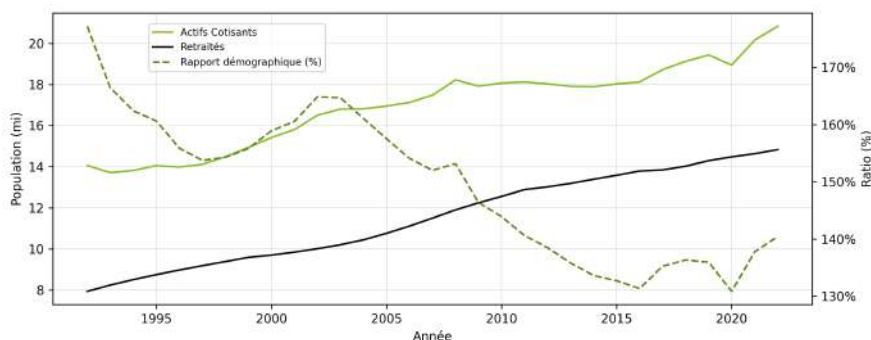
Vieillesse de la population

En effet, en 2024, plus d'un quart de la population française avait 60 ans ou plus, contre moins de 20% en 1994¹. Cette évolution résulte à la fois d'un accroissement de la natalité après la Seconde Guerre mondiale (*baby-boom*) et d'un allongement significatif de l'espérance de vie aux âges élevés. D'après les projections de l'INSEE², ce vieillissement de la population devrait se poursuivre au moins jusqu'en 2040 avec notamment une augmentation de 50% de la population de plus de 75 ans d'ici 2030.

Ratio de dépendance et cotisants *vs.* retraités

Cette transition démographique a des conséquences majeures sur le ratio démographique (de 65 ans), défini comme le rapport entre la population âgée de plus de 65 ans et celle des 20-64 ans. En 2021, ce ratio s'élevait à hauteur de 40%, et devrait continuer à croître dans les prochaines décennies selon l'INSEE.

FIGURE 1.1 – Evolution du ratio de dépendance en France



Source : [DRESS, 2023]

Le ratio démographique influence directement le rapport entre le nombre de cotisants et le nombre de retraités, un indicateur clé de la soutenabilité du système de retraite, basé sur la solidarité intergénérationnelle. Pour le régime général, ce ratio s'élevait à 140% en 2022

1. [Rapport d'évaluation des politiques de Sécurité Sociale, 2023].

2. [INSEE, IFOP et Sociovision, 2020].

(soit 20,8 millions de cotisants pour 14,8 millions de retraités), contre une moyenne de 120% sur l'ensemble des régimes. Ce rapport connaît une tendance baissière depuis les années 1970, époque à laquelle il atteignait 400%. Quelques hausses ponctuelles ont toutefois été observées, notamment au début des années 2000 et entre 2017 et 2019, dans un contexte économique favorable et du fait du relèvement de l'âge légal de départ à la retraite. Néanmoins, la hausse continue de l'espérance de vie combinée à l'évolution démographique laisse présager une poursuite de la dégradation de ce ratio.

Patrimoine immobilier : un levier

Les éléments discutés ci-avant tendent à réduire le niveau de vie à la retraite, comme en témoigne l'évolution du taux de conversion des revenus avant et après cessation d'activité. En effet, selon le COR 2024³, le taux de remplacement net à la liquidation du cas type de non-cadre du secteur privé passerait de 75% pour la génération 1960 à moins de 70% pour les générations postérieures à 1975, sous le scénario de référence⁴. Pour préserver un niveau de vie adéquat, les seniors sont ainsi conduits à rechercher de nouvelles sources de financement. Parmi leurs actifs, le patrimoine immobilier constitue souvent le poste le plus important (à hauteur de 60% du patrimoine total selon l'INSEE). C'est donc naturellement vers cette ressource que se tournent les réflexions pour développer des solutions de sécurisation financière pérennes, intégrant une dimension immobilière au cœur des stratégies.

1.1.2 L'immobilier

Caractéristiques du marché de l'immobilier

L'immobilier est un marché très singulier, il représente l'ensemble des transactions liées à l'achat, la vente, la location et la gestion des biens immobiliers. Ce secteur est l'un des piliers de l'économie et est influencé par divers facteurs socio-économiques, politiques et environnementaux. En effet, des grandeurs macro-économiques telles que le niveau de croissance économique, les taux d'intérêt et le taux de chômage affectent la demande et l'offre de biens immobiliers.

Le marché immobilier, notamment français, a connu des évolutions importantes ces dernières années. Actuellement 20% des propriétaires sont des primo-accédants, ce qui traduit une dynamique de renouvellement. De plus, le marché a été fortement impacté par les données macro-économiques, notamment l'évolution des taux. L'impact des taux est particulièrement visible lors des grandes phases de variation monétaire, en particulier :

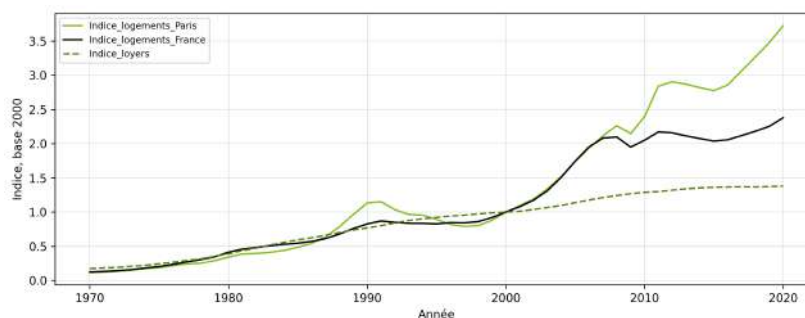
- Les pics élevés des années 1980 et début 1990, avec une forte inflation mondiale, freinant le marché immobilier,
- La forte baisse des taux post-2010, jusqu'à environ 2019 ; la présence de taux négatifs a favorisé la hausse des prix immobiliers,
- La remontée rapide des taux d'intérêt pour lutter contre l'inflation post-covid a commencé à freiner la dynamique immobilière.

Comme le montre l'évolution de l'indice des logements et des loyers, le marché immobilier a connu une forte tendance haussière depuis les années 1970, malgré plusieurs crises. Ainsi, le pouvoir d'achat immobilier a diminué de 25m² entre 1999 et 2023.

3. [Rapport annuel du COR, 2024].

4. Une description du scénario de référence est faite dans la publication [Rapport annuel du COR, 2024], qui prend elle-même pour source la DREES, modèle Trajectoire : *hypothèses démographiques centrales de l'INSEE (poursuite des gains d'espérance de vie, fécondité de 1,8 enfant par femme et solde migratoire net de 70 000 personnes par an), croissance annuelle de la productivité horaire du travail de 1,0% (à partir de 2040) et taux de chômage de 5,0% (à partir de 2030)*.

FIGURE 1.2 – Indice des logements et des loyers



Source : [J. Friggit, 2025]

Le marché de l'immobilier se distingue des autres actifs par plusieurs caractéristiques :

- **Faible liquidité** : L'immobilier est considéré comme un marché peu liquide. En effet, la possibilité de convertir rapidement les biens en liquidités est limitée - contrairement aux actions ou aux obligations par exemple. Cette faible liquidité s'explique par plusieurs raisons : les procédures administratives sont complexes, des négociations prolongées sont nécessaires entre vendeurs et acheteurs, et un grand nombre de facteurs sont à prendre en compte afin d'évaluer un bien immobilier.
- **Hétérogénéité** : L'immobilier est également caractérisé par une forte hétérogénéité. En effet, chaque bien est unique en termes de taille, état, localisation et autres caractéristiques spécifiques. On peut considérer que les biens immobiliers sont difficilement substituables. La valeur d'un bien immobilier dépend de son emplacement géographique, de l'accès aux services et infrastructures, de la qualité du quartier ainsi que des caractéristiques physiques de la propriété elle-même. Cette diversité oblige les acteurs du marché à effectuer des analyses approfondies pour déterminer la valeur d'un bien.
- **Coût de transaction élevé** : De manière générale, les coûts de transaction dans l'immobilier (frais d'agence immobilière, taxes municipales et provinciales, frais de notaires, coût du financement...) représentent une forte part de la valeur d'un bien, notamment en comparaison avec d'autres marchés. Les acteurs du marché doivent prendre en compte ces coûts pour garantir la rentabilité des transactions immobilières.
- **Durée de détention longue** : Les biens immobiliers sont de manière générale détenus sur des périodes longues, en comparaison avec d'autres types d'actifs. Cette longue durée de détention est souvent associée à des stratégies d'investissement prudentes et à une stabilité financière accrue, supposant une appréciation stable et lente.
- **Nécessité de gestion** : Les biens immobiliers nécessitent une gestion active permettant de maintenir leur valeur et générer des rendements, au contraire des investissements passifs. En effet, la gestion locative, la rénovation des biens ou à minima leur entretien représentent un certain coût de gestion.
- **Sensibilité aux facteurs extérieurs** : Le marché immobilier est fortement influencé par des facteurs exogènes tels que l'environnement économique (de par les taux d'intérêt notamment), les tendances démographiques ou encore les politiques gouvernementales. En effet, la demande et les prix sont fortement impactés par les variations de ces éléments. Les acteurs se doivent donc de surveiller la conjoncture économique et réglementaire pour adapter leurs stratégies et prendre des décisions éclairées.

Ces caractéristiques spécifiques expliquent pourquoi l'analyse fine du marché immobilier nécessite des approches adaptées et rigoureuses. Plusieurs études approfondies ont été réalisées ;

les travaux de J. Friggit constituent une référence clé pour analyser les évolutions de long terme du marché immobilier français.

Données et études Friggit

Plusieurs études d'évolution du prix de l'immobilier d'habitation à long terme ont été réalisées par J. Friggit et permettent de dresser un panorama de l'évolution du marché immobilier en France. Dans ses travaux, J. Friggit construit un indice des logements anciens en France (y.c. logements parisiens), un indice des logements parisiens et un indice des loyers (c.f. 1.2).

Depuis les années 1970, l'indice des loyers à structure constante (base 1970) rapporté au revenu disponible des ménages reste relativement stable ; ce phénomène est appelé *tunnel de Friggit* dans la littérature. Il en est de même pour l'indice des logements à structure constante (base 1970) rapporté au revenu disponible des ménages, jusqu'aux années 2000. Une forte augmentation de l'indice sur les années 2000-2005 a eu pour impact la sortie de l'indice de ce 'tunnel'. Cet indicateur fluctue depuis les années 2010 autour de 1,75. Le contraste entre les deux indices présentés ci-avant est dû à l'effet inflationniste de l'environnement financier et notamment à l'environnement de taux bas.

Cependant, la part du loyer dans le revenu disponible des ménages n'a fait qu'augmenter depuis 1970⁵. D'autres indicateurs tels que la vitesse de rotation du marché immobilier, qui s'obtient en divisant le nombre de transactions de logements anciens par le nombre de logements du parc, sont intéressants. Cet indicateur a doublé entre 1970 et 2000 et stagne autour de 3,2%, défini comme la tendance long terme de cet indicateur⁶.

Un autre indicateur essentiel de l'activité du marché immobilier est le volume des transactions immobilières. Il permet d'apprécier la liquidité du marché, la confiance des acheteurs et vendeurs, ainsi que les effets de conjoncture macroéconomique ou réglementaire. Depuis les années 1970, le nombre annuel de transactions sur les logements anciens en France a plus que triplé.

Comparés à d'autres classes d'actifs, les actifs immobiliers ont enregistré des rendements particulièrement élevés au cours des dernières décennies, notamment en Île-de-France. Bien que cette hausse des prix soit marquée par une forte hétérogénéité géographique, elle s'observe sur l'ensemble du territoire national⁷. Cette dynamique a contribué à un enrichissement significatif des ménages propriétaires, en parallèle d'un accroissement des inégalités, notamment au détriment des locataires. Les seniors, en particulier, détiennent aujourd'hui une part importante de leur patrimoine sous forme d'actifs immobiliers, ce qui renforce l'intérêt d'une réflexion autour de la mobilisation de ce patrimoine dans une optique de financement de la retraite.

Par ailleurs, les données de l'INSEE pour les années 2019 et 2020 mettent en évidence une dynamique forte entre les seniors et la question du logement :

- 81% des retraités sont propriétaires de leur résidence principale,
- 85% des seniors souhaitent vivre leur retraite à domicile,

5. [IGF-CGEDD, 2021].

6. [J. Friggit, 2025].

7. Le lecteur intéressé pourra trouver en Annexe 5 la carte des évolutions de prix à l'échelle départementale, de 2014 à 2024, en France, basée sur les données DVF.

- Un retraité sur deux déclare rencontrer des difficultés financières ; les coûts liés à une perte d'autonomie varient entre 400€ et 2 500€ par mois,
- 8 seniors sur 10 se disent opposés à la vente en viager ; 34% se déclarent néanmoins ouverts à d'autres solutions de monétisation de leur patrimoine.

Ces éléments sont notamment mis en avant par Mirabelle, un nouvel acteur de la *Silver Economy*, dans son communiqué de presse de décembre 2024⁸. Dans ce contexte, le prêt viager hypothécaire (PVH) apparaît comme une piste intéressante pour concilier maintien à domicile et besoin de liquidité.

1.2 Le prêt viager hypothécaire

Le marché du viager et des prêts viagers hypothécaires offre une illustration concrète de la connexion entre immobilier et retraite, permettant aux seniors de mobiliser leur patrimoine immobilier pour financer leurs besoins.

Contrairement à une vente classique ou à un viager traditionnel, le PVH permet aux seniors de conserver l'entière propriété de leur bien immobilier tout en bénéficiant d'un capital ou de revenus réguliers. Le développement des PVH s'appuie également sur un cadre sécurisant pour les emprunteurs et leurs héritiers, de par la clause de non-dépassement (*Non Negative Equity Guarantee*, NNEG). Par ailleurs, l'absence de remboursement tant que l'emprunteur est en vie et le recours à une hypothèque sur le bien immobilier comme garantie renforcent l'attractivité de ce produit, le distinguant des autres formes de crédit ou d'arbitrage patrimonial.

1.2.1 Fonctionnement du prêt viager hypothécaire

Généralités

Le prêt viager hypothécaire peut être défini comme un prêt — le plus souvent à remboursement in fine — accordé exclusivement aux propriétaires immobiliers, relativement âgés, dans lequel l'emprunteur met son bien en hypothèque en garantie du remboursement futur du capital prêté.

À l'instar de la vente en viager, le PVH permet de mobiliser le capital immobilisé dans un bien immobilier sans en nécessiter le départ immédiat. Le montant prêté peut être obtenu en totalité au début du prêt (*lump sum*) ou sous forme d'un montant en début de prêt et de tirages durant la vie du prêt. L'emprunteur dispose alors d'une partie du montant total disponible pour le prêt et d'un droit de tirage jusqu'au montant maximal prêté défini en amont. Le mécanisme de tirage permet de diminuer l'effet capitalisation des intérêts composés du prêt.

Par le mécanisme de garantie de l'hypothèque, ce type de prêt permet aux personnes âgées (âge minimal fixé à 60 ou 65 ans par la plupart des acteurs du marché) d'accéder à un prêt permettant une libre utilisation des fonds. Contrairement aux prêts classiques, le PVH ne prévoit pas d'échéancier de remboursement fixe : le remboursement du capital, ainsi que des intérêts éventuellement capitalisés, intervient au moment de la revente du bien, que celle-ci soit réalisée du vivant de l'emprunteur ou à son décès. Comme tout type de prêt, le coût du financement se traduit par des intérêts. L'emprunteur peut choisir de rembourser ces intérêts tout au long de la vie du prêt sous forme de mensualités (très peu courant sur le marché), ou bien d'opter pour un remboursement in fine. Dans ce second cas, les intérêts sont capitalisés, c'est-à-dire ajoutés au capital restant dû (CRD) et remboursés à la clôture

8. [Mirabelle, 2024]

du prêt. Ce mécanisme de capitalisation, reposant sur les intérêts composés, peut entraîner une forte augmentation du montant total à rembourser à long terme.

Garantie de non-dépassement (NNEG)

Un des mécanismes clés du PVH est la garantie de non-dépassement - ou *Non Negative Equity Guarantee* (NNEG) en anglais - présente dans les contrats pour protéger l'emprunteur et ses héritiers. Cette garantie impose que, au moment du remboursement du prêt, le montant exigible ne peut excéder la valeur du bien immobilier mis en hypothèque, quelle que soit l'évolution du marché immobilier ou la durée du prêt. Autrement dit, même en cas de dépréciation significative du bien ou de très longue durée de vie de l'emprunteur (ce qui accroît mécaniquement le capital dû via la capitalisation des intérêts), les héritiers ne seront jamais redevables d'un montant supérieur à la valeur de revente du bien.

Couple *Loan-To-Value* - Taux d'intérêt

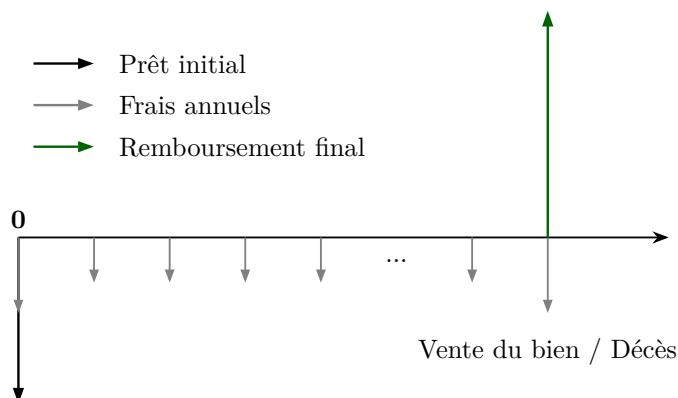
Un paramètre central dans la structuration d'un prêt viager hypothécaire est la *Loan-To-Value* (LTV). Elle traduit le montant du prêt et est définie comme le rapport entre le montant prêté à t_0 et la valeur du bien immobilier à cette même date. Ce ratio, généralement compris entre 15% et 70% selon les caractéristiques de l'emprunteur et de son bien, permet au prêteur de se protéger contre le risque de non-recouvrement (lié à la NNEG), notamment en cas de longévité élevée ou de baisse des prix de l'immobilier. En parallèle, le taux d'intérêt appliqué au prêt – souvent fixe et supérieur à celui des prêts immobiliers classiques en raison de la durée et des risques associés – a un effet direct sur le rythme de croissance du capital restant dû. Un taux proche de 6% apparaît comme la référence actuelle du marché ; il peut néanmoins être ajusté en modulant la LTV ou d'autres paramètres du prêt.

L'interaction entre la LTV et le taux d'intérêt est déterminante pour la probabilité de déclenchement de la garantie de non-dépassement : une LTV élevée combinée à un taux d'intérêt important augmente fortement le risque que le capital dû excède la valeur future du bien, surtout si la croissance immobilière est faible. Ainsi, la tarification du PVH repose sur un arbitrage délicat entre niveau de LTV et le taux d'intérêt, selon les caractéristiques de l'emprunteur et la valorisation anticipée du bien.

Flux financiers liés au PVH

Le prêt viager hypothécaire repose sur des flux financiers bien définis entre l'emprunteur et le prêteur.

FIGURE 1.3 – Flux financiers liés au PVH



Le schéma global des flux met en évidence trois catégories principales de flux :

- Le prêt initial : à t_0 , le prêteur verse un montant initial correspondant au capital prêté, c'est un flux sortant. Ce montant est égal à la valeur initiale du bien multipliée par la LTV du prêt,
- Les frais : tout au long de la vie du prêt, le prêteur fait face à des frais. Ces frais traduisent les frais d'apporteur d'affaire à t_0 , les frais de gestion durant la vie du contrat et les frais de clotûre du prêt (et liés à la vente), à T ,
- Le remboursement final : au décès de l'emprunteur ou à la vente du bien, soit à T , un flux positif correspondant au remboursement final du prêt est versé au prêteur. Ce montant correspond à la valeur minimale entre la valeur du bien et celle du prêt en T .

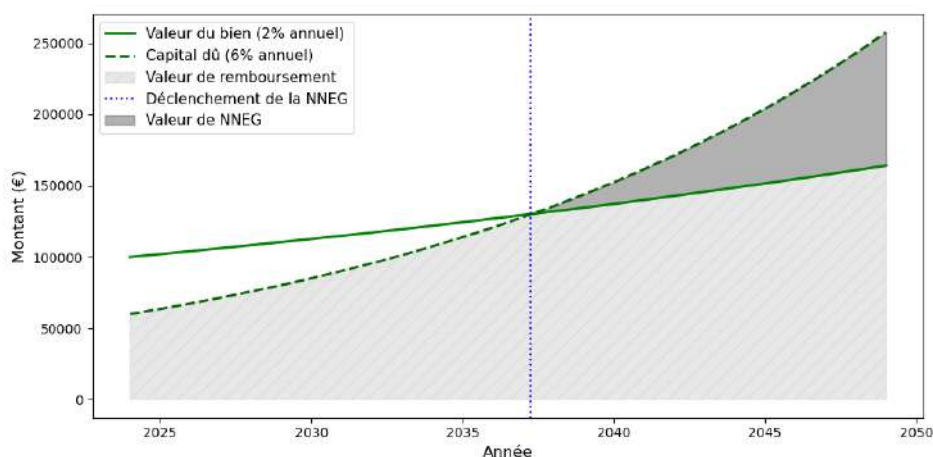
Schémas explicatifs

Le remboursement du PVH repose sur la comparaison en date T , à la date de décès ou de vente du bien, de la valeur du bien immobilier et du capital dû par l'emprunteur. Ces deux grandeurs évoluent différemment dans le temps :

- La valeur du bien suit une trajectoire de croissance liée au rendement immobilier. Dans le schéma ci-dessous, ce rendement est supposé constant à 2% par an.
- Le capital dû augmente au taux d'intérêt du prêt, ici fixé à 6% par an⁹.

Ces deux courbes peuvent se croiser : c'est le déclenchement de la garantie de non-dépassement (NNEG). Au-delà du point de croisement, le capital dû devient supérieur à la valeur du bien. Alors, la garantie de non-dépassement s'applique : le prêteur ne pourra pas récupérer plus que la valeur du bien au moment du décès.

FIGURE 1.4 – Schéma général du PVH



L'impact des paramètres de prêt ainsi que des caractéristiques individuelles des emprunteurs sur le fonctionnement du PVH est analysé ci-dessous.

Impact d'une augmentation de la LTV Une augmentation de la LTV initiale accroît mécaniquement le capital emprunté, et donc le capital dû projeté dans le temps. Cela augmente le risque pour le prêteur que la valeur du bien devienne insuffisante pour couvrir la dette en fin de prêt. Ainsi, plus la LTV est élevée, plus le croisement avec la courbe de la

9. A noter que le taux d'emprunt d'un PVH peut être variable, mais ce type de prêt n'est pas étudié ici.

valeur du bien se produit tôt (donc plus la probabilité de déclenchement de la NNEG est élevée).

FIGURE 1.5 – LTV = 60%

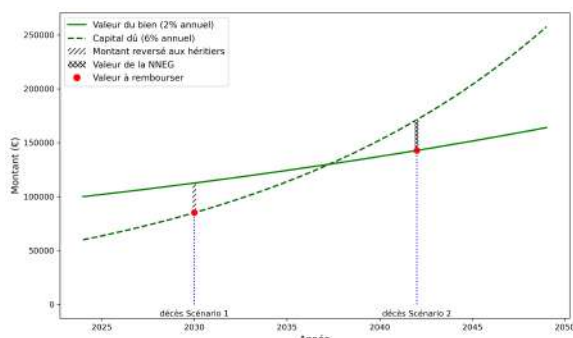
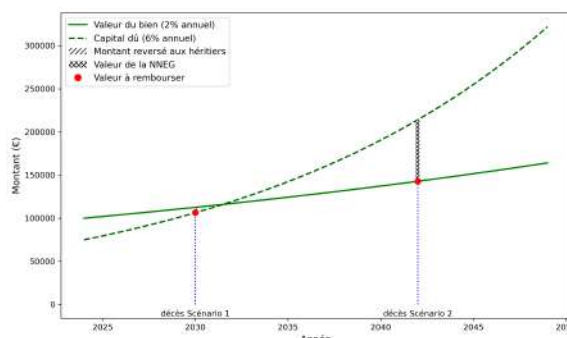


FIGURE 1.6 – LTV = 75%



Impact de la date de décès (ou de vente du bien) Lorsque le décès intervient avant le point de croisement entre la courbe de la valeur du bien et celle du capital dû, le bien est alors suffisamment valorisé pour couvrir intégralement le capital restant dû, et l'excédent éventuel est reversé aux héritiers. Le prêteur récupère 100% du capital dû, soit :

$$\text{Valeur de remboursement} = \text{LTV} \times \text{Valeur du bien} \times (1 + \text{taux})^T$$

Lorsque le décès survient après ce point de croisement, le capital dû dépasse alors la valeur du bien. Grâce à la NNEG, les héritiers n'ont rien à payer : le prêteur absorbe son manque à gagner, ne récupérant que la valeur du bien, même si elle est inférieure à la dette. La rentabilité pour le prêteur est alors altérée. Si la valeur du bien à T est inférieure au montant prêté, le prêteur est alors en perte.

FIGURE 1.7 – Décès avant le croisement

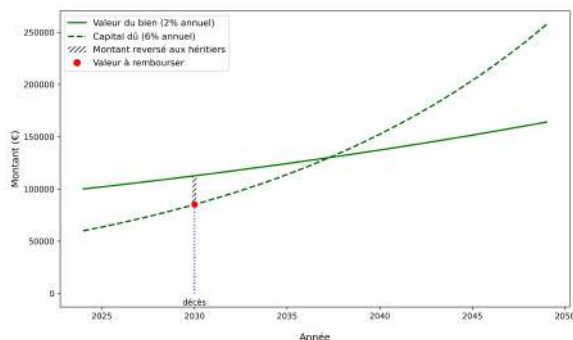
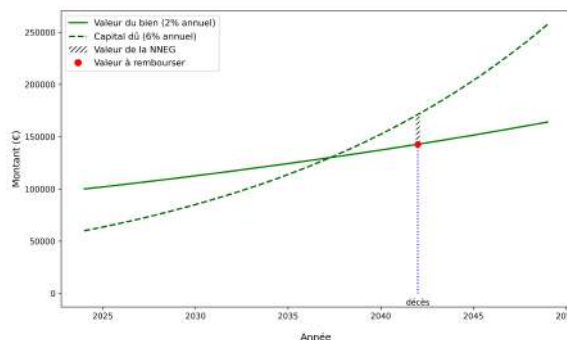


FIGURE 1.8 – Décès après le croisement



Ces graphiques mettent en évidence le rôle clé de la croissance relative des deux composantes (valeur du bien et dette) dans la structuration du risque du PVH. Le risque pour le prêteur est amplifié lorsque l'horizon s'allonge, surtout si le rendement immobilier est insuffisant pour compenser les intérêts composés. Ainsi, l'âge de l'emprunteur joue un rôle déterminant dans le calibrage de la LTV : plus l'âge est élevé, plus l'horizon de remboursement est court, ce qui limite la capitalisation de la dette et permet mécaniquement une LTV plus élevée.

Impact d'une augmentation du rendement immobilier Lorsque le rendement immobilier augmente, la courbe de la valeur du bien croît plus rapidement, ce qui repousse la date de déclenchement de NNEG (et réduit sa probabilité d'occurrence en moyenne). Autrement

dit, le risque pour le prêteur diminue, même pour des durées de vie longues. En réponse, le prêteur est favorable à augmenter l'assiette de prêt, anticipant une probabilité plus faible que le capital dû dépasse la valeur du bien au moment du décès, avec un rendement augmentant en parallèle.

Cet effet est particulièrement marqué chez les emprunteurs jeunes, car l'effet cumulatif des intérêts composés s'exprime davantage sur de longues durées. Ainsi, une hypothèse de rendement immobilier élevée justifie mécaniquement une LTV initiale plus élevée, même si le capital reste longtemps en circulation.

FIGURE 1.9 – Rdt annuel fixe 2% annuel

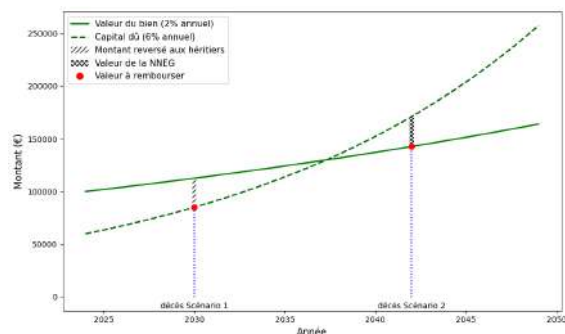
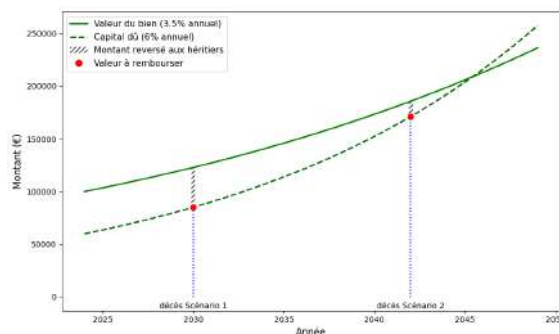


FIGURE 1.10 – Rdt annuel fixe 3,5% annuel



1.2.2 Marché du prêt viager hypothécaire

Marché français

Le marché du PVH en France reste encore peu développé par rapport à d'autres pays comme le Royaume-Uni ou les États-Unis, où les produits dits d'*Equity Release* - ou *Equity Release Mortgage* - sont largement diffusés. Introduit en France par la loi n° 2006-872 du 13 juillet 2006¹⁰, le PVH a longtemps souffert d'un cadre juridique et opérationnel peu attractif pour les établissements prêteurs.

Après la fermeture du PVH commercialisé par le Crédit Foncier en 2019, seuls quelques acteurs traditionnels proposent des offres sur ce segment. La Banque Populaire¹¹ commercialise ce type de prêt via son produit *Foncier Reversimmo*. Toutefois, ces dernières années ont vu l'émergence de nouveaux acteurs spécialisés, comme Arrago, ou encore Mirabelle - qui a déposé son dossier en vue d'obtenir l'agrément de Société de financement par l'ACPR dans l'octroi de PVH - qui ambitionnent de relancer le marché avec des offres modernisées. Ce regain d'intérêt, couplé à la conjoncture décrite en section 1.1.1, pourrait marquer un tournant dans le développement du PVH en France.

L'offre reste cependant restreinte, avec des volumes de prêts encore marginaux au regard du potentiel représenté par le patrimoine immobilier détenu par les seniors. Toutefois, le vieillissement démographique, la montée des besoins en financement à la retraite et l'essor des solutions de monétisation patrimoniale pourraient stimuler ce marché dans les années à venir.

Marché étranger

À l'international, le marché du prêt viager hypothécaire est beaucoup plus développé que sur le territoire français. Le Royaume-Uni constitue l'un des exemples les plus aboutis : depuis

10. [Legifrance, 2006].

11. Le Groupe BPCE, l'organe central du réseau Banque Populaire, détient 100% du Crédit Foncier.

1991, plus de 675 000 ménages y ont mobilisé environ 49 mds£ pour soutenir leur situation financière.

Entre 2006 et 2022, le marché de l'*Equity Release* s'est fortement développé, passant d'un volume de 1 md£ à 5 mds£, avant qu'un choc économique ne vienne affecter l'ensemble du secteur hypothécaire¹². Depuis cet épisode, l'économie britannique fait face à une forte inflation et à une hausse des taux d'intérêt. Toutefois, les prix de l'immobilier sont restés solides, ce qui signifie que les actifs immobiliers continuent de représenter une forte part du patrimoine total des ménages. Le dernier rapport trimestriel de l'*Equity Release Council* (T4 24) montre une hausse du nombre de clients souscrivant à de nouveaux produits de type *Equity Release* ainsi qu'une hausse du montant total prêté par rapport à la fin 2023. Le produit est largement distribué par des institutions financières, encadré par un cadre réglementaire clair et soutenu par des mécanismes de garantie bien définis.

TABLE 1.1 – Evolution du marché de l'*Equity Release* au Royaume-Uni

Indicateur	T4 23	T4 24	Évol. Ann.
Montant total prêté	535m£	622m£	+16 %
Nombre total de clients	13 651	15 073	+10 %

Source : [Equity Release Council, 2024a]

Aux États-Unis, les *reverse mortgages*, principalement proposés via le programme fédéral HECM (*Home Equity Conversion Mortgage*), connaissent également un essor important, bien qu'encadré plus strictement depuis la crise financière. D'autres pays comme le Canada, les Pays-Bas ou l'Australie ont également développé des formes variées de financement hypothécaire similaires au PVH. Cette dynamique étrangère offre ainsi un référentiel riche pour évaluer les perspectives d'évolution du PVH en France, tant en matière de conception produit que de diffusion commerciale.

1.2.3 Risques liés au prêt viager hypothécaire

Le prêt viager hypothécaire est un produit risqué, car il combine des incertitudes liées à la longévité des emprunteurs, à l'évolution du marché immobilier et aux comportements individuels. Cette asymétrie des flux, accentuée par la garantie de non-dépassement (NNEG), expose le prêteur à une diversité de risques qu'il doit anticiper et gérer avec rigueur.

Risque de longévité

Le risque de longévité constitue l'un des enjeux majeurs pour le prêteur dans un contrat de prêt viager hypothécaire. En effet, plus l'emprunteur vit longtemps, plus la durée du prêt s'allonge, ce qui accentue la capitalisation des intérêts en cas de remboursement in fine. Cette dynamique peut conduire à une croissance exponentielle du capital restant dû, d'autant plus marquée lorsque les taux d'intérêt sont élevés. Si la durée de vie de l'emprunteur excède largement les hypothèses initiales, cela augmente mécaniquement le risque de déclenchement de la garantie de non-dépassement, c'est-à-dire le risque que le montant dû excède la valeur du bien au moment de la revente. Ainsi, le risque de longévité dans le PVH est un risque asymétrique et non mutualisable, qui nécessite des outils de projection robustes intégrant des tables de mortalité prospectives, ainsi qu'une gestion prudente de la tarification.

12. [Equity Release Council, 2024b].

Risque de dépréciation immobilière

Le risque de dépréciation immobilière constitue un facteur de vulnérabilité important pour le prêteur dans le cadre d'un prêt viager hypothécaire. Ce risque se matérialise lorsque la valeur du bien immobilier au moment du remboursement est inférieure au capital restant dû. Dans ce cas, et en présence de la NNEG, le prêteur subit une perte en capital relativement à un prêt classique au même taux. Si la valeur du bien atteint en T une valeur inférieure au montant prêté, alors le prêteur essuie une perte réelle. Ce risque est amplifié si le prêt est de longue durée, si le bien est situé dans une zone à forte volatilité ou à marché illiquide, ou encore en période de crise économique affectant la demande immobilière. Contrairement à un emprunt classique, le prêteur se retrouve ici exposé directement à un actif physique, dont l'évolution dépend de multiples facteurs exogènes (démographie, politiques publiques, cycles économiques). Pour en atténuer l'exposition, il est essentiel de recourir à des modèles de projection immobilière, de diversifier géographiquement les portefeuilles, et de calibrer le *Loan-To-Value* en fonction du risque de marché associé au bien. Pour évaluer correctement ce risque et en saisir l'asymétrie, il est indispensable de recourir à des scénarios économiques stochastiques. Ainsi, la maîtrise du risque immobilier est au cœur de la viabilité économique du PVH sur le long terme.

Risque de taux

Le risque de taux dans un prêt viager hypothécaire se manifeste à deux niveaux distincts mais interdépendants.

D'une part, le risque de taux renvoie à l'impact des variations des taux d'intérêt sur la valeur actuelle des flux futurs liés à un portefeuille de prêts viagers hypothécaires. Dans un cadre de valorisation économique, le remboursement futur du prêt est actualisé à un taux représentatif du coût du capital ou des conditions de marché. Lorsque les taux d'actualisation augmentent, la valeur actuelle des montants futurs diminue : le prêt a alors moins de valeur pour le prêteur, car les flux espérés dans 15 ou 20 ans sont plus fortement *discountés*. Inversement, si les taux baissent, la valeur actualisée du prêt augmente, car les flux futurs ont un poids plus élevé dans l'évaluation actuelle. Ce phénomène expose donc le prêteur à une volatilité de la valorisation de son portefeuille.

D'autre part, le taux d'intérêt contractuel du prêt, qui détermine la croissance du capital restant dû (CRD), joue un rôle central dans le risque de dépassement de la valeur du bien. Si ce taux est fixe, une hausse des taux de marché crée un décalage défavorable pour le prêteur : il perçoit des intérêts à un taux devenu inférieur au marché, tout en voyant la valeur actualisée de ses flux diminuer. Si le taux est variable ou indexé, le prêteur est mieux protégé contre la hausse des taux, mais le CRD augmente plus vite, ce qui accroît le risque d'activation de la NNEG en cas de faible performance du marché immobilier. À l'inverse, une baisse des taux rend les prêts fixes plus attractifs pour le prêteur (car mieux rémunérés que le marché), mais peut inciter à des rachats anticipés ou négociations, réduisant la rentabilité attendue. Ainsi, le risque de taux affecte à la fois la valorisation actuarielle du portefeuille et la dynamique financière du prêt.

Risque de rachat

Le risque de rachat dans un PVH correspond à la possibilité que l'emprunteur décide de rembourser de manière anticipée son prêt, soit par la vente volontaire de son bien, soit par refinancement. Pour le prêteur, un rachat anticipé peut entraîner une perte d'intérêts futurs, et donc une moindre rentabilité du prêt. Ce type de cas peut être problématique si le prêt a été initialement structuré avec une longue durée de capitalisation, censée générer des intérêts

élevés à terme. Ce risque est d'autant plus pénalisant que le coût initial de mise en place du prêt (frais de dossier, commission, étude notariale) est généralement élevé, concentré sur la première année, ce qui pèse fortement sur la rentabilité en cas de sortie prématurée. Par ailleurs, la nature même du PVH implique une durée longue : la rupture prématurée du contrat modifie brutalement cette durée et peut déséquilibrer la structure de financement ou la stratégie de couverture mise en place par le prêteur.

À l'inverse, dans certains cas, le rachat peut aussi limiter l'exposition au risque de longévité ou à la dépréciation immobilière.

Risque d'aléa moral

Le risque d'aléa moral dans le cadre du prêt viager hypothécaire renvoie à la possibilité que le comportement de l'emprunteur change après l'octroi du prêt, du fait de la structure même du produit. En effet, sachant que le remboursement du capital ne lui incombera pas directement mais sera pris en charge à son décès par la valeur du bien, l'emprunteur peut être moins incité à entretenir correctement le logement, ni à en préserver la valeur sur le long terme. Cela peut entraîner une dégradation du bien et donc une baisse de sa valeur de revente, augmentant mécaniquement le risque pour le prêteur.

Risque d'anti-sélection

Le risque d'anti-sélection désigne une situation où les emprunteurs disposent d'une meilleure information que le prêteur sur leur propre situation (santé, espérance de vie, situation familiale, état du bien immobilier, etc.) et adaptent leur comportement en conséquence. Ils choisissent de souscrire au produit uniquement s'il leur est particulièrement favorable, ou au contraire, de ne pas y recourir lorsqu'il leur semble désavantageux.

Dans le cas du prêt viager hypothécaire, ce phénomène se traduit par exemple par le refus de souscription de la part d'individus qui anticipent une espérance de vie plus courte que la moyenne, car ils jugent que le produit serait peu rentable pour eux ou pour leurs héritiers. À l'inverse, des emprunteurs en bonne santé et ayant une bonne hygiène de vie pourraient y souscrire de manière préférentielle.

Cette asymétrie d'information crée un biais de sélection défavorable pour le prêteur, qui se retrouve confronté à une population non représentative de la moyenne attendue, et donc potentiellement plus risquée au regard des hypothèses utilisées dans la tarification et le provisionnement. À terme, une telle dérive peut fragiliser l'équilibre technique du portefeuille et augmenter la fréquence d'activation de la garantie de non-dépassement.

1.3 Objectifs du mémoire

Parmi les risques structurels auxquels s'expose le prêteur dans prêt viager hypothécaire, le risque de longévité et de dépréciation immobilière occupent une place majeure. Le risque de longévité a déjà été étudié dans différents mémoires d'actuariat. Le mémoire IA [G. Parat, 2016] *Couverture assurantielle du risque de longévité dans le cadre d'un Prêt Viager Hypothécaire* s'attarde notamment sur ce risque. Le risque de dépréciation immobilière ne semble pas avoir fait l'objet d'étude poussée sur le secteur français.

Ce risque se matérialise lorsque la valeur du bien est inférieure au capital restant dû au moment du remboursement. Dans un contexte où la garantie de non-dépassement interdit tout recours aux héritiers, cette situation conduit à un manque à gagner pour le prêteur, qui récupère une valeur de réalisation insuffisante pour couvrir l'encours du prêt. Ce risque

est particulièrement fort dans les prêts longue durée, à remboursement in fine, où les effets composés des intérêts peuvent largement dépasser la valeur initiale du bien si celui-ci se déprécie, stagne, ou s'apprécie moins vite que le taux fixe de l'emprunt.

Dans de nombreux travaux et pratiques de marché, la valeur future du bien est modélisée à l'aide d'indices immobiliers agrégés (type Notaires-INSEE) au niveau national, selon des modèles mathématiques type Vasicek pour les rendements ou Black & Scholes pour l'indice lui-même. Cette approche ne prend pas en compte la forte hétérogénéité des dynamiques immobilières, tant sur le plan géographique que structurel. En réalité, deux biens de même valeur initiale peuvent suivre des trajectoires radicalement différentes selon leur localisation, leur caractéristiques intrinsèques ou leur attractivité locative par exemple.

La soutenabilité économique d'un portefeuille de PVH repose en grande partie sur la capacité à anticiper correctement l'évolution de la valeur de chaque bien sous-jacent. Cela implique la mise en place d'une modélisation immobilière plus fine, intégrant des variables locales et intrinsèques au bien (surface, état, type, ...). Ce niveau de granularité permet non seulement une estimation plus réaliste de la valeur, mais aussi un meilleur calibrage de la *Loan-To-Value* (LTV) et une tarification différenciée du risque selon le profil du bien.

Ce constat fonde l'objet du présent mémoire : construire une approche fine de projection immobilière et mesurer son impact sur la tarification, sous contrainte de rentabilité et de gestion de risque, du point de vue du prêteur.

L'objectif ici n'est pas de prédire l'évolution du prix d'un bien, mais de simuler N trajectoires de prix plausibles. Une telle modélisation permettrait d'améliorer la robustesse des modèles de valorisation du PVH, de mieux maîtriser le risque de perte associé à la garantie NNEG, et, in fine, de contribuer à la viabilité et au développement de ce produit dans un contexte de vieillissement démographique.

Méthodes de projection envisagées

Plusieurs approches ont été envisagées pour projeter les rendements segmentés à long terme, en cohérence avec des scénarios économiques. Leur sélection a reposé sur deux critères principaux : la capacité à intégrer les trajectoires macroéconomiques parallèles et à produire des scénarios différenciés capable de déclencher l'asymétrie liée aux garanties du PVH.

Les PVH présentent une asymétrie importante dans leur valorisation : les scénarios économiques adverses (hausse des taux c.f. 1.2.3, baisse de l'immobilier notamment c.f. 1.2.3) entraînent une baisse de rendement pour le prêteur, tandis que les scénarios favorables sont plafonnés. Il est donc crucial que les trajectoires immobilières projetées reflètent réellement les différences entre contextes macroéconomiques positifs et négatifs. Cette exigence écarte les méthodes qui tendent à lisser les dynamiques ou à ne projeter qu'une tendance moyenne.

Séries temporelles Compte tenu de la nature des données et de l'objectif de projection sous scénarios, les modèles classiques de séries temporelles sont écartés. D'une part, la faible profondeur d'historique par segment et l'hétérogénéité entre 182 segments rendent la sélection d'ordres et l'estimation instables, favorisant le sur-ajustement et des paramètres peu robustes. D'autre part, ces modèles reposent sur des hypothèses fortes rarement vérifiées en présence de ruptures de régime macroéconomiques. En projection, leur mécanisme tend à produire des trajectoires simplifiées (retour à la moyenne, tendance linéaire ou marche aléatoire) qui amortissent les chocs et peinent à restituer des effets persistants ou non linéaires. Enfin, ils n'exploitent pas le partage d'information inter-segments, tandis que des variantes multivariées deviennent rapidement coûteuses et peu identifiables.

GSEs segmentés Au contraire, la projection des rendements immobiliers à l'aide d'un GSE monde réel, calibré à l'échelle segmentée, s'est imposée comme une approche naturelle et cohérente dans le cadre de cette étude. Elle s'inscrit pleinement dans une logique actuarielle, en traitant le rendement immobilier de chaque segment comme une variable intégrée au sein d'un dispositif de simulation économique global, et est cohérente avec la modélisation courante du marché. Cette méthode permet de générer rapidement des trajectoires différenciées, directement liées aux paramètres historiques du segment concerné. Cette architecture garantit à la fois une simplicité d'implémentation, une flexibilité des projections et leur cohérence avec les scénarios économiques globaux, tout en assurant la compatibilité avec la structure asymétrique du modèle de prêt en question.

Régressions Les régressions multivariées (linéaires ou régularisées) ont été utilisées dans le cadre du projet, notamment pour le *backcast* via une régression de type Elastic Net. Cette approche offre un bon compromis entre interprétabilité et robustesse, notamment face à la multicolinéarité des variables macroéconomiques. La régression Elastic Net n'a pas été utilisée pour projeter le futur, mais pour 'consolider' le passé en approfondissant la base historique essentielle pour le calibrage du modèle.

Machine Learning Les méthodes d'apprentissage automatique classiques comme les forêts aléatoires (*Random Forest*) ou les modèles de *gradient boosting* ont également été écartées. Bien qu'efficaces sur des tâches de prévision à court terme, ces modèles ne sont pas conçus pour des projections à long terme. Leur usage nécessiterait l'emploi d'une architecture réursive non native et ne permettrait pas une intégration fluide des trajectoires macroéconomiques projetées dans le temps. En effet, les modèles de type arbre sont puissants pour expliquer ou prédire un instantané, mais peu adaptés pour simuler des trajectoires longues.

Réseaux de neurones récurrents À l'inverse, les réseaux de neurones récurrents, et en particulier les LSTM (*Long Short-Term Memory*), sont apparus comme les plus adaptés à cette problématique et répondent à ces limites. Leur architecture permet de modéliser des séquences temporelles de manière dynamique, en intégrant les variables macroéconomiques projetées. Ainsi, le second type de modèle retenu repose sur un LSTM multivarié avec *embedding* de segment (département \times type de bien), permettant d'apprendre des dynamiques propres à chaque segment tout en profitant d'un entraînement global. Il autorise une projection réursive réaliste, différenciée selon les scénarios économiques, ce qui est indispensable pour faire émerger l'asymétrie du produit PVH.

Approche *top-down* Enfin, les approches *top-down* — qui consistent à projeter un rendement immobilier national, puis à le décliner à la maille segmentée via des retraitements a posteriori — ont été envisagées mais non retenues, la granularité d'information disponible dans le portefeuille construit ne permettant pas de justifier de manière robuste l'application d'écarts différenciés par segment (ou si possible individu) sur un rendement global. Pour autant, cette approche reste pertinente dans le cadre d'un portefeuille réel bien renseigné. Elle permettrait alors des projections très fines, potentiellement à la maille individuelle, mais au prix d'un important travail de calibrage, de suivi et de justification des écarts appliqués.

En synthèse, les méthodes retenues (Elastic Net pour le *backcast*, GSEs segmentés et LSTM pour la projection) ont été choisies pour leur capacité respective à répondre aux objectifs spécifiques de l'étude : reconstitution crédible de l'historique et projection cohérente, flexible et différenciée sur le long terme, des rendements immobiliers segmentés, en ligne avec les contraintes économiques du PVH.

Chapitre 2

Construction du portefeuille

Cette section présente la méthodologie adoptée pour construire le portefeuille de référence utilisé dans l'ensemble des analyses de cette étude. Afin de refléter au mieux les caractéristiques du marché, un échantillon de 10 000 polices a été simulé. Chaque police est définie par un ensemble de variables nécessaires à la modélisation dynamique du portefeuille, telles que l'âge, le sexe, la situation géographique ou encore la valeur du bien immobilier. Ce portefeuille conçu pour représenter fidèlement la population des emprunteurs de PVH constitue la base de l'ensemble des calculs et simulations menés par la suite.

Les individus simulés comme contractants un prêt sont des personnes dites « seniors ». En pratique, les acteurs du marché fixent un âge minimal pour pouvoir accéder à ce type de produit. Dans le cadre de cette étude, cet âge minimal est fixé à 65 ans. Les patrimoines immobiliers construits ici sont supposés composés d'un unique bien immobilier, représentant 100% de la valeur simulée.

2.1 Généralités

Comme décrit précédemment, le prêt viager hypothécaire repose sur le couple *Loan-To-Value* - taux d'intérêt. En effet, ces deux éléments impactent à la fois l'attractivité commerciale du produit, la rentabilité attendue pour le prêteur et les risques que ce dernier devra assumer. Ce couple de paramètres est directement influencé par les caractéristiques du prêt en question, c'est à dire celles de l'emprunteur et du bien immobilier sous-jacent.

Dans le but d'évaluer l'impact d'une modélisation plus fine du risque immobilier sur la rentabilité du produit et sur la gestion du risque, il est nécessaire de construire un portefeuille fictif d'emprunteur et leur patrimoine immobilier. La modélisation fine du risque immobilier implique l'utilisation de données à une maille précise ; l'utilisation de *model points* a donc été écartée. Ainsi, le portefeuille simulé est composé de 10 000 lignes chacune traduisant les caractéristiques d'un emprunteur et du bien sous-jacent.

Les hypothèses et simplifications adoptées dans la construction de ce portefeuille fictif imposent de l'aborder avec un regard critique. Les limites susceptibles de restreindre la portée ou la robustesse des résultats obtenus seront détaillées dans la section 2.3.4.

2.2 Démarche suivie

La construction de ce portefeuille s'est organisée en trois étapes :

- Simulation des caractéristiques des seniors,
- Simulation des caractéristiques du patrimoine immobilier conditionnellement aux caractéristiques des seniors,
- Retraitements liés aux spécificités du produit.

En effet, afin d’assurer la pertinence des analyses menées par la suite, le portefeuille modélisé doit, avant tout retraitement, rester cohérent avec les données nationales relatives aux seniors et à leur patrimoine immobilier. Dans un premier temps, un portefeuille fictif a été construit en tenant uniquement compte de l’interaction entre les caractéristiques des seniors et celles de leur patrimoine immobilier, sans chercher à refléter l’appétence pour un tel produit. Dans un second temps, des ajustements ont été réalisés afin de rendre ce portefeuille représentatif d’une population réelle d’emprunteur de PVH. En effet, le profil type d’un senior contractant un PVH diffère de celui d’un senior moyen en France.

2.2.1 Caractéristiques des seniors

Données utilisées

Dans un souci de cohérence avec les statistiques nationales relatives aux seniors, et en l’absence de données *open-source* spécifiques au PVH, le portefeuille fictif utilisé dans cette étude a été construit à partir de données publiques.

En particulier, les données de recensement publiées par l’INSEE à la maille communale ont servi de référence. Le jeu de données TD_POP4_2021 publié en 2024 a servi de base dans le cadre de cette modélisation et fournit des informations détaillées sur la population par sexe, tranche d’âge, situation matrimoniale et commune de résidence.

Sexe et tranches d’âge

Les variables majeures que sont le sexe, l’âge et la localisation ont pu être construites en se basant sur les publications de l’INSEE - Population en 2021 : Recensement de la population (hors Corse, Alsace, Moselle, DOM-TOM). Ces données mettent en évidence que près de 57% des seniors de plus de 65 ans sont des femmes. Ce déséquilibre lié au sexe est cohérent avec la différence de longévité hommes-femmes. Ainsi, la variable sexe est construite telle que les femmes représentent environ 57% du portefeuille et les hommes 43%.

Les âges ont été construits conditionnellement aux sexes. Les données de recensement INSEE permettent de croiser les caractéristiques des seniors. L’âge a été modélisé de 65 à 80 ans par tranche de 5 ans (ce qui a nécessité un retraitement de manière à disposer d’une segmentation plus fine), en cohérence avec les données INSEE 2021. La différence d’espérance de vie hommes-femmes se traduit aussi sur cette variable.

TABLE 2.1 – Répartition des seniors par sexe et âge

Âge	Hommes (%)	Femmes (%)
65	30,77	26,31
70	26,57	23,39
75	17,71	16,78
≥80	24,95	33,52

Source : INSEE - Recensement de la population - TD_POP4_2021

Passage d’âges segmentés à des âges entiers Les âges construits suivent les segments présents dans les bases INSEE utilisées. Ces segments permettent de prendre en compte l’âge lors de la construction des différentes variables liées aux seniors, décrites ci-après. Cependant, compte tenu des spécificités du PVH et notamment du risque de longévité, une segmentation plus fine des âges est nécessaire.

En se basant sur la pyramide des âges par sexe de l'INSEE, les segments de 5 ans ont été répartis en âges entiers. Le segment correspondant aux individus âgés de plus de 80 ans a été réparti sur des âges de 80 à 100 ans. Par avis d'expert, les âges supérieurs à 90 ans ont été supprimés. En effet, l'application d'un âge maximal de souscription au PVH permet de garantir une durée minimale d'investissement du capital prêté, et permet de capitaliser sur le principe des intérêts composés. La répartition obtenue est présentée ci-après (c.f. figure 2.5).

Situation matrimoniale

La situation matrimoniale est réduite à deux modalités — seul ou en couple — et est construite conditionnellement au sexe et à l'âge, à partir du recensement INSEE 2021.

TABLE 2.2 – Répartition des seniors par sexe, âge et situation matrimoniale

Âge	Hommes		Femmes	
	En couple (%)	Seul (%)	En couple (%)	Seule (%)
65	23,47	7,06	17,25	9,55
70	20,70	5,58	13,28	9,61
75	14,46	3,85	8,39	8,67
≥80	16,08	8,80	8,34	24,91

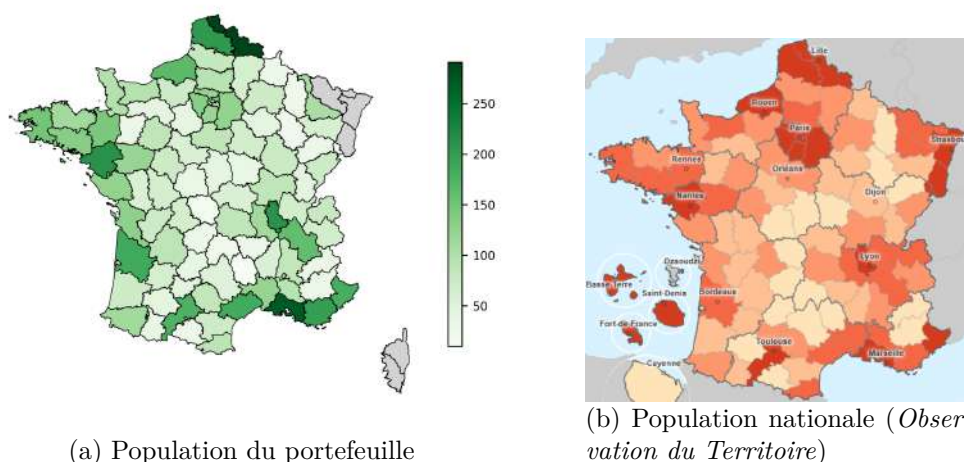
Source : INSEE - Recensement de la population - TD_POP4_2021

La répartition des seniors met elle aussi en évidence la différence d'espérance de vie hommes-femmes. Près d'un quart des femmes de plus de 80 ans sont seules, contre moins de 10% pour les hommes. Tout âge confondu, un quart des hommes sont seuls alors que plus de la moitié des femmes le sont. A noter que pour les couples, l'âge du conjoint ou de la conjointe de l'individu contractant est supposé égal à celui du contractant.

Localisation

La localisation (région, département, commune) a été construite conditionnellement aux caractéristiques individuelles que sont l'âge, le sexe et la situation matrimoniale. Afin de limiter les écarts de distribution, la répartition régionale a d'abord été simulée, avant de procéder à un affinement progressif. Ainsi, la répartition par département a été construite conditionnellement à la région, puis celle par commune en fonction du département.

FIGURE 2.1 – Comparaison territoriale du portefeuille et de la population nationale



A noter que les données utilisées pour créer les variables de localisation ne contiennent aucune information sur les DOM-TOM, la Corse, l'Alsace ou la Moselle. Aucune police n'est donc simulée sur ces départements.

Catégorie socioprofessionnelle

La catégorie socioprofessionnelle (CSP) a été construite conditionnellement au département, en cohérence avec la distribution départementale des CSP des actifs. En effet, la répartition géographique des retraités selon leur ancienne CSP n'étant pas disponible publiquement, cette dernière a été supposée similaire à celle des actifs (INSEE, TCRD_005). De manière à rester cohérent avec les publications INSEE au sujet de la relation ancienne CSP des retraités - patrimoine, seulement 3 CSP différents sont modélisés :

- Retraité ancien agriculteur, artisan, commerçant, chef d'entreprise,
- Retraité ancien cadre ou profession libérale,
- Retraité ancien autre salarié.

Les simplifications utilisées lors de la construction de cette variable créent un biais générationnel. Ce dernier est étudié en section 2.3.4.

2.2.2 Caractéristiques du patrimoine immobilier

Une fois les caractéristiques des emprunteurs construites, la simulation se concentre sur celles de leur bien immobilier. Le patrimoine immobilier est ici supposé constitué à 100% d'un unique bien, dont les attributs sont décrits par les variables détaillées ci-après.

Données utilisées

Les données utilisées ici correspondent aux données Demandes de Valeurs Foncières (DVF) publiées par data.gouv¹. Ce jeu de données est publié annuellement par la direction générale des finances publiques et répertorie les transactions immobilières depuis 2014 sur le territoire métropolitain hors Alsace, Moselle et Mayotte. Les données DVF sont disponibles à la maille commune. Les données DVF sont aussi publiées à la maille [No voie, Code voie, Voie et Code postal] mais uniquement pour la période 2020-2024, ce qui constitue un historique trop limité pour être exploitable dans cette étude.

En se basant sur la base DVF, data.gouv.fr publie des indicateurs immobiliers par commune et par année. Entre 2014 et 2023, plusieurs indicateurs ont été centralisés tels que :

- Le nombre de mutations,
- Le nombre de ventes de maison et d'appartements,
- La proportion de ventes de maisons et d'appartements,
- Le prix moyen des biens vendus,
- Le prix moyen au m² des biens vendus,
- La surface moyenne des biens vendus.

Les mutations exclues pour le calcul de ces indicateurs sont les ventes en lots, les prix inférieurs à 15k€ ou supérieurs à 10m€, les surfaces des appartements et des maisons inférieures à 10m² ou supérieurs à 250m² pour les appartements et 400m² pour les maisons et les prix au m² inférieurs à 330€/m² ou supérieurs à 15k€/m².

1. Le lecteur intéressé peut trouver une description du jeu de données en Annexe c.f. A.2.

Cette étude se base aussi sur les publications de l'INSEE relatives au recensement des résidences principales, et notamment TD_PRINC4_2021 qui décompte les résidences principales par type de logement, nombre de pièces et superficie. Les distributions construites dans cette étude se basent sur ce jeu de données, agrégé par département.

Type du bien

Le type de bien - maison ou appartement - impacte fortement la surface et donc la valeur du bien et est lui-même impacté par le secteur géographique. La répartition des biens immobiliers par type de bien respecte les données INSEE TD_PRINC4_2021, conditionnellement au département.

Surface

La valeur d'un bien immobilier dépend fortement de sa surface. Lors de la construction de ce portefeuille, cette dernière est estimée à partir de la surface moyenne des dernières transactions réalisées dans la commune concernée, selon les données issues de la base DVF.

Un ajustement est ensuite appliqué à la maille département, en fonction du type de bien, afin de traduire les écarts significatifs de superficie moyenne existant entre les maisons et les appartements. Pour prendre en compte cette différence, un coefficient de correction est calculé à partir du recensement INSEE des résidences principales, ventilé par commune, type de bien et tranche de surface. Ce coefficient permet d'ajuster la surface estimée afin de refléter plus fidèlement les spécificités du type de logement considéré.

Ainsi, pour i un couple département \times type, le coefficient appliqué est calculé comme le ratio suivant :

$$Coeff_i = \frac{Surface\ moyenne_i}{Surface\ moyenne\ du\ département}$$

Les ordres de grandeurs de coefficients appliqués sont les suivants :

TABLE 2.3 – Coefficient moyen de correction de la surface par type de bien

Type	Coefficient min	Coefficient max
Maison	105,1 %	154,1 %
Appartement	60,4 %	99,5 %

A titre d'exemple, le département des Deux-Sèvres est le département pour lequel la réduction de la surface est la plus forte sur les appartements, avec un coefficient de 60,42%.

Afin de refléter l'hétérogénéité des biens immobiliers, la surface moyenne issue des données DVF est ajustée individuellement à l'aide d'un facteur multiplicatif aléatoire, noté ε_S . Ce facteur est tiré selon une loi normale centrée en 1 (de manière à ne pas modifier la moyenne), avec un écart-type de 10%, de manière à représenter les différences de configuration, de typologie ou d'aménagement entre logements de même commune. Cette approche permet d'introduire une variabilité réaliste tout en conservant les tendances structurelles observées au niveau agrégé.

Nombre de pièces

Le nombre de pièces de chaque logement n'étant pas directement disponible dans les données DVF, il a été estimé à partir de la surface simulée pour chaque ligne de prêt, en

tenant compte du type de bien.

Les références utilisées proviennent des données INSEE : la surface moyenne d'une maison principale est d'environ 112m² pour 5,6 pièces. Pour les appartements, la surface moyenne est de 65m² pour 4,2 pièces, soit environ 15m² par pièce. Le nombre de pièces de chaque bien immobilier a donc été calculé comme la surface du bien divisée par la surface d'une pièce pour le type de bien en question. Les moyennes issues de la simulation sont présentées ci-dessous et permettent de vérifier la cohérence des résultats simulés avec les données réelles.

TABLE 2.4 – Statistiques simulées par type de bien

Type de bien	Surface moyenne	Nb moy pièces	Surface moy pièce
Appartement	58,23 m ²	3,76	15,48 m ²
Maison	90,12 m ²	4,50	20,03 m ²

Il convient de noter que ce nombre de pièces ne correspond pas aux désignations de type *T2*, *T3*, *T4*, mais constitue une estimation utilisée exclusivement à des fins de modélisation².

Prix au mètre carré

Le prix au mètre carré, de la même manière que la surface du bien, est obtenu en se basant sur les indicateurs liés aux données DVF décrits ci-dessus (c.f. 2.2.2), directement à la maille communale.

Ce premier prix simulé est multiplié par un coefficient aléatoire (centré en 1) dans le but de traduire l'existence d'effets de structures propres au logement autres que la taille (tel que le confort par exemple) ou d'autres effets de structures.

Afin de refléter le fait que les petites surfaces tendent à se valoriser plus cher au m², une corrélation négative ($\rho = -0.7$) a été introduite entre les facteurs aléatoires affectant la surface (ε_S) et le prix au mètre carré ($\varepsilon_{P_{m^2}}$). Les deux perturbations sont modélisées comme un tirage conjoint selon une loi normale bidimensionnelle centrée, avec un écart-type de 10%. Cette corrélation négative permet de refléter l'effet de structure inverse mis en évidence par J. Friggit dans ses travaux : *Le prix d'un logement augmente avec sa surface mais pas proportionnellement : une augmentation de 10% de la surface, toutes choses égales par ailleurs, augmente le prix de vente de 7%³*. Ainsi, en modélisant les perturbations ε_S (surface) et $\varepsilon_{P_{m^2}}$ (prix/m²) avec une corrélation négative, on reproduit cette relation inverse observée empiriquement, tout en générant une hétérogénéité cohérente intra-communale. Le niveau de dispersion retenu (10%) constitue un compromis pertinent entre réalisme statistique et stabilité des résultats agrégés. Ce choix de modélisation permet d'introduire une variabilité individuelle cohérente avec les caractéristiques observées sur le marché, tout en garantissant que les prix au mètre carré simulés restent globalement alignés avec les niveaux de marché.

TABLE 2.5 – Prix au mètre carré par type de bien et département (simulation)

Type de bien	Max		Min		Prix/m ² France
	Dép.	Prix/m ²	Dép.	Prix/m ²	
Appartement	75	10 254	23	912	2 506
Maison	92	7 624	36	1 176	2 351

Dép. 75 Paris, 92 Hauts-de-Seine, 23 Creuse, 36 Indre.

2. Cette variable n'est utilisée que dans le modèle de *backcast*, et n'est à aucun moment comparée à des dénominations commerciales de type *T2*, *T3*, etc.

3. [J. Friggit, 2025].

2.2.3 Intéraction senior - patrimoine immobilier

La simulation du bien immobilier détenu par un senior se doit de dépendre des caractéristiques du senior en question.

Correction liée à la CSP

La CSP permet de classer les individus selon leur activité et leur profession. L'appartenance à ces classes a un impact direct sur le niveau de vie ainsi que sur le patrimoine. En effet, le patrimoine diffère grandement entre agriculteurs exploitants, artisans, commerçants, chefs d'entreprises ou encore employés. Comme le montre le tableau ci-dessous, l'ancienne CSP des retraités implique une grande différence de patrimoine :

TABLE 2.6 – Montant de patrimoine net début par CSP, début 2021

CSP	Moyenne	1er décile	Médiane	9e décile	Coefficient
Retraité anc. agric., artisan, comm., chef d'entr.	478 700	8 800	279 700	925 200	109,1 %
Retraité anc. cadre, prof. libérale	621 500	79 600	422 600	1 142 000	141,7 %
Retraité anc. autre salarié	216 100	3 000	149 200	502 000	49,3 %

Source : [INSEE, 2023]

Un coefficient multiplicatif en lien avec ces différences est calculé comme :

$$Coeff_i = \frac{\text{Patrimoine moyen des retraités de l'ancien CSP}_i}{\text{Patrimoine moyen des retraités}}$$

La racine carrée de ce coefficient est appliquée directement sur la surface et le prix/m² du bien de l'individu en question. Ainsi, ce coefficient impacte de manière égale les deux attributs qui servent à calculer le patrimoine immobilier final (calculé comme le produit de la surface et du prix/m²).

Correction liée à l'âge

De la même manière que pour la CSP, les statistiques produites par l'INSEE mettent en évidence l'impact de l'âge sur le patrimoine :

TABLE 2.7 – Montant de patrimoine net par classe d'âge, début 2021

Classe d'âge	Moyenne	1 ^{er} décile	Médiane	9 ^{ème} décile	Coefficient
Inférieur à 70 ans	361 400	3 700	214 200	842 700	102,5 %
Supérieur à 70 ans	343 800	6 800	209 900	742 500	97,5 %

Source : [INSEE, 2023]

Un retraitement est appliqué, traduisant la différence de patrimoine des seniors de moins et de plus de 70 ans. Ce retraitement est calculé selon la formule

$$Coeff_i = \frac{\text{Patrimoine moyen des retraités de plus (ou moins) de 70 ans}}{\text{Patrimoine moyen des retraités}}$$

Ce coefficient multiplicatif est appliqué au prix/m². Il permet également de refléter, de manière indirecte, la dépréciation potentielle du bien liée à son vieillissement ou à son absence d'entretien, en supposant qu'aucun travaux de rénovation n'est réalisé.

Récapitulatif de construction du patrimoine immobilier

Ainsi, la formule finale de la variable de valeur de patrimoine immobilier est :

$$\text{Patrimoine immobilier} = \text{Surface} \cdot \text{Prix}_{m^2}, \text{ où}$$

$$\text{Surface} = S_{moy\ comm.} \cdot \text{Corr}_{dep.\ et\ type} \cdot \sqrt{\text{Corr}_{csp} \cdot \varepsilon_S}$$

$$\text{Prix}_{m^2} = P_{m^2\ moy\ comm.} \cdot \text{Corr}_{age} \cdot \sqrt{\text{Corr}_{csp} \cdot \varepsilon_{P_{m^2}}}$$

avec :

- $S_{moy\ comm.}$ est la surface moyenne par commune,
- $\text{Corr}_{dep.\ et\ type}$ est le coefficient multiplicatif lié au type de bien par département,
- Corr_{csp} est le coefficient multiplicatif lié à la CSP,
- $P_{m^2\ moy\ comm.}$ est le prix/m² moyen par commune,
- Corr_{age} est le coefficient multiplicatif lié à l'âge,
- ε_S et $\varepsilon_{P_{m^2}}$ sont les coefficients aléatoires multiplicatifs respectivement appliqués à la surface et au prix/m², suivant une loi normale de moyenne 1 et d'écart type 10%. Ils sont corrélés à -0,70.

2.2.4 Retraitements liés aux spécificités du prêt viager hypothécaire

Le portefeuille tel que décrit jusqu'à maintenant est fidèle à la population globale des seniors français. Au regard des spécificités du PVH, le portefeuille s'écarte volontairement des statistiques de la population nationale : certains profils sont en effet plus enclins à monétiser la valeur immobilisée de leur logement. Ainsi, certains profils sont éliminés :

- Patrimoines faibles : L'hypothèse selon laquelle les personnes vivant avec un faible niveau de vie et pour qui le patrimoine immobilier représente une part du patrimoine global importante seraient plus enclins à souscrire un PVH semble cohérente. En effet, la part de patrimoine liquide étant faible pour ces profils, le PVH peut apparaître comme une solution pour dégager des ressources financières en mobilisant leur actif principale : leur résidence principale.
- Patrimoines élevés : Au contraire, en posant l'hypothèse selon laquelle le patrimoine immobilier est proportionnel au patrimoine global d'un ménage (environ 60% selon l'INSEE, pour les ménages situés entre le 4^{ème} et le 9^{ème} décile de niveau de vie), les individus à hauts patrimoines sembleraient moins enclins à souscrire un PVH. En effet, disposant de liquidités suffisantes pour maintenir leur niveau de vie, ces profils auraient moins intérêt à recourir à ce type de financement.
- Patrimoines particulièrement modestes : Il semble aussi cohérent d'éliminer les profils ayant un patrimoine particulièrement faible. En effet, les profils pour lesquels le bien immobilier est très faiblement valorisé (dans une commune peu attractive en terme d'immobilier, avec un faible prix au m² et avec une petite surface) ne sont en capacité de ne libérer qu'un montant très faible via le PVH.

Dans l'objectif de refléter ces éléments dans la structure du portefeuille simulé, un retraitement est effectué de manière à traduire le fait que les seniors à hauts patrimoines et ceux à très faibles patrimoines ne soient pas attirés ou n'aient pas accès au PVH. Les montants des bornes appliqués ont été définis par avis d'expert : les profils aux patrimoines inférieurs à 50k€ ou supérieurs à 500k€ sont retirés du portefeuille établi précédemment. Après retraitement, le portefeuille est composé de 8 187 individus.

2.3 Analyses et regard critique

2.3.1 Analyses globales

L'âge moyen du portefeuille est de 74,5 ans pour une espérance de vie moyenne de l'ordre de 21 ans, niveau élevé qui s'explique par la part importante de couples, la forte proportion de femmes et la sur-pondération des âges 'jeunes' par rapport aux plus âgés. Le patrimoine moyen s'élève à environ 187k€. Ce niveau relativement modeste du patrimoine moyen reflète le profil typique des emprunteurs de prêts viagers hypothécaires : des seniors souvent exclus du crédit classique, disposant d'un patrimoine immobilier parfois hérité, mais d'une capacité d'épargne ou de revenu limité.

TABLE 2.8 – Statistiques patrimoniales

Nombre de contrats	8 187
Patrimoine moyen (€)	187 051
Patrimoine minimum (€)	50 109
Patrimoine maximum (€)	499 965
1 ^{er} quantile	70 181
9 ^{ème} quantile	360 099
Médiane	154 856

En moyenne, les biens en portefeuille ont une surface de 77,8m² pour un prix/m² d'environ 2 500€/m², ce qui situe ces logements dans une gamme intermédiaire, en cohérence avec le profil d'emprunteur observé précédemment. La distinction entre maisons et appartements met en évidence deux segments aux caractéristiques contrastées :

- Les maisons présentent une surface moyenne plus élevée (90,3m²) mais un prix au mètre carré inférieur (2 088€/m²), ce qui reflète une localisation périurbaine ou rurale, où l'espace est plus accessible mais la valorisation moindre,
- À l'inverse, les appartements sont en moyenne plus petits (58,2m²), mais leur prix au mètre carré est sensiblement plus élevé (3 145€/m²), ce qui suggère une plus forte concentration en zones urbaines, voire dans des grandes agglomérations.

TABLE 2.9 – Statistiques des actifs immobiliers par type de bien

Catégorie	Surface moyenne (m ²)	Prix/m ² moyen (€)
Tout type de bien	77,79	2 499,3
Maisons	90,3	2 087,9
Appartements	58,2	3 144,6

2.3.2 Cartes

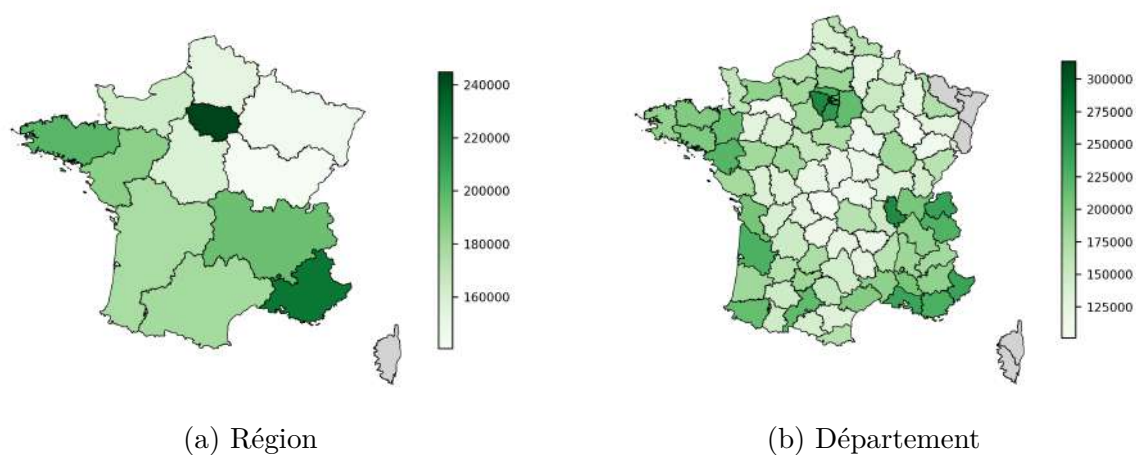
Les cartes représentant le patrimoine moyen par région et par département illustrent clairement les disparités géographiques de valorisation immobilière au sein du portefeuille. Ces disparités sont cohérentes avec la réalité socio-économique française et confirment la représentativité du portefeuille simulé. On observe que :

- L'Île-de-France se distingue nettement, avec un patrimoine moyen largement supérieur à la moyenne nationale, ce qui est attendu compte tenu des prix élevés de l'immobilier en région parisienne,

- Des régions comme la Bretagne et la région PACA apparaissent également bien valorisées, reflétant une attractivité résidentielle et/ou secondaire croissante,
- À l'échelle départementale, on retrouve une survalorisation sur les littoraux atlantiques et méditerranéens, autour de Lyon, dans les Alpes proches de la Suisse, et à Paris,
- À l'opposé, la *diagonale du vide* se dessine nettement : une bande de départements allant du Sud-Ouest au Nord-Est, caractérisée par des valeurs patrimoniales faibles, en lien avec une demande immobilière limitée et une moindre dynamique démographique.

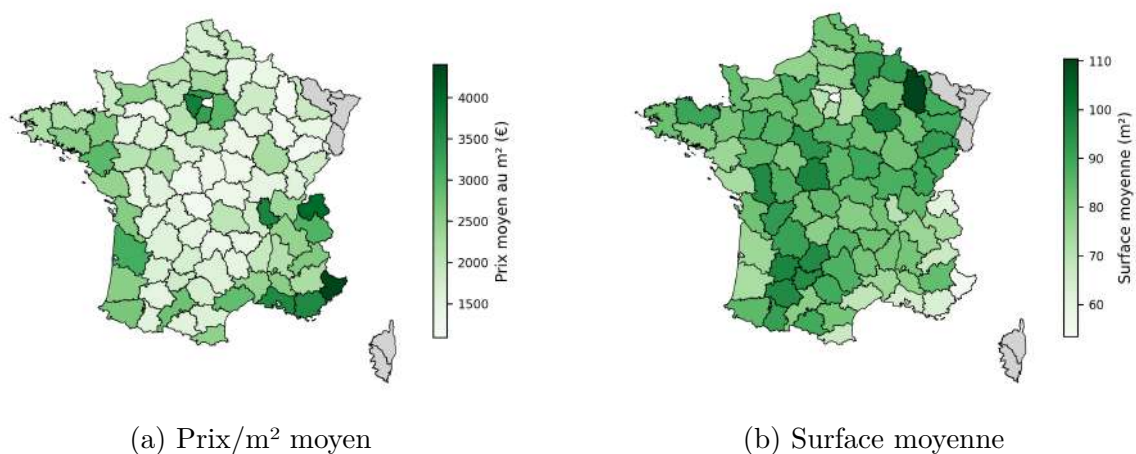
Sur le plan quantitatif, le patrimoine moyen par région varie de 140k€ à 240k€, tandis qu'au niveau départemental, l'écart est encore plus marqué, allant de 100k€ à plus de 300k€.

FIGURE 2.2 – Cartes des patrimoines immobiliers moyens



Les cartes du prix au m² et de la surface moyenne par département (hors Paris, par soucis de lisibilité) restituent les principales tendances spatiales du marché immobilier français. Concernant le prix au m², la carte montre des niveaux allant d'environ 1 000€ à 5 000€, avec des valeurs élevées sur les zones tendues : littoraux (Côte d'Azur, Atlantique), grandes métropoles (Lyon, Bordeaux, Nantes) et zones frontalières attractives (Alsace, Savoie) et des prix faibles dans des départements ruraux ou peu denses, en particulier dans la diagonale du vide. La surface moyenne, qui varie globalement entre 50m² et 100m², suit une tendance inverse au prix/m².

FIGURE 2.3 – Cartes départementales des Prix/m² et Surface moyenne (hors Paris)



2.3.3 Statistiques

L'analyse de la composition du portefeuille selon l'âge fait apparaître une évolution marquée du statut matrimonial des emprunteurs. Sur les tranches d'âges 'jeunes' (60–70 ans), une proportion significative d'emprunteurs en couple est observée. À mesure que l'âge progresse, notamment au-delà de 75 ans, la proportion de personnes seules augmente fortement. Ce constat est cohérent avec l'espérance de vie plus courte des hommes : les femmes étant majoritaires à survivre à leur conjoint, les emprunteurs seuls à âge élevé sont en grande majorité des veuves.

L'analyse géographique des âges ne fait pas apparaître de dynamique départementale évidente, hormis une tendance à des emprunteurs légèrement plus jeunes dans les départements côtiers (notamment côte Atlantique) ou dans le Sud (Occitanie et PACA). Les âges moyens s'échelonnent globalement entre 71 et 77 ans, avec peu de contrastes nets à l'échelle nationale. La pyramide des âges du portefeuille étudié est cohérente avec celle de la population nationale sur la tranche 65–90 ans. On observe une forme globalement décroissante avec l'âge, ce qui est attendu. Par ailleurs, la proportion de femmes y est systématiquement supérieure à celle des hommes, ce qui reflète également une réalité démographique.

Statistiques relatives à la situation matrimoniale, au sexe et à l'âge

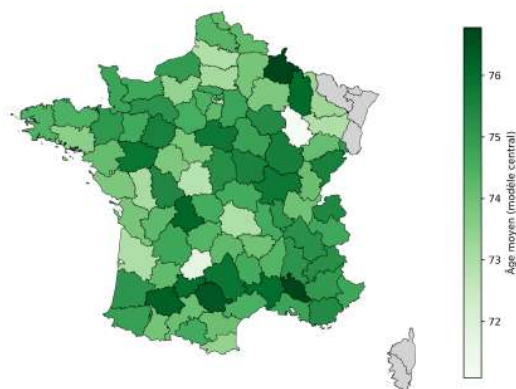


FIGURE 2.4 – Carte des âges moyens

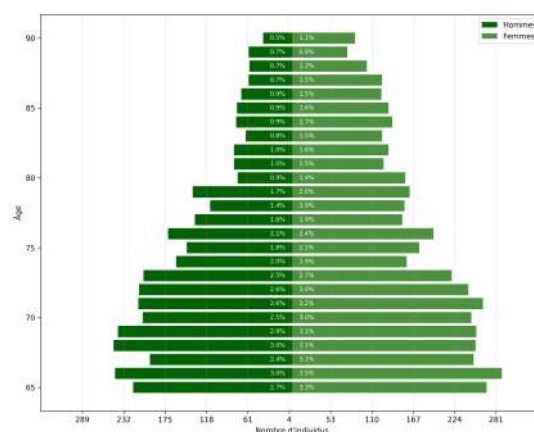


FIGURE 2.5 – Pyramide des âges par sexe

2.3.4 Limites

Le portefeuille créé comporte des biais, dûs aux données et aux hypothèses ou simplifications utilisées.

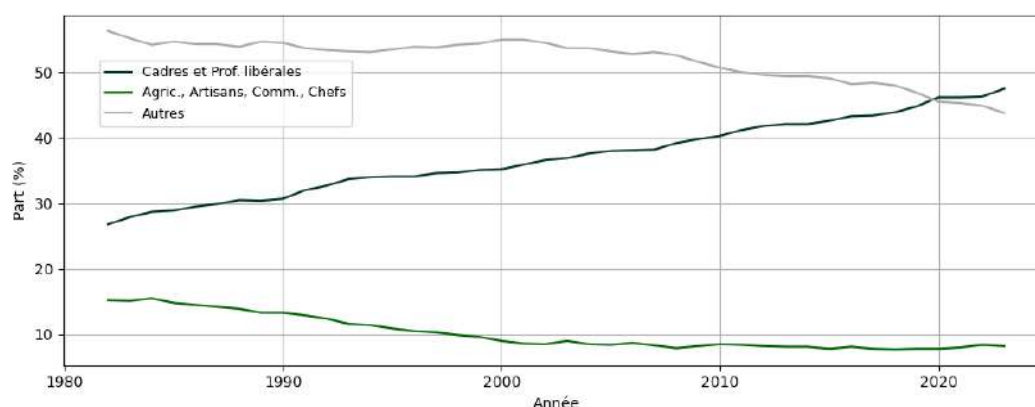
Qualité des données publiques utilisées L'utilisation d'une maille géographique fine est nécessaire pour pouvoir simuler précisément l'évolution des rendements immobiliers dans le temps. À la maille communale, les données publiées par l'INSEE ou par data.gouv ne sont pas très profondes. Par exemple, les moyennes de prix/m² ou de surfaces publiées sont calculées sur un historique de vente court.

Sexe et Situation matrimoniale sans impact sur le patrimoine Aucun impact de la situation matrimoniale ou du sexe sur le patrimoine immobilier n'est présent dans le portefeuille construit. En réalité, la situation matrimoniale et le sexe influencent le patrimoine immobilier. En effet, les couples bénéficient d'économies d'échelle, facilitant l'accès à la propriété contrairement aux personnes seules, souvent limitées par des revenus plus faibles. Les

femmes, avec des salaires en moyenne 15-20% inférieurs d'après l'INSEE, et les familles monoparentales, majoritairement féminines, accumulent moins de patrimoine immobilier. Dans le cadre d'une simulation de portefeuille de PVH, l'absence de prise en compte de la situation matrimoniale et du sexe sur le patrimoine immobilier n'est pas critique, puisque sans doute écrasé par l'effet longévité, mais limite la représentativité des résultats.

Construction des anciennes CSP des retraités Faute de données et statistiques publiques détaillées, la construction des anciennes CSP des emprunteurs seniors repose sur deux hypothèses limitantes (c.f. 2.2.1). La première suppose que les dynamiques patrimoniales des retraités peuvent être regroupées en trois CSP, et la seconde que la répartition géographique des retraités selon leur ancienne CSP est similaire à celle observée chez les actifs actuels. Ces hypothèses doivent être considérées avec précaution. En effet, le passage à la retraite s'accompagne fréquemment de changements de résidence, notamment des centres urbains vers des zones plus rurales ou périurbaines. La structure du marché du travail a aussi connu de profondes transformations, modifiant la répartition des actifs entre les secteurs (développement du tertiaire, baisse du secteur agricole, évolution de l'artisanat, ...). Ces dynamiques introduisent un biais générationnel dans le portefeuille construit. Les données de l'INSEE permettent de mettre en évidence les évolutions mentionnées ci-dessus :

FIGURE 2.6 – Evolution de la répartition de la population active par CSP



Source : Statista, Répartition de la population active en emploi selon la CSP

En supposant la répartition des anciennes CSP des retraités similaires à celles des CSP des actifs, la part d'employés et d'ouvriers est largement sous-estimée, là où celles des anciens cadres ou des professions libérales est surestimée. Or, en moyenne, un ancien cadre possède un patrimoine supérieur au patrimoine moyen à l'inverse des anciens ouvriers ou salariés (c.f. 2.6) : le patrimoine moyen du portefeuille est surestimé.

Cependant, les retraitements effectués dans le but de refléter les caractéristiques d'un portefeuille de PVH ont éliminé les patrimoines élevés (c.f. 2.2.4), majoritairement anciens cadres ou professions libérales. Le trop fort poids des anciens cadres ou personnes ayant exercé une profession libérale dans le portefeuille est donc contrebalancé par les retraitements effectués a posteriori.

Age du conjoint L'hypothèse selon laquelle les couples sont uniquement formés de personnes de mêmes âges est simplificatrice. Elle implique un biais structurel puisque les statistiques montrent que les hommes sont généralement plus vieux que leur conjointe ; et les femmes ont une espérance de vie plus élevée que celle des hommes. Ainsi, l'espérance de

vie des couples dont la personne principale est un homme (considérée comme un groupe au dernier décès) est sous-estimée, et celui dont c'est une femme est sur-estimée.

Concrètement, la sous-estimation de la durée de vie réduit l'impact des intérêts composés et décale la date de remboursement. Le risque de NNEG est aussi sous-estimé en fréquence car la valeur du prêt est sous-estimée étant donnée la date de décès précoce par rapport à la réalité observée. A l'échelle du portefeuille, cette sous-estimation est a minima compensée par la sur-estimation symétrique dû aux couples dont la personne principale est un homme.

Equipondération selon le patrimoine Afin de tenir compte de l'influence du niveau de richesse sur la décision de souscrire à un prêt viager hypothécaire, une borne inférieure et une borne supérieure sur le patrimoine immobilier ont été appliquées sur le premier portefeuille obtenu, fixées respectivement à 50k€ et 500k€. Ce filtre permet d'exclure d'un côté les propriétaires dont la valeur du bien est trop faible pour générer des liquidités significatives, et de l'autre les ménages fortunés, plus susceptibles de recourir à d'autres mécanismes de financement ou de transmission patrimoniale. Toutefois, cette contrainte ne permet pas de traduire la différence de distribution des patrimoines d'un portefeuille de PVH (au sein des bornes) et d'une population de senior 'classique'.

Au-delà de ces seuils, aucune pondération différenciée selon le niveau de patrimoine global des individus n'a été appliquée. Cette approche revient implicitement à considérer que la probabilité de souscrire un prêt viager hypothécaire est identique pour tous les ménages remplissant les critères d'éligibilité retenus. Pourtant, il paraît cohérent que ce type de prêt attire plus particulièrement les ménages modestes, ou qui disposent d'un patrimoine immobilier significatif mais de peu de revenus liquides.

Il aurait été intéressant de pondérer les individus en fonction du poids relatif de leur patrimoine immobilier dans leur patrimoine total. Cette clé de pondération aurait permis de donner plus de poids aux ménages dont l'immobilier représente la composante principale de leur richesse, ceux-là mêmes qui sont les plus incités à mobiliser cette richesse via un PVH. Malheureusement, les données nécessaires à une telle analyse — en particulier la ventilation du patrimoine total par nature d'actifs (immobilier, financier, professionnel, etc.) à un niveau géographique ou socio-économique fin, - ne sont pas disponibles. Faute de telles données, l'hypothèse d'équipondération simple est utilisée, tout en gardant à l'esprit les limites que cela implique en termes de réalisme comportemental.

La construction du portefeuille de référence a permis d'établir une base représentative des caractéristiques démographiques, patrimoniales et géographiques des emprunteurs potentiels. Cette étape constitue le socle indispensable pour appréhender la diversité des profils et identifier les principaux déterminants de risque associés au PVH.

L'enjeu est maintenant de développer un outil de tarification capable de traduire ces caractéristiques en conditions financières cohérentes et prudentes. La construction de cet outil repose sur une articulation entre modélisation (mortalité, rendement immobilier, taux) et paramétrage tarifaire (LTV, taux d'intérêt), afin d'assurer un équilibre technique et une gestion maîtrisée du risque pour le prêteur.

Construction du portefeuille de référence

Cet encadré a pour objectif de résumer de façon synthétique les éléments présentés dans le chapitre, en offrant une vision d'ensemble sans entrer dans le détail. Des encadrés similaires figurent à la fin des chapitres suivants afin de récapituler la démarche et les principaux résultats.

Cette étude repose sur la modélisation d'un portefeuille fictif, conçu pour refléter fidèlement la population des souscripteurs de PVH.

1. Caractéristiques des seniors :

Les individus sont simulés à partir des données INSEE :

- Sexe, âge et situation matrimoniale cohérents avec les statistiques nationales,
- Localisation géographique simulée par maille région, département, commune,
- Ancienne catégorie socioprofessionnelle (CSP) estimée à partir de la répartition des actifs (3 CSP retenues).

2. Caractéristiques immobilières :

Chaque senior est supposé détenir un bien immobilier unique, résidence principale, dont la valeur est proportionnelle à son patrimoine total, simulé à partir des bases DVF et INSEE :

- Type de bien simulé conditionnellement au département,
- Surface et prix/m² estimés à la maille communale, ajustés :
 - selon le type de bien (maille départemental),
 - selon la CSP et l'âge (coefficients multiplicatifs),
 - par une variabilité aléatoire anticorrélée entre l'ajustement de la surface et du prix/m².

Le patrimoine immobilier est le produit de la surface par le prix au m².

Pour mieux refléter les spécificités du produit, les profils à patrimoine très faible (< 50 000€) ou très élevé (> 500 000€) sont exclus.

3. Limites :

La construction de ce portefeuille présente certaines limites, notamment dûes au manque ou au niveau de qualité des données publiques :

- Aucun impact du sexe ou de la situation matrimoniale sur la valeur du bien,
- Les anciennes CSP des retraités sont supposées avoir la même distribution que la distribution des CSP des actifs,
- Aucun ajustement probabiliste n'est fait sur la propension à souscrire selon le niveau de patrimoine (équ pondération implicite),
- L'âge du conjoint est supposé égal à celui du contractant.

Chapitre 3

Outil de tarification en environnement central

3.1 Méthodologies de tarification

La méthodologie de calcul de la LTV optimale s'inspire du mémoire [A. Bruchet, 2014] *Détermination du taux et des montants prêtés pour un prêt viager hypothécaire à taux fixe sous contrainte de rentabilité dans le cadre juridique français*. Cette méthodologie consiste à fixer un taux d'intérêt contractuel donné, puis à déterminer la valeur maximale de LTV acceptable pour le prêteur au regard des attentes de rendement et de l'appétence au risque définies en amont.

Autrement dit, pour un taux $r_{emprunt}$ fixé (ici 6%¹), la LTV optimale est la LTV maximale qui respecte les contraintes posées. Cette approche est particulièrement pertinente dans un contexte de marché où le taux proposé à l'emprunteur est contraint par un fort environnement concurrentiel, et où l'on souhaite maximiser le capital octroyé tout en respectant les objectifs économiques du prêteur.

Ici les contraintes sont définies comme :

- Un seuil sur le rendement annuel moyen,
- Un rendement minimal dans les $\alpha\%$ scénarios les plus adverses (VaR),
- Une probabilité maximale de déclenchement de la NNEG,
- Une valeur maximale de NNEG, exprimée en % du montant prêté.

De manière générale, l'outil simule N (ici 10 000²) scénarios économiques composés de chroniques de taux courts, taux longs, inflation et de rendements immobiliers conjoints. Puis pour chaque prêt, l'outil :

1. Simule N scénarios de mortalité, selon le sexe, l'âge et la situation matrimoniale (un couple est assimilé à un groupe au dernier décès),
2. Projette les flux financiers sous ces scénarios de mortalité \times économique,
3. Calcule les indicateurs pertinents pour une grille de LTV et un taux d'emprunt fixé,
4. Identifie la LTV optimale sous les contraintes énumérées ci-avant.

3.2 Généralités

Dans un PVH, le prêteur agit comme un investisseur confronté à un double objectif : maximiser sa rentabilité tout en maîtrisant les risques associés au prêt ; c'est la notion de couple rendement-risque. Dans cette optique, le couple LTV – Taux d'intérêt constitue le principal levier de tarification du contrat.

1. 6% est cohérent avec les pratiques du marché ; Deux sensibilités ont été réalisées en fixant le taux d'emprunt à 5% et 7% (c.f. 3.6.1).

2. La vitesse de convergence des VAN est étudiée en Annexe.4).

La problématique de détermination du couple LTV - Taux d'intérêt a fait l'objet du mémoire IA déjà cité³. Comme mentionné ci-dessus, la méthodologie utilisée pour le calcul de la LTV optimale décrite ci-après s'inspire des résultats énoncés dans ce mémoire et repose sur la méthode de taux fixé et calcul de LTV maximale sous contrainte.

Le type de PVH étudié ici répond aux caractéristiques suivantes :

- Frais liés aux apporteurs d'affaires supposés égaux à 1%⁴ du montant prêté, payés à t_0 ,
- Frais liés à la gestion du contrat supposés égaux à 0.05% du montant prêté, payés annuellement durant toute la vie du prêt,
- Frais liés à la clôture du contrat supposés égaux à 1% de la valeur de remboursement, payés à la clôture du contrat,
- Aucun autre frais n'est pris en compte (frais d'agence, remise en l'état, frais juridiques...),
- *Lump sum* : 100% du montant prêté est versé à t_0 ,
- Taux d'emprunt fixe,
- Ventes instantanées du bien suite aux décès,
- Les décès ont lieu en fin d'année, par prudence.

3.2.1 Valeur Actuelle Nette

La Valeur Actuelle Nette (VAN) est l'indicateur central de valorisation d'un PVH. Elle représente la somme des flux futurs probabilisés et actualisés. De manière générale :

$$VAN = \sum_{t=0}^{\infty} \left[\frac{p_t \cdot (CF_t - f_t)}{\prod_{i=0}^t (1 + r_i)} \right]$$

Pour le cas du PVH, la VAN représente la différence entre la valeur actuelle des remboursements espérés et le montant prêté initialement sommé des frais actualisés, où :

- p_t est la probabilité que le remboursement intervienne à l'année t ,
- CF_t représente le flux financier à date t , soit $-C_0$ à t_0 et le montant récupéré à la date t (typiquement $\min(CRD_t, V_t)$, avec CRD_t le capital restant dû soit la valeur du prêt hors NNEG et V_t la valeur du bien à t),
- f_t représente les frais imputés au prêteur,
- r_i est le taux d'intérêt à l'instant t_i .

Ici, une VAN positive traduit une espérance de gain positif pour le prêteur, nette de frais⁵.

3.2.2 Taux Effectif Global

Le Taux Effectif Global (TEG) est une mesure normalisée du coût d'un prêt pour l'emprunteur. Il inclut non seulement le taux nominal du prêt, mais aussi les frais annexes (frais de dossier, assurance, notaire). Dans une optique actuarielle, on peut également le considérer comme le taux d'intérêt implicite qui annule la VAN, c'est-à-dire le taux i^* tel que :

3. [A. Bruchet, 2014].

4. Une sensibilité avec des frais à 5% à t_0 a été réalisée, sans choquer les frais à t_s pour $s > 0$, c.f. Annexe.4.

5. En moyenne, à la maille portefeuille, la structure de frais utilisée lors de cette étude et décrite en section précédente implique un montant de frais représentant près de 8% de la valeur du bien à t_0 .

$$C_0 = \sum_{t=1}^{\infty} \frac{p_t \cdot CF_t}{(1 + i^*)^t}$$

Ce taux se doit d'être inférieur au taux d'usure en vigueur à la souscription du prêt.

3.2.3 Value at Risk

La *Value at Risk* (VaR) est un outil central en gestion des risques, largement utilisé dans les domaines bancaires, assurantiels et financiers. Elle mesure, pour une variable aléatoire X représentant une perte (ou ici un rendement), la valeur seuil telle que la probabilité de subir une perte plus importante est inférieure à un niveau donné α :

$$VaR_{\alpha}(X) = \inf \{x \in \mathbb{R} \mid \mathbb{P}(X \leq x) \geq \alpha\}$$

Si $\alpha = 5\%$, la *VaR* correspond au rendement minimal observé dans les pires 5% des cas simulés.

Dans le cadre de la modélisation du prêt viager hypothécaire, il est crucial de ne pas se limiter à une analyse en moyenne, mais d'intégrer explicitement la notion de queue de distribution. Intégrer la VaR permet de quantifier les risques extrêmes associés à différentes métriques clés du contrat, cependant, bien qu'intuitive, la VaR ne donne aucune information sur la gravité des pertes situées au-delà du quantile α .

3.3 Scénarios de mortalités

Le principal risque auquel est exposé un établissement prêteur dans le cadre d'un prêt viager hypothécaire est le risque de longévité. En effet, la durée du prêt, et donc l'accumulation des intérêts potentiellement non remboursés, dépend directement de la date de décès de l'emprunteur (du dernier survivant en cas de couple). Une longévité supérieure à l'hypothèse retenue peut ainsi se traduire par une baisse de rentabilité pour le prêteur.

Afin de quantifier ce risque et d'évaluer l'impact de l'incertitude sur la durée de vie des emprunteurs, une simulation probabiliste de scénarios de mortalité est mise en œuvre. Cette approche permet de capturer la variabilité individuelle des trajectoires de survie tout en conservant un cadre statistique fidèle aux tables de mortalité de référence.

3.3.1 Tables de mortalité et probabilités sur une tête

La modélisation de la durée de vie de l'emprunteur repose sur l'utilisation des tables règlementaires de mortalité prospectives TGH05 (pour les hommes) et TGF05 (pour les femmes). Ces tables sont spécifiquement conçues pour représenter la longévité des populations rentières, qui vivent généralement plus longtemps que la population générale.

Dans un cadre actuariel classique, les probabilités de survie et de décès sont dérivées des fonctions de mortalité sous-jacentes aux tables TGH05 et TGF05. On note l_x l'effectif survivant à l'âge x , ce qui permet de définir les grandeurs suivantes :

- $p_x = \frac{l_{x+1}}{l_x}$: probabilité de survivre de l'âge x à $x + 1$,
- $q_x = 1 - p_x$: probabilité de décéder entre les âges x et $x + 1$,
- ${}_t p_x = \frac{l_{x+t}}{l_x}$: probabilité de survivre t années après l'âge x ,
- ${}_t q_x = 1 - {}_t p_x$: probabilité de décéder avant $x + t$,
- ${}_t | q_x = {}_t p_x \cdot q_{x+t}$: probabilité de décéder à l'âge $x + t$,

— $e_x = \sum_{t=0}^{\infty} {}_t p_x$: espérance de vie résiduelle à l'âge x , dans une approximation discrète.

Dans le cas d'un emprunteur isolé, ces probabilités permettent de simuler la durée de vie résiduelle à partir de la fonction de répartition $F_x(t) = {}_t q_x$, et donc de modéliser la durée du contrat PVH.

Dans le cas d'un emprunteur en couple, la simulation de la mortalité repose sur les principes de la modélisation des groupes au dernier décès. Ce changement de perspective modifie profondément la dynamique du prêt : il ne s'agit plus de simuler la survie d'un individu isolé, mais celle du couple considéré comme une unité, dont la durée de vie correspond à celle du dernier survivant.

Cette hypothèse conduit à une prolongation de la durée effective du prêt, avec des conséquences notables sur la capitalisation des intérêts, le risque de non-recouvrement (NNEG), et in fine sur la tarification du prêt. La section suivante présente le cadre retenu pour la modélisation de la survie conjointe, ainsi que la méthodologie de simulation des trajectoires de vie jusqu'au décès du dernier membre du couple.

3.3.2 Probabilités sur deux têtes

Lorsque le contrat est souscrit par un couple, la durée du prêt est déterminée par le décès du dernier survivant. La durée du contrat est alors modélisée par :

$$T = \max(T_x, T_y)$$

où T_x et T_y sont les durées de vie résiduelles des deux conjoints, âgés respectivement de x et y ans. On note :

- ${}_t p_{\overline{xy}} = \mathbb{P}(T > t)$: probabilité qu'au moins un des deux soit encore en vie à l'instant t ,
- $F_{\overline{xy}}(t) = \mathbb{P}(T \leq t) = 1 - {}_t p_{\overline{xy}}$: fonction de répartition du dernier décès,
- ${}_t |q_{\overline{xy}} = \mathbb{P}(T \in [t, t + 1]) = {}_t p_{\overline{xy}} - {}_{t+1} p_{\overline{xy}}$,
- $e_{\overline{xy}} = \sum_{t=0}^{\infty} {}_t p_{\overline{xy}}$: espérance de vie résiduelle du groupe au dernier décès.

En supposant l'indépendance des durées de vie T_x et T_y , on a :

$${}_t p_{\overline{xy}} = {}_t p_x + {}_t p_y - {}_t p_x \cdot {}_t p_y,$$

c'est-à-dire la probabilité qu'au moins un des deux individus soit encore en vie à l'instant t . Ainsi,

$${}_t |q_{\overline{xy}} = {}_t p_{\overline{xy}} - {}_{t+1} p_{\overline{xy}}$$

où :

- ${}_t p_{\overline{xy}}$: probabilité qu'au moins un des deux soit encore en vie à l'instant t (dernier décès après t),
- ${}_t |q_{\overline{xy}}$: probabilité que le dernier décès survienne dans l'intervalle $[t, t + 1]$.

Après avoir posé le cadre théorique de la modélisation de la mortalité, il convient désormais de détailler la mise en œuvre pratique des simulations de trajectoires de vie. Cette étape opérationnelle permet de traduire les hypothèses démographiques en scénarios individuels de longévité, base indispensable au calcul des flux financiers futurs, de la valorisation actuarielle des contrats, et de l'évaluation des risques associés. La section suivante présente la méthodologie retenue pour la simulation des décès, ainsi que les hypothèses techniques et les choix de modélisation sous-jacents.

3.3.3 Simulations

La génération de la durée de vie repose sur un tirage aléatoire selon la fonction de répartition de la variable aléatoire représentant l'âge au décès. À partir des tables de mortalité TGH05/TGF05, on construit la fonction de répartition empirique $F(t) = \mathbb{P}(T \leq t)$ pour les âges entiers, puis on procède à un tirage par inversion : on simule une variable uniforme $U \sim \mathcal{U}(0, 1)$, et on en déduit la durée de vie $T = F^{-1}(U)$. Cette méthode permet de reproduire la variabilité naturelle des durées de vie et d'intégrer l'aléa de longévité dans les scénarios utilisés pour les calculs décrits ci-après. Lors de cette étude, 10 000 tirages sont réalisés par individu. Ce cadre de simulation est essentiel pour évaluer correctement la valeur actuelle nette du contrat, le risque de longévité, et la probabilité d'activation de la garantie de non-dépassement.

Une fois les trajectoires de mortalité simulées, permettant de déterminer la durée effective de chaque prêt, il est nécessaire de compléter la modélisation par des scénarios économiques cohérents, afin d'évaluer les flux financiers futurs. En effet, les paramètres macroéconomiques tels que les taux d'intérêt, l'inflation ou encore l'évolution du marché immobilier ont un impact direct sur la capitalisation des intérêts, la valorisation du bien au moment du remboursement, ainsi que sur les indicateurs de rentabilité du portefeuille. La section suivante présente la génération des scénarios économiques, ainsi que les modèles retenus pour projeter l'environnement financier et immobilier sur lequel la tarification réalisée.

3.4 Scénarios économiques

La valorisation d'un PVH repose sur des projections à long terme. À ce titre, les rendements immobiliers et les taux d'actualisation constituent les variables économiques utilisées dans le but de capter l'incertitude liée à la garantie de non-dépassement. Lors de ce mémoire, une variante du modèle Ahlgrim est développée (sans modélisation des actifs de type action), en se basant sur [Ahlgrim et al., 2005], [F. Planchet et al., 2012] et [K. Armel, 2025].

3.4.1 Modèle d'Ahlgrim

Le modèle d'Ahlgrim a été construit par la *Casualty Actuarial Society* et la *Society of Actuaries*. Il permet de simuler les variables énumérées ci-dessous de manière conjointe afin de refléter la dépendance statistique entre les séries, en monde réel :

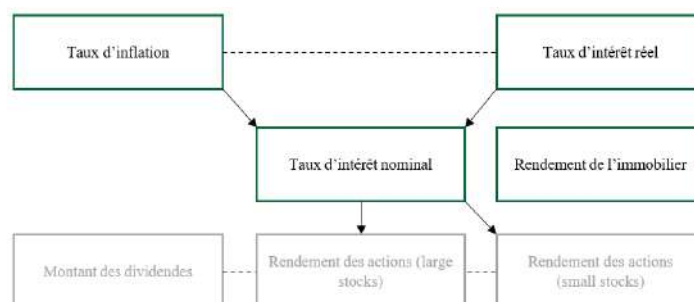
- l'inflation,
- le taux d'intérêt court terme,
- le taux d'intérêt long terme,
- le rendement action,
- le rendement immobilier.

L'objectif de ce modèle n'est pas de prévoir précisément chaque variable, mais de générer un ensemble cohérent de trajectoires économiques possibles. Au sein de ce modèle, une certaine importance est portée à l'interdépendance des variables simulées ; elle est traduite par une structure de corrélation.

Dans le cadre spécifique de la valorisation d'un portefeuille de prêts viagers hypothécaires, l'intégration des actions dans les scénarios économiques n'est pas jugée nécessaire. En effet, les flux associés à ces prêts ne sont pas directement sensibles aux marchés financiers actions.

Dans les travaux [F. Planchet et al., 2012], la structure de ce modèle est représentée par le schéma ci-dessous, ici privé de la partie liée aux actions :

FIGURE 3.1 – Structure du modèle Ahlgrim - sans action



Le modèle d'Ahlgrim est utilisé dans le cadre d'analyses de rentabilité, de solvabilité ou de projection d'engagements, en monde réel. La construction des scénarios économiques repose sur une méthodologie en plusieurs étapes structurées :

1. Estimation des dynamiques individuelles : pour chaque variable X , une régression autorégressive est réalisée sur les données historiques. Les résidus ε_t sont extraits à partir des écarts entre la valeur observée et la valeur prédite.
2. Projection des résidus, selon une structure de dépendance basée sur la corrélation : la matrice empirique de corrélation entre les résidus est calculée ; des vecteurs de bruit standard $Z \sim \mathcal{N}(0, I)$ sont générés, puis transformés pour respecter la structure de dépendance. Cela est fait en appliquant une décomposition de Cholesky à la matrice de corrélation.
3. Simulation des trajectoires stochastiques : la dynamique discrétisée du modèle de Hull and White à deux facteurs est utilisée pour les taux, de Vasicek pour l'inflation et l'immobilier. Cette étape est répétée pour chaque scénario, en injectant les résidus simulés corrélés.

La modélisation propre à chaque variable est détaillée par la suite, en commençant par les rendements immobiliers, qui constituent l'élément central de ce mémoire.

3.4.2 Modélisation des rendements immobiliers

La dynamique de l'immobilier est modélisée à travers le taux de variation logarithmique de la valeur du bien, notée R_t^{immo} , défini comme :

$$R_{t+1}^{\text{immo}} = \ln \left(\frac{V_{t+1}}{V_t} \right)$$

où V_t est la valeur du bien à la date t .

Les valeurs des biens sont ensuite obtenues par accumulation exponentielle des rendements simulés :

$$V_{t+1} = V_t \cdot e^{R_{t+1}^{\text{immo}}}$$

Étant définis comme les taux de variation logarithmiques de la valeur du bien, les rendements immobiliers doivent être modélisés selon une dynamique temporelle cohérente. À cette fin, la modélisation retenue s'appuie sur le cadre stochastique proposé par Ahlgrim⁶.

6. [Ahlgrim et al., 2005].

Modèle et calibrage

Les log-rendements immobiliers sont modélisés par un processus de Vasicek à un facteur, selon la dynamique continue :

$$dx_t = \theta \cdot (\mu - x_t) dt + \sigma dW_t$$

où x_t représente la variable modélisée à l'instant t et :

- θ est la vitesse de retour à la moyenne,
- μ est la valeur d'équilibre à long terme,
- σ est la volatilité,
- W_t est un mouvement brownien standard.

Avec un pas annuel, le processus peut être discrétisé selon la relation suivante :

$$x_{t+1} = x_t \cdot e^{-\theta_x} + \mu_x \cdot (1 - e^{-\theta_x}) + \sigma_x \cdot \sqrt{\frac{1 - e^{-2\theta_x}}{2\theta_x}} \cdot \varepsilon_{x,t} = \alpha_x \cdot x_t + \beta_x + \gamma_x \cdot \varepsilon_{x,t}$$

où $\varepsilon_{x,t} \sim \mathcal{N}(0, 1)$, i.i.d..

Les relations entre les coefficients de régression et les paramètres du modèle continu sont les suivantes :

$$\theta_x = -\ln(\alpha_x) \quad (3.1)$$

$$\mu_x = \frac{\alpha_x}{1 - \beta_x} \quad (3.2)$$

$$\sigma_x = \gamma_x \cdot \sqrt{\frac{2\theta_x}{1 - \exp(-2\theta_x)}} \quad (3.3)$$

La projection des variables repose sur l'estimation de ces paramètres, qui caractérisent les dynamiques de retour à la moyenne, de niveau de long terme et de volatilité. Le calibrage est effectué par une régression linéaire des séries historiques, selon la forme autorégressive suivante :

$$x_{t+1} = \alpha \cdot x_t + \beta_x + \gamma_x \cdot \varepsilon_{x,t}$$

Une fois les coefficients α, β, γ estimés, les paramètres θ, μ, σ du processus de Vasicek étudié sont déduits à l'aide des relations ci-dessus.

La procédure de calibrage adoptée suit donc les étapes suivantes :

1. Estimer les coefficients α, β et l'écart-type des résidus γ à l'aide d'une régression linéaire des log-rendements annuels,
2. Calculer les paramètres du processus continu de Vasicek (θ, μ, σ) à partir des équations (3.1)–(3.3),
3. Utiliser ces paramètres pour simuler les trajectoires futures des variables dans le cadre du générateur de scénarios économiques.

Cette approche permet de garantir la cohérence entre les dynamiques historiques observées et les trajectoires simulées sur des horizons longs, tout en assurant une interprétation économique des paramètres du modèle : retour à l'équilibre, volatilité macroéconomique et niveaux de long terme.

Le calibrage des paramètres du processus de Vasicek se fait à partir de données représentatives. La section suivante décrit ainsi les sources de données utilisées, ainsi que la période d'observation retenue.

Données

Les données utilisées pour modéliser les rendements immobiliers proviennent de deux sources. Jusqu'en 2020, les indices sont issus des données [J. Friggit, 2025], qui fournissent des séries longues et homogènes sur les prix de l'immobilier en France. À partir de 2020, la série est prolongée à l'aide de l'indice [INSEE, 2024] des prix des logements anciens. Les scénarios économiques simulés présentés ci-après sont construits à partir de données annuelles, en s'appuyant sur un historique long.

Le choix de la période d'observation retenue pour le calibrage des modèles économiques constitue un enjeu crucial, dans la mesure où il influence fortement les paramètres estimés et, par conséquent, les trajectoires simulées. Il est essentiel que l'historique utilisé soit cohérent avec l'horizon de projection considéré. En particulier, dans le cadre de projections de long terme, un historique trop court peut s'avérer limitant, notamment en raison du risque d'absence de certains événements extrêmes ou cycles économiques significatifs. À l'inverse, un historique trop long peut poser problème si les structures économiques fondamentales ont évolué de manière significative sur la période.

Plusieurs profondeurs d'historiques ont été testées, notamment entre 1965 et 1980, et une profondeur de 45 années, entre 1980 et 2024 a été retenue. Lorsque l'on intègre des données plus anciennes, on inclut des périodes économiques, sociales et structurelles parfois très différentes du contexte actuel. Dans ce cas précis, l'intégration d'années antérieures à 1980 a pour effet d'augmenter sensiblement la moyenne long terme du rendement immobilier, la portant à un niveau autour de 5%, ce qui implique par la suite une rentabilité jugée trop optimiste. De plus, cela ne permet pas de refléter fidèlement les dynamiques économiques récentes ni d'assurer la prudence nécessaire dans les projections. Ainsi, par souci de prudence et afin de garantir que la moyenne long terme reste représentative des tendances structurelles observées sur les dernières décennies sans être biaisée par des périodes atypiques ou moins pertinentes, la profondeur de l'historique a été limitée à 1980.

Après avoir modélisé la dynamique des rendements immobiliers, il convient d'introduire la variable macroéconomique centrale qu'est l'inflation, qui intervient dans la revalorisation des flux financiers en valeur réelle. À l'instar de la modélisation des rendements immobiliers, l'inflation est représentée à travers un processus de Vasicek, calibré selon la même approche.

3.4.3 Modélisation de l'inflation

L'inflation désigne l'augmentation générale et durable des prix des biens et services dans une économie. Elle est généralement mesurée à l'aide de l'indice des prix à la consommation (IPC), qui reflète l'évolution du coût d'un panier représentatif de biens consommés par les ménages. Une inflation modérée est considérée comme un signe de bonne santé économique, mais sa variabilité a un impact direct sur la valeur réelle des flux financiers futurs, en influençant à la fois les taux d'intérêt réels et le pouvoir d'achat des montants remboursés. Sa modélisation est donc essentielle pour évaluer de manière cohérente la rentabilité et le risque associés aux PVH, notamment dans une perspective de projection de long terme.

Modèle, calibrage et données

La dynamique de l'inflation est modélisée à travers le taux de variation logarithmique de l'IPC, notée R_t^{infl} , défini comme :

$$R_{t+1}^{\text{infl}} = \ln \left(\frac{IPC_{t+1}}{IPC_t} \right)$$

En termes de modélisation et de calibrage, la même méthodologie que pour l'immobilier a été appliquée : un modèle de Vasicek est calibré à l'aide de moindres carrés ordinaires (MCO) pour estimer les paramètres du modèle d'inflation. Les données utilisées proviennent également de deux sources différentes. Jusqu'en 2021, les indices sont issus de la base Friggit et à partir de 2021, la série est prolongée avec les données de l'INSEE sur l'indice des prix à la consommation hors tabac. Le paramétrage de l'inflation, comme les autres variables, se base sur un historique remontant jusqu'à 1980.

Après la modélisation de l'inflation, il convient d'intégrer les taux d'intérêt nominaux, qui jouent un rôle clé dans la valorisation financière des flux futurs et la capitalisation des intérêts.

3.4.4 Modélisation conjointe des taux court et long terme

Les taux courts et taux longs sont modélisés selon l'approche multi-factorielle de Hull and White, en cohérence avec le modèle d'Ahlgrim, faisant tendre les taux courts vers les taux long.

Modèles et calibrage

Le modèle d'Ahlgrim propose une modélisation cohérente des taux d'intérêt à court et long termes en capturant leur dynamique propre tout en tenant compte de leurs interactions structurelles. L'approche repose sur l'hypothèse que ces taux évoluent dans le monde réel selon le modèle de Hull and White à deux facteurs. Le taux d'intérêt court terme, noté r_t , et le taux d'intérêt long terme, noté l_t , réels, sont modélisés selon la forme :

$$\begin{aligned} dl_t &= \theta_l \cdot (\mu_l - l_t) dt + \sigma_l dW_t^l \\ dr_t &= \theta_r \cdot (l_t - r_t) dt + \sigma_r dW_t^r \end{aligned}$$

où :

- r_t est le taux court,
- l_t est le taux long,
- θ_x est la vitesse de retour à la moyenne de la variable x ,
- μ_l est le niveau d'équilibre de long terme de la variable l ,
- σ_x est la volatilité de la variable x ,
- W_t^x est un mouvement brownien standard propre à la variable x , avec dW_t^r et dW_t^l définis sur le même espace de probabilité et supposés corrélés avec un coefficient $\rho \in [-1, 1]$ non nul.

Une telle structure implique que le taux long suit une dynamique de retour à la moyenne classique, tandis que le taux court converge vers le taux long, ce dernier jouant alors le rôle de niveau de long terme dans le processus suivi par le taux court. Cette structure hiérarchique

reflète l'idée selon laquelle les taux courts sont plus volatils et influencés par la politique monétaire de court terme, tandis que les taux longs captent des anticipations structurelles de marché.

Avec un pas annuel, le modèle peut être discrétisé selon la relation suivante :

$$r_{t+1} = r_t \cdot e^{-\theta_r} + l_x \cdot (1 - e^{-\theta_r}) + \sigma_r \cdot \sqrt{\frac{1 - e^{-2\theta_r}}{2\theta_r}} \cdot \varepsilon_{r,t}$$

$$l_{t+1} = l_t \cdot e^{-\theta_l} + \mu_l \cdot (1 - e^{-\theta_l}) + \sigma_l \cdot \sqrt{\frac{1 - e^{-2\theta_l}}{2\theta_l}} \cdot \varepsilon_{l,t}$$

avec $(\varepsilon_{r,t})_{t \geq 0}$ et $(\varepsilon_{l,t})_{t \geq 0}$ deux suites i.i.d. de variables $\mathcal{N}(0, 1)$, et pour tout t , $\mathbb{E}[\varepsilon_{r,t} \varepsilon_{l,t}] = \rho$, où $\rho \in [-1, 1]$.

Le modèle d'Ahlgrim se base sur les Doubles Moindres Carrés pour procéder à l'estimation des paramètres. Cette procédure se décompose en deux étapes :

— **Taux long :**

$$l_{t+1} = \alpha_l \cdot l_t + \beta_l + \gamma_l \cdot \varepsilon_{l,t}$$

La vitesse de retour à la moyenne, le niveau d'équilibre de long terme et l'écart type de l'erreur du modèle de taux à long terme sont calculés sous les mêmes égalités que les équations 3.1, 3.2 et 3.3 énoncées plus haut,

— **Taux court :**

$$r_{t+1} = \alpha_r \cdot l_t + \beta_r \cdot r_t + \gamma_r \cdot \varepsilon_{r,t}$$

La présence de deux variables endogènes modifie l'approche à utiliser. Les coefficients sont :

$$\beta_r = e^{-\theta_r} \tag{3.4}$$

$$\alpha_r = 1 - e^{-\theta_r} = 1 - \beta_r \tag{3.5}$$

L'estimation des paramètres propres aux taux courts réels se base alors sur :

$$\Delta r_{t+1} = \alpha_r \cdot (l_t - r_t) + \varepsilon_{r,t}$$

Les taux réels ainsi modélisés constituent la base de projection des trajectoires. Pour les besoins de valorisation financière en monnaie courante, il est toutefois nécessaire de convertir ces taux réels en taux nominaux, en tenant compte de l'inflation anticipée. Cette conversion s'appuie sur la formule de Fisher, qui relie taux réel, taux d'inflation et taux nominal, comme détaillé ci-après.

Taux nominaux et réels

Le taux nominal correspond au taux d'intérêt affiché par un contrat de prêt ou un investissement, sans tenir compte de l'inflation. En revanche, le taux réel mesure le rendement effectif en prenant en compte l'évolution des prix. Il représente donc le pouvoir d'achat réel généré ou perdu par l'investisseur ou l'emprunteur.

La relation entre les taux nominaux i , les taux réels r et le taux d'inflation π est donnée par la relation de Fisher :

$$1 + i = (1 + r)(1 + \pi)$$

La transformation des taux réels en taux nominaux via la formule de Fisher permet ainsi d'obtenir des trajectoires cohérentes de taux d'intérêt exprimés en monnaie courante, nécessaires pour la valorisation des flux économiques projetés.

La section suivante décrit les sources de données et l'historique utilisés pour la modélisation de ces variables et les choix méthodologiques retenus pour assurer la robustesse du calibrage.

Données

Pour le taux long terme, le taux de référence retenu est l'OAT 10 ans. Pour le court terme, les taux interbancaires ont été utilisés (Euribor 3 mois). Comme pour les rendements immobiliers, le choix de la période d'observation est important, car il conditionne le calibrage des modèles et l'interprétabilité des résultats. La profondeur d'historique choisie correspond à celle utilisée pour l'immobilier, soit 1980-2024, par souci de cohérence.

Une fois l'inflation et les taux d'intérêt à court et long termes modélisés, il convient de définir les taux d'actualisation utilisés pour la valorisation des flux futurs. Ces taux constituent un élément central du modèle économique, dans la mesure où ils permettent de ramener l'ensemble des flux à leur équivalent en monnaie courante. La section suivante précise la méthodologie retenue pour la construction et l'utilisation des taux d'actualisation dans le cadre des simulations.

3.4.5 Taux zéro-coupon et construction des taux d'actualisation

La valorisation de flux financiers futurs repose sur un principe fondamental en finance : la valeur temps de l'argent. Un euro reçu aujourd'hui n'a pas la même valeur qu'un euro reçu dans plusieurs années, en raison de l'incertitude, de l'inflation et des opportunités d'investissement. Pour intégrer cet effet temporel dans les simulations, il est nécessaire de définir une structure de taux d'actualisation permettant de convertir tout flux futur en valeur actuelle. La construction de cette courbe repose sur les taux zéro-coupon, qui constituent la référence pour l'actualisation sans paiements intermédiaires.

Définition d'un taux zéro-coupon

Le taux zéro-coupon (ZC) est le taux d'intérêt d'un instrument financier qui ne verse aucun flux intermédiaire, et qui ne paye qu'un montant unique d'une unité à l'échéance. Pour une maturité M , le prix d'un zéro-coupon est noté $P(0, M)$. Le taux zéro-coupon correspondant, appelé taux *spot*, est défini par :

$$P(0, M) = \exp(-M \cdot ZC(0, M))$$

où $ZC(0, M)$ est le taux zéro-coupon annualisé en continu. Ainsi, ce taux représente le taux d'actualisation constant qui permet de ramener un flux unique à l'instant initial, sans paiements intermédiaires.

Dans le cadre de la valorisation des flux futurs, il est essentiel de disposer d'une courbe permettant d'actualiser les paiements à toutes les échéances pertinentes. Le modèle d'Ahlgrim étant formulé dans le monde réel et sur la base de variables macroéconomiques simulées, la construction des taux zéro-coupon repose sur les taux d'intérêt générés dans chaque scénario.

Construction des taux zéro-coupon

Après avoir calibré les dynamiques des taux d'intérêt et de l'inflation, on cherche à construire, pour chaque scénario, les courbes zéro-coupon nécessaires à l'actualisation des flux futurs. Ces courbes donnent, à chaque échéance t , le prix d'un instrument sans flux intermédiaire permettant de ramener un flux unique à sa valeur actuelle simulée.

Le modèle de Hull and White à deux facteurs offre une expression analytique fermée pour le prix d'un zéro-coupon (c.f. [K. Armel, 2025], [J. Hibbert et al., 2001]). Pour une maturité M (exprimée en années) et un taux court initial r_0 , ce prix est donné par :

$$P(0, M) = \exp(A(M) - B(M) \cdot r_0)$$

où :

- $B(M)$ dépend uniquement de la vitesse de retour à la moyenne θ ,
- $A(M)$ dépend des trois paramètres du modèle : θ , μ , et σ .

Les expressions fermées de $B(M)$ et $A(M)$ sont énoncées dans [K. Armel, 2025].

Dans un modèle de type Ahlgrim, les composantes réelles et nominales du taux d'intérêt sont modélisées séparément. Plus précisément, le taux d'intérêt réel r_t et le taux d'inflation π_t évoluent chacun selon un processus stochastique propre, calibré sur données historiques. Pour projeter des flux économiques futurs, deux courbes de taux zéro-coupon sont alors construites à chaque date t :

- La courbe réelle $P_{\text{réel}}(t, M)$, qui correspond à la valeur actualisée d'un euro réel versé à l'échéance M , est construite à partir des taux d'intérêt réels simulés ; les coefficients $A_{\text{réel}}, B_{\text{réel}}$ sont calculés à partir des paramètres calibrés sur le taux réel (θ, μ, σ).
- La courbe d'inflation $P_{\text{infl}}(t, M)$, qui reflète la revalorisation des prix attendue sur la période $[t, M]$, est calculée à partir des taux d'inflation projetés ; de la même manière que $P_{\text{réel}}(t, M)$, les coefficients $A_{\text{infl}}, B_{\text{infl}}$ sont issus des paramètres de l'inflation.

Ces deux composantes sont ensuite combinées pour actualiser les flux nominaux, revalorisés par l'inflation. Le prix d'un zéro-coupon nominal résulte alors du produit :

$$P_{\text{nom}}(t, M) = P_{\text{réel}}(t, M) \cdot P_{\text{infl}}(t, M)$$

Ce produit permet d'obtenir une courbe d'actualisation nominale cohérente avec les dynamiques simulées. Ces calculs sont effectués pour chaque année de projection, permettant de simuler les ZC pour tous les scénarios. Enfin, les déflateurs cumulés $D(0, t)$ sont construits par un produit successif des ZC à un an glissant :

$$D(0, t) = \prod_{u=1}^t P_{\text{nom}}(u-1, 1)$$

Chaque facteur $P_{\text{nom}}(u-1, 1)$ est calculé à partir des taux simulés au pas $u-1$, garantissant la cohérence dynamique de l'actualisation.

Le cadre théorique étant désormais défini, la section suivante s'attache à l'analyse critique des scénarios simulés ainsi qu'à la validation des trajectoires économiques générées.

3.4.6 Simulations

Paramètres

La Table 3.1 présente les paramètres calibrés des quatre variables économiques modélisées. Ces paramètres – vitesse de retour à la moyenne (θ), niveau de long terme (μ) et volatilité (σ) – ont été estimés à partir des données historiques selon la méthodologie décrite en section 3.4. Le lecteur intéressé pourra trouver en annexe les résultats détaillés des tests statistiques⁷, globalement satisfaisants, associés à ces calibrages (c.f. Annexe.5).

TABLE 3.1 – Paramètres calibrés des modèles économiques

Paramètres	θ	μ	σ	X_0
Taux long	0,095	0,030	0,011	2,01 %
Taux court	0,054	0,030	0,012	1,55 %
Inflation	0,216	0,016	0,012	1,17 %
Immobilier	0,417	0,033	0,043	-0,94 %

Afin de garantir des scénarios cohérents avec les dynamiques réelles, toutes les simulations démarrent au même niveau initial (X_0), correspondant aux valeurs observées au 31.12.2024. Compte tenu de la dynamique négative de l'immobilier au cours des dernières années, ce niveau de départ est fixé à environ -1% , ce qui correspond à l'évolution moyenne des prix de l'immobilier ancien en France sur l'année 2024 selon les données de l'INSEE.

Il peut paraître étonnant que les taux longs et courts aient des valeurs de niveau de long terme (μ) similaires, alors que la théorie économique suggère généralement un écart structurel entre ces taux. Cela s'explique par la construction même du modèle d'Ahlgrim, où le taux long joue le rôle de niveau de long terme dans le processus suivi par le taux court. Plus précisément, le modèle suppose que le taux court r_t suit un processus d'Ornstein-Uhlenbeck, dont le niveau de long terme est le taux long l_t lui-même modélisé selon ce processus, avec :

$$dr_t = \theta_r \cdot (l_t - r_t) dt + \sigma_r dW_t^r$$

Ainsi, en régime stationnaire, lorsque l_t fluctue autour de sa propre moyenne μ_l , le taux court r_t tend lui aussi à converger vers cette même valeur moyenne, du fait de la structure du processus. Ce comportement est conforme à la théorie du taux d'intérêt, qui prévoit une convergence du taux court vers le taux long dans les phases d'équilibre monétaire. Ainsi, les deux paramètres μ estimés à partir de leurs dynamiques respectives sont naturellement proches, sans que cela contredise la théorie, mais au contraire car :

- La convergence du taux court vers le taux long est un effet attendu en régime stationnaire,
- Sur des horizons longs, la différence entre les taux (la prime de terme) devient faible (notamment dans la cas d'une inflation maîtrisée). Les taux courts sont fortement influencés par la politique monétaire et tendent à converger à long terme vers les taux longs en période d'équilibre.

Enfin, ces hypothèses de convergence jouent un rôle de stabilisation des projections de long terme. En limitant les écarts persistants entre taux courts et longs, elles évitent la dérive excessive des scénarios d'actualisation ou de réinvestissement, ce qui est souhaitable dans un cadre prospectif prudent.

7. Les indicateurs usuels d'évaluation de la qualité d'ajustement sont calculés et analysés : le R^2 ajusté, le test de Fisher, la moyenne des résidus, les tests de Breusch-Godfrey et d'hétéroscédasticité conditionnelle (ARCH(1)), ainsi que le test de Jarque-Bera. Cette démarche s'inscrit dans la lignée des travaux appliquant la méthodologie d'Ahlgrim, et vise à valider le paramétrage retenu.

Corrélation et structure de dépendance

Afin de générer des trajectoires économiques cohérentes dans le temps et entre les variables, la projection des variables macroéconomiques repose non seulement sur leurs dynamiques propres (issues du calibrage de modèles), mais également sur la prise en compte de la structure de dépendance entre les chocs aléatoires. Concrètement, les résidus simulés à chaque date, pour chaque variable (taux d'intérêt court, long, inflation, rendement immobilier), sont supposés corrélés entre eux, afin de refléter les interdépendances structurelles observées dans les données historiques.

Cette dépendance est modélisée à l'aide d'une décomposition de Cholesky de la matrice de corrélation des résidus estimée sur les données historiques des quatre variables modélisées. Cette méthode consiste à factoriser une matrice symétrique définie positive $\Sigma \in \mathbb{R}^{4 \times 4}$ sous la forme :

$$\Sigma = LL^\top$$

où L est une matrice triangulaire inférieure. Cette décomposition est unique et garantit que les produits vectoriels conservent la structure de corrélation initiale. On génère à chaque pas de temps un vecteur $\varepsilon_t \sim \mathcal{N}(0, I_4)$, composé de chocs indépendants et centrés réduits. Ces chocs sont ensuite transformés en un vecteur de chocs corrélés $\tilde{\varepsilon}_t \in \mathbb{R}^4$ via l'opération linéaire suivante :

$$\tilde{\varepsilon}_t = L \cdot \varepsilon_t, \quad \text{par construction : } \mathbb{E}[\tilde{\varepsilon}_t] = 0 \quad \text{et} \quad \text{Var}(\tilde{\varepsilon}_t) = L \cdot \text{Var}(\varepsilon_t) \cdot L^\top = LL^\top = \Sigma$$

Cela garantit que les simulations reproduisent fidèlement la dépendance empirique observée entre les variables. Cette approche est largement utilisée en finance, actuariat et modélisation économique.

La table 3.2 présente les matrices de corrélations historiques et simulées. La comparaison des matrices de corrélation des résidus permet d'évaluer dans quelle mesure le simulateur reproduit fidèlement les dépendances statistiques résiduelles observées dans les données historiques. Globalement, la structure simulée reste cohérente en niveau et en signe avec la matrice empirique. Les écarts sont modérés, et aucun renversement de signe n'est observé, ce qui témoigne d'une bonne préservation de la structure de dépendance entre les résidus.

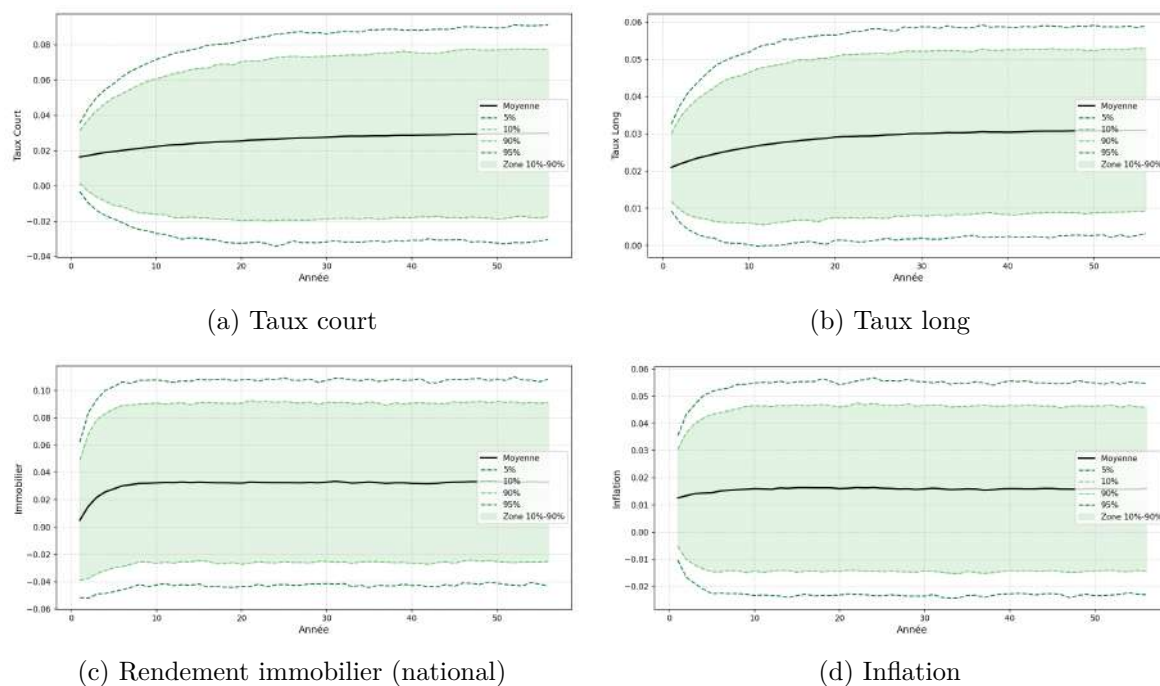
TABLE 3.2 – Comparaison des matrices de corrélation des résidus

	Historique				Simulé				
	Inflation	Immo	T. long	T. court	Inflation	Immo	T. long	T. court	
Inflation	1				Inflation	1			
Immo	0,347	1			Immo	0,324	1		
T. long	-0,720	-0,391	1		T. long	-0,568	-0,140	1	
T. court	-0,185	0,074	0,156	1	T. court	-0,142	0,154	0,099	1

Validation

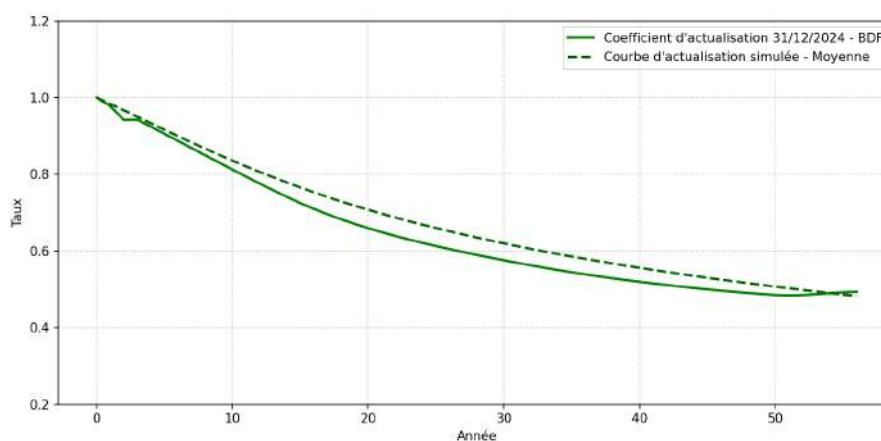
Les ensembles de scénarios obtenus sont illustrés ici par leur moyenne et des intervalles de confiance, permettant de quantifier l'incertitude autour des projections économiques. Ces visualisations offrent ainsi un cadre complet pour évaluer la cohérence des scénarios économiques utilisés dans la modélisation, et ainsi les valider ou non. Les courbes simulées pour l'immobilier, les taux et l'inflation apparaissent globalement cohérentes avec les tendances observées sur les marchés, validant la pertinence du modèle pour capturer la dynamique des principales variables économiques.

FIGURE 3.2 – Scénarios économiques simulés



La courbe des taux d'actualisation moyenne est comparée à la courbe d'actualisation obtenue avec les taux ZC de la Banque de France (BDF) au 31.12.2024. On note que la moyenne des courbes d'actualisation simulées surestime la courbe BDF. L'écart reste relativement faible et tend à s'annuler en fin de projection.

FIGURE 3.3 – Courbe d'actualisation



Source : [Banque de France, 2024]

Les espérances et variances des séries simulées ont été comparées aux moments empiriques des séries historiques, conformément aux recommandations de la littérature au sujet des GSEs monde réel⁸.

8. [Ahlgrim et al., 2005] et [F. Planchet et al., 2012].

TABLE 3.3 – Comparaison des moments simulés *vs.* historiques

Variable	E hist.	E sim.	ΔE	V hist.	V sim.	ΔV
Inflation	0,0291	0,0157	-0,0134	0,00094	0,00054	-0,00040
Immo	0,0444	0,0307	-0,0137	0,00286	0,00206	-0,00080
Taux long	0,0283	0,0287	+0,0004	0,00064	0,00027	-0,00036
Taux court	0,0169	0,0259	+0,0091	0,00076	0,00115	+0,00039

Cette comparaison met en évidence plusieurs écarts, notamment sur le rendement immobilier moyen ou sur l'inflation et les taux courts. En effet, comme mentionné et justifié ci-avant, le choix de profondeur d'historique a été réalisé dans le but de modéliser des scénarios économiques cohérents, avec une moyenne de rendements immobiliers plus faible que dans le passé (proche de 3% contre 4-5% en historique). L'objectif d'un GSE en monde réel n'est pas de coller parfaitement aux réalisations passées, mais de générer des trajectoires plausibles cohérentes avec les régimes économiques observés. Ces vérifications permettent donc de valider la capacité du modèle à reproduire des niveaux et des volatilités réalistes, ce qui est essentiel pour des applications actuarielles telles que celles qui suivent.

Focus rendements immobiliers

Les rendements immobiliers simulés présentent une distribution centrée autour de 3,3%, avec 95% des observations comprises entre -4% et +11%. Ces résultats traduisent la volatilité inhérente au marché immobilier ainsi que la présence de scénarios optimistes et pessimistes intégrés dans les projections. Cette dispersion permet de capturer la nature cyclique et les incertitudes liées à l'immobilier, tout en offrant une moyenne cohérente (mais tout de même relativement haute) avec les tendances observées sur le long terme.

La modélisation économique présentée précédemment constitue le socle sur lequel reposent les projections des variables clés : rendements immobiliers, inflation, taux d'intérêt réels et nominaux. Ces dynamiques permettent de simuler des scénarios économiques cohérents et réalistes sur le long terme, dans lesquels s'inscrivent les flux futurs liés aux prêts viagers hypothécaires. En s'appuyant sur ce cadre, l'objectif est désormais de construire un modèle de tarification robuste, capable de déterminer la valeur optimale de LTV. La section suivante introduit ainsi la formalisation de ces contraintes, puis la méthodologie retenue pour identifier la LTV maximale compatible avec les exigences de rentabilité et de risque fixées par l'organisme prêteur.

3.5 Contrainte d'optimalité de la *Loan-To-Value*

L'objectif central du prêteur dans le cadre d'un produit de prêt viager hypothécaire est d'allouer le capital le plus élevé possible à l'emprunteur, de manière à être compétitif sur marché, tout en garantissant une rentabilité suffisante et une gestion de risque solide. Cette logique repose sur un arbitrage délicat entre le montant prêté et la soutenabilité économique du contrat. Pour satisfaire cette logique, des contraintes économiques et prudentielles sont fixées en amont. Ces dernières permettent de borner le niveau de risque pris par le prêteur tout en assurant un niveau de rentabilité conforme à ses objectifs et une attractivité maximale. La LTV optimale est alors définie comme la valeur maximale de LTV pouvant être octroyée, sous réserve que l'ensemble de ces contraintes soit respecté. Cette section détaille précisément la nature de ces contraintes, leur formulation mathématique et leur interprétation

économique.

3.5.1 Rentabilité

Dans le cadre du prêt viager hypothécaire, le prêteur engage un capital pour une durée incertaine, avec un remboursement différé et plafonné par la valeur du bien. Afin d'assurer la viabilité économique du produit, il est donc indispensable de fixer les paramètres contractuels de manière à garantir une rentabilité suffisante au regard des risques sous-jacents.

Rendement moyen

La première contrainte de rentabilité choisie est basée sur le taux de rendement annuel. Ce dernier représente le taux constant équivalent qui aurait permis d'obtenir le même rendement global sur la période considérée, sous forme de coupon. Il est défini par la relation suivante, en gardant les mêmes notations que définies précédemment, et où T_i est la durée de vie sur le scénario i :

$$R_{\text{annuel}} = \frac{1}{N} \cdot \sum_{i=0}^N \frac{R_{\text{total}_i}}{T_i}$$

avec

$$R_{\text{total}} = \frac{\min(CRD_T, V_T) - C_0 - \sum_{t=0}^T CF_{\text{frais}_t}}{C_0 + CF_{\text{frais}_0}}$$

Concrètement, pour chaque police et chaque scénario simulé, le taux de rendement annuel attendu est calculé en utilisant la formule énoncée ci-avant, puis la contrainte suivante est testée :

$$\mathbb{E}(R_{\text{annuel}}) \geq R_{\text{cible}}$$

où R_{cible} est le niveau de rentabilité exigé (par la suite fixé à 6,5%),

Si la contrainte est respectée, la LTV est jugée admissible. La LTV optimale sous cette contrainte est la LTV admissible maximale.

Rentabilité minimale dans les scénarios adverses

Cette contrainte vise à contrôler le risque de sous-performance dans les situations extrêmes. L'idée est de garantir qu'en cas de conditions économiques ou de mortalité très défavorables — représentées par les $\alpha\%$ pires scénarios simulés — le taux de rendement annuel effectif du prêt ne descende pas sous un certain seuil de rentabilité minimale.

En effet, dans un contexte de *payoff* asymétrique pour le prêteur, il est insuffisant d'évaluer la rentabilité d'un produit financier uniquement à partir de la moyenne de rendement attendue. Il est fondamental d'intégrer également la variabilité et le risque de sous-performance, notamment dans les queues de distribution.

Mathématiquement, cette contrainte peut être formulée comme suit :

$$\mathbb{P}(R_{\text{annuel}} \geq r_{\text{seuil}}) \geq 1 - \alpha\%$$

Autrement dit, le rendement annuel doit dépasser une valeur minimale r_{seuil} dans au moins $(100 - \alpha)\%$ des cas. Cela équivaut à dire que le $\alpha\%$ -quantile de la distribution du rendement — c'est-à-dire la VaR au seuil $\alpha\%$ — doit être supérieur à ce seuil minimal :

$$VaR_{\alpha}(R_{\text{annuel}}) \geq r_{\text{seuil}}$$

où r_{seuil} est le rendement minimal accepté et $\alpha\%$ la probabilité maximale sous laquelle ce rendement minimal peut être atteint.

Cette contrainte traduit une aversion au risque de sous-performance. Par exemple, avec $\alpha = 5\%$ et $r_{\text{seuil}} = 2,5\%$, cela signifie que, même dans les 5% les plus adverses (scénarios de longévité extrême, marché immobilier défavorable, etc.), le rendement ne doit pas tomber en dessous de 2,5%. Le calibrage de la valeur de r_{seuil} reflète alors le niveau minimal de performance que le prêteur juge acceptable, même en cas de réalisation d'un scénario fortement défavorable.

Après s'être attardé sur les contraintes de rentabilité, deux contraintes de gestion de risque sont définies dans la section suivante. Ces deux types de contraintes sont complémentaires et répondent différemment selon le profil de risque, notamment l'âge. Les considérer conjointement permet donc de couvrir l'ensemble du spectre des risques et d'assurer un équilibre entre performance et sécurité.

3.5.2 Gestion de risque

Probabilité de déclenchement de la NNEG

Une autre contrainte imposée dans la tarification du contrat de prêt concerne la probabilité de déclenchement de la garantie de non-dépassement. Cette garantie implique que, si à l'échéance du prêt la valeur du bien immobilier est inférieure au capital restant dû (CRD), le prêteur ne peut récupérer que la valeur du bien. Il s'agit alors d'un scénario adverse. Mathématiquement, cela revient à étudier :

$$\mathbb{P}(CRD_T > V_T)$$

où CRD_T est le capital restant dû à la date de clotûre du prêt T , et V_T la valeur du bien immobilier au même horizon. Dans ce cas, le remboursement effectif pour le prêteur est :

$$\min(CRD_T, V_T) = V_T$$

Le risque financier associé à cette garantie est intégré via une contrainte probabiliste sur sa survenance. Pour chaque scénario simulé, la condition $CRD_T > V_T$ est testée. Pour des simulations i.i.d., la probabilité empirique de déclenchement de la NNEG est estimée comme :

$$\mathbb{P}(CRD_T > V_T) \approx \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\{CRD_T^{(i)} > V_T^{(i)}\}}$$

Cette probabilité est ensuite comparée à un seuil de tolérance β ; on étudie donc :

$$\mathbb{P}(V_T > CRD) \geq 1 - \beta$$

Soit :

$$VaR_{\beta}(V_T - CRD) > 0$$

où V_T est la valeur du bien à T , CRD l'encours du prêt et β la probabilité maximale de déclenchement (par exemple 45%).

De la même manière que pour la contrainte de rentabilité, la LTV optimale est considérée comme la LTV maximale vérifiant cette contrainte.

Valeur de la NNEG

La valeur de la garantie NNEG est calculée comme l'espérance des manques à gagner actualisés pondérés par la probabilité de survénance à chaque date, c'est-à-dire :

$$Valeur_{NNEG} = \sum_{t=0}^T {}_t|q_x \cdot \frac{\max(CRD_t - V_t, 0)}{DF_t}$$

où :

- ${}_t|q_x$ est la probabilité que l'emprunteur décède exactement à l'année t ,
- CRD_t est le capital restant dû à la date t ,
- V_t est la valeur du bien immobilier à la même date,
- DF_t est le déflateur, avec $DF_0=1$,
- T est le nombre d'année dont découle l'âge maximal considéré.

La quantité $\max(CRD_t - V_t, 0)$ représente le manque à gagner subi par le prêteur dans le cas où la garantie est activée, c'est-à-dire lorsque la valeur du bien ne suffit pas à couvrir le remboursement du prêt. Ce manque à gagner est actualisé à la date d'origine, puis pondéré par la probabilité de survénance du décès à cette date.

Ainsi, la valeur de la NNEG représente le coût économique moyen de la protection offerte par la garantie de non-dépassement, intégrée dans la valorisation du prêt comme une option implicite au bénéfice de l'emprunteur.

Dans le cadre de cette étude, la valorisation est effectuée de manière empirique. Ainsi, la probabilité de décès à une date t est traduite dans le modèle par une indicatrice sur la date de décès. La valeur de la NNEG est alors calculée en considérant uniquement la perte actualisée observée dans le scénario au moment du décès.

Dans le but de déterminer la LTV optimale, la valeur de la NNEG est exprimée en fonction du montant prêté ce qui permet d'annuler l'effet patrimoine. Cette contrainte traduit la volonté de contenir le risque de perte extrême induit par la NNEG ; l'expression étudiée est :

$$\mathbb{E} \left[\frac{NNEG}{C_0} \right] \leq \gamma$$

où NNEG est le coût de la garantie, C_0 le montant prêté à t_0 et γ le ratio maximal toléré, par exemple fixé à 8,5%.

Après avoir posé les contraintes encadrant le calcul de la LTV optimale modélisé un jeu de scénarios, il est possible de réaliser une première tarification centrale. Les quatre contraintes ici décrites encadrent la détermination de la LTV optimale et guident la construction de la tarification, présentée dans la section suivante.

3.6 Analyses et regard critique

Cette section présente les principaux indicateurs issus des simulations. Les LTV optimales obtenues sont très sensibles, par construction, au modèle de tarification construit. D'autres approches, par exemple prenant en compte le point de vue emprunteur au delà d'une simple maximisation sous contraintes du point de vue prêteur, auraient pu être utilisées et auraient présenté des divergences dans les résultats.

3.6.1 Analyses à la maille portefeuille

L'objectif est de quantifier les performances du portefeuille simulé et d'analyser les dynamiques de tarification. Cette lecture globale permet d'apprécier les effets conjoints des hypothèses techniques et du calibrage économique sur la viabilité et la rentabilité du produit. Les seuils de contraintes déterminés par avis d'expert, de manière à déclencher chacune des 4 contraintes, selon les profils, sont récapitulés dans le tableau suivant :

TABLE 3.4 – Seuils de contrainte

Seuils de contraintes	Valeur
(i) Rendement <i>flat</i> minimal	6,5 %
(ii) Rendement <i>flat</i> au 5 ^{ème} percentile minimal	2,5 %
(iii) Probabilité maximale de NNEG	45,0 %
(iv) Valeur maximale de la NNEG (%prêt)	8,5 %

Indicateurs macro - Immobilier

La valeur moyenne du portefeuille immobilier sous-jacent aux PVH est de l'ordre de 187k€ à t_0 et croît jusqu'à environ 415k€ à $t_{\text{décès}}$. L'augmentation de la valeur du portefeuille immobilier est de +121% sur toute la durée des prêts. Cette valorisation correspond à un rendement immobilier capitalisé moyen de 3,77%. Ces résultats confirment une croissance du patrimoine immobilier dans le temps, en moyenne, cohérente avec les tendances historiques.

TABLE 3.5 – Indicateurs globaux - Portefeuille immobilier

Indicateur	Valeur
Patrimoine moyen à t_0 (€)	187 051
Espérance de vie moyenne	21,39 ans
Patrimoine moyen à $t_{\text{décès}}$ (€)	412 872
Évolution du patrimoine	+121 %
Rendement immobilier <i>flat</i>	5,64 %
Rendement immobilier capitalisé	3,77 %

Indicateurs macro - PVH

Avec un taux d'emprunt fixé à 6%, le rendement moyen sur la base des flux nominaux atteint 9,8%, et 5,3% en rendement annualisé capitalisé. Ces niveaux de rentabilité reflètent l'effet des intérêts composés sur des durées longues. Le rendement du portefeuille de prêt est donc bien supérieur au rendement immobilier sous-jacent, en cohérence avec les risques pris. Le montant moyen de prêt est de 88k€, représentant une *Loan-To-Value* de 47,2%⁹. Le montant de prêt moyen peut être jugé relativement faible mais est cohérent avec le patrimoine moyen du portefeuille construit. En réalité, certains acteurs du marché refusent tout prêt inférieur à 50k€. Une LTV de 47% semble d'ailleurs plus généreuse que le marché. La VAN moyenne ressort élevée, ce qui s'explique par la structure du portefeuille et la structure des scénarios économiques en sortie de modèle. La probabilité moyenne de déclenchement de la clause NNEG est de 35% et représente une contrainte significative lors de la maximisation de la LTV puisqu'elle est limitante dans près d'un cas sur deux. La contrainte liée à la valeur NNEG (en % de valeur prêtée) est déclenchée dans un tiers des cas. Ainsi, pour la grande

9. A noter que la satisfaction du point de vue emprunteur n'est pas étudiée mais seulement prise en compte au travers de la maximisation de la LTV, ce qui constitue une limite de cette étude (c.f. 3.6.2).

majorité des individus, les contraintes liées à la garantie NNEG sont bloquantes. A noter que certains individus (0,4%) se voient refuser leur prêt sous ces conditions.

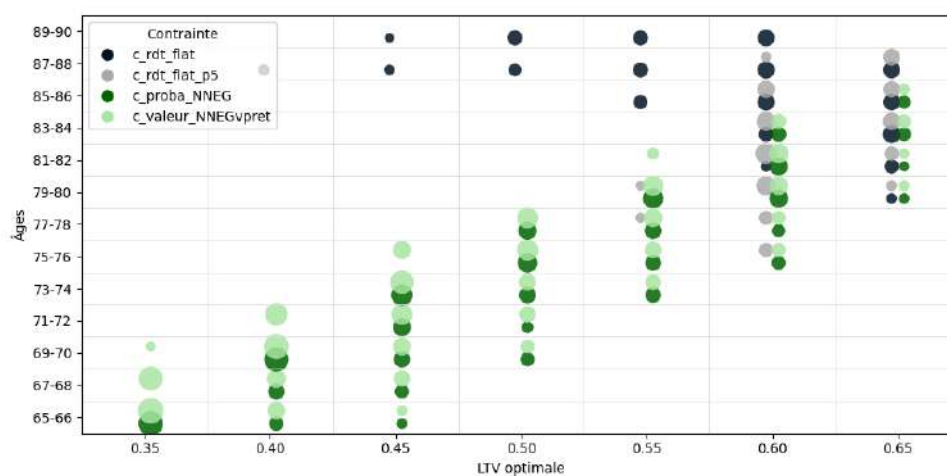
TABLE 3.6 – Indicateurs globaux - Portefeuille PVH

Indicateur	Valeur
Rendement <i>flat</i>	9,8 %
Rendement capitalisé	5,3 %
Prêt moyen (€)	88 329
VAN moyenne (€)	104 175
Rendement <i>flat</i> au 5 ^{ème} percentile	4,4 %
Probabilité NNEG	35,0 %
LTV moyenne	47,2 %
Proportion du portefeuille contrainte par le rdt. <i>flat</i>	13,2 %
Prop. du ptf. contrainte par le rdt. <i>flat</i> p5	6,6 %
Prop. du ptf. contrainte par la proba. NNEG	46,2 %
Prop. du ptf. contrainte par la valeur NNEG	33,5 %
Prop. du ptf. refusée	0,4 %
Proportion du portefeuille - Total	100 %

Contraintes limitantes dans le calcul de *Loan-To-Value* optimale

Globalement, l'âge et la LTV sont fortement corrélés positivement, sauf aux âges les plus élevés où la relation s'inverse, la LTV étant alors surtout déterminée par les contraintes liées à la NNEG. Au contraire, pour les espérances de vie courtes (âges élevés), les contraintes de rendement sont le plus souvent activées, ce qui s'explique par un effet coûts couplé à des intérêts composés moins forts, rendant plus difficile l'atteinte du seuil de rentabilité fixé. Diminuer la LTV permet de dépasser la valeur du prêt plus tardivement, et d'avoir un rendement de 6% plus longtemps. Ce dernier point explique la cassure observée autour de 80 ans. Cette dynamique est mise en évidence grâce à la figure 3.4 ci-dessous, permettant de visualiser les contraintes effectivement limitantes par LTV et âge. La taille des points y est proportionnelle au nombre d'individus appartenant au couple Âge-LTV de la cellule correspondante, et pour lesquels la contrainte associée (indiquée par la couleur) est limitante.

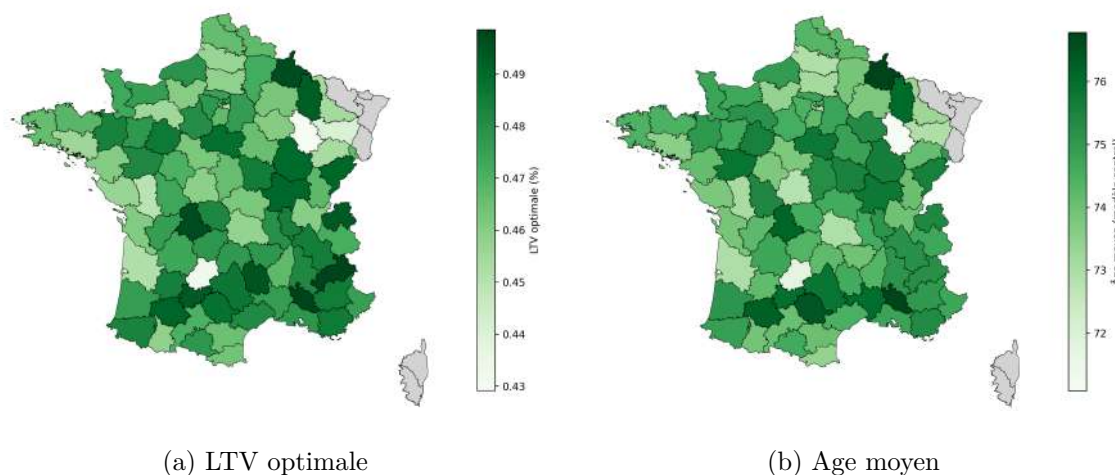
FIGURE 3.4 – Contraintes déclenchées par couple Age - LTV optimale



Note : Cette figure se lit tel que, pour les individus de 65 ou 66 ans qui ont une LTV optimale de 35% (cellule en bas à gauche), les contraintes de rendement *flat* ou de percentile de rendement n'ont jamais été déclenchées.

L'amplitude des LTV optimales moyennes par département n'est que de 8%, ce qui montre que les projections immobilières utilisées ici n'intègrent pas l'hétérogénéité territoriale du marché. Dans ce cadre, seuls les effets de longévité influencent la tarification ; le lien entre l'âge et la LTV optimale apparaît nettement à la lecture des cartes ci-dessous.

FIGURE 3.5 – Cartes des départementales LTV optimales et des âges moyens



Sensibilités aux caractéristiques de l'emprunteur

L'étude de l'impact de certaines modalités sur les montants prêtés et les indicateurs de rendement et de gestion risque permet de mieux comprendre les dynamiques de tarification.

Sexe et situation matrimoniale La segmentation selon le sexe et la situation matrimoniale permet de distinguer trois groupes présentant de réelles différences de longévité :

- hommes seuls,
- femmes seules,
- couples.

Elle met en évidence des écarts marqués dans les dynamiques de tarification. En moyenne, une LTV plus importante est calculée pour les femmes seules ou hommes seuls que pour les couples, mais les contraintes limitantes sont très différentes. Pour les personnes seules, le rendement est limitant dans environ un quart des cas, contre seulement un cas sur vingt pour les couples. Au contraire, les contraintes liées à la NNEG sont très limitantes pour les couples, en lien avec la durée moyenne de leur prêt et l'effet composé des intérêts. Selon le processus de décision décrit en section 3.5, la capacité maximale de prêt est calculée comme environ 50% de la valeur du bien aux individus seuls, contre 45% pour un couple.

A noter que 4% des hommes seuls se voient refuser¹⁰ leur prêt. Leur courte espérance de vie limite la durée de vie moyenne du prêt (et les intérêts composés n'ont pas le temps de compenser les frais), ce qui rend l'opération non acceptable au regard du rendement minimal.

10. Un refus traduit ici une absence de solution du système de maximisation : aucune LTV ne permet de satisfaire les quatre contraintes. Ce refus est dû aux niveaux de contraintes, donc à des critères économiques, non à des caractéristiques du bien en question ou des emprunteurs eux-mêmes.

TABLE 3.7 – Impact du sexe et de la situation matrimoniale sur les indicateurs PVH

Indicateur	Global	Femme seule	Homme seul	Couple
Rendement <i>flat</i>	9,8 %	8,6 %	8,6 %	10,6 %
Rendement capitalisé	5,3 %	5,2 %	5,2 %	5,3 %
Prêt moyen (€)	88 329	96 720	91 829	83 704
VAN moyenne (€)	104 175	86 716	83 760	115 977
Rendement <i>flat</i> au 5 ^{ème} percentile	4,4 %	3,9 %	3,9 %	4,8 %
Probabilité NNEG	35,0 %	33,0 %	32,8 %	36,5 %
LTV moyenne	47,2 %	51,7 %	49,8 %	44,6 %
Prop. du ptf. contrainte par le rdt. <i>flat</i>	13,2 %	26,3 %	22,3 %	5,3 %
Prop. du ptf. contrainte par le rdt. <i>flat</i> p5	6,6 %	10,4 %	7,9 %	4,6 %
Prop. du ptf. contrainte par la proba. NNEG	46,2 %	31,1 %	26,9 %	56,9 %
Prop. du ptf. contrainte par la valeur NNEG	33,5 %	32,2 %	38,9 %	33,2 %
Prop. du ptf. refusée	0,4 %	–	4,0 %	–

Age Au delà de la relation LTV optimale-Age lisible sur les figures 3.4 et 3.5, il est intéressant de s'attarder sur les indicateurs du portefeuille par âge. Trois segments ont été créés :

- inférieur à 70 ans,
- entre 70 et 80 ans,
- supérieur à 80 ans.

Chez les emprunteurs les plus âgés [>80], la rentabilité est plus faible : les intérêts composés ne s'accumulent que sur une courte durée, ce qui réduit les rendements. Cela explique que ce groupe soit majoritairement contraint par le rendement minimum, avec un taux de refus de 1,7%. À l'inverse, les assurés les plus jeunes [<70] présentent une très forte rentabilité, mais aussi une forte probabilité de NNEG (50,8%) malgré une LTV moyenne faible (37,3%).

TABLE 3.8 – Impact de l'âge sur les indicateurs PVH

Indicateur	Global	< 70	70–80	> 80
Rendement <i>flat</i>	9,8 %	13,2 %	9,8 %	7,0 %
Rendement capitalisé	5,3 %	5,4 %	5,3 %	5,1 %
Prêt moyen (€)	88 329	70 341	89 341	107 962
VAN moyenne (€)	104 175	144 395	105 328	53 349
Rendement <i>flat</i> au 5 ^{ème} percentile	4,4 %	6,0 %	4,3 %	3,2 %
Probabilité NNEG	35,0 %	36,5 %	37,5 %	30,0 %
LTV moyenne	47,2 %	37,3 %	47,5 %	58,9 %
Prop. du ptf. contrainte par le rdt. <i>flat</i>	13,2 %	–	–	52,4 %
Prop. du ptf. contrainte par le rdt. <i>flat</i> p5	6,6 %	–	1,3 %	24,0 %
Prop. du ptf. contrainte par la proba. NNEG	46,2 %	50,8 %	57,3 %	21,1 %
Prop. du ptf. contrainte par la valeur NNEG	33,5 %	49,2 %	41,4 %	0,8 %
Prop. du ptf. refusée	0,4 %	–	–	1,7 %

Les valeurs de seuils de contraintes pourraient être différenciées par âge pour mieux répondre à ces défis. Avec ces seuils, la LTV pour le segment [<70] est faible ce qui semble éloigné des attentes des emprunteurs, alors que le rendement moyen affiché est très élevé.

Valeur du bien Le niveau de patrimoine n'a pas été explicitement pris en compte dans la définition des contraintes du modèle, ce qui explique des résultats globalement homogènes

entre les différentes tranches patrimoniales. Malgré des différences notables dans les montants empruntés, les taux de refus et les contraintes restent très proches. Cela souligne une des limites du cadre actuel, abordée en section 3.6.2 : la non prise en compte de la valeur du bien et de son impact en termes de risque. En effet, on observe une espérance de vie plus élevée dans les quantiles supérieurs de patrimoine (c.f. tables de mortalité par quantile de niveau de vie : [INED, 2016]), ce qui met en évidence une inégalité de longévité en fonction du niveau de richesse. Cette limite peut entraîner une surestimation de la mortalité, conduisant ainsi à une sous-estimation de la probabilité de déclenchement de la NNEG. De plus, le modèle ne tient pas compte des effets de mutualisation : prêter dix fois 100k€ n'implique pas les mêmes risques que prêter une fois 1m€(c.f. section 3.6.2).

TABLE 3.9 – Impact du niveau de patrimoine sur les indicateurs PVH

Indicateur	Global	< 150k	150k–350k	> 350k
Rendement <i>flat</i>	9,8 %	9,8 %	9,7 %	9,8 %
Rendement capitalisé	5,3 %	5,3 %	5,3 %	5,3 %
Prêt moyen (€)	88 329	45 865	110 184	194 747
VAN moyenne (€)	104 175	53 941	128 961	233 654
Rendement <i>flat</i> au 5 ^{ème} percentile	4,4 %	4,4 %	4,4 %	4,4 %
Probabilité NNEG	35,0 %	34,8 %	34,8 %	35,3 %
LTV moyenne	47,2 %	47,2 %	47,3 %	47,0 %
Prop. du ptf. contrainte par le rdt. <i>flat</i>	13,2 %	13,3 %	13,7 %	10,9 %
Prop. du ptf. contrainte par le rdt. <i>flat</i> p5	6,6 %	6,8 %	6,6 %	5,9 %
Prop. du ptf. contrainte par la proba. NNEG	46,2 %	46,0 %	45,6 %	49,7 %
Prop. du ptf. contrainte par la valeur NNEG	33,5 %	33,3 %	33,9 %	33,0 %
Prop. du ptf. refusée	0,4 %	0,6 %	0,3 %	0,4 %

Sensibilités aux paramètres de constructions

Le calcul de LTV optimale est très sensible aux paramètres utilisés, et notamment au taux d'emprunt fixé et aux seuils de contraintes :

Taux d'emprunt On observe que le rendement global augmente avec le taux d'emprunt ; ce phénomène est décrit en section 1.2.1. Les contraintes limitantes évoluent avec le taux : à 5%, 17% du portefeuille est exclu à cause d'un rendement insuffisant, tandis qu'à 7% les clauses liées à la NNEG qui sont limitantes. Le prêteur n'est pas prêt à prêter plus, bien que le rendement fixé en amont soit élevé, si c'est pour être plafonné par le rendement immobilier.

TABLE 3.10 – Impact du taux d'emprunt sur les indicateurs PVH

Indicateur	5 %	6 %	7 %
Rendement <i>flat</i>	7,9 %	9,8 %	12,6 %
Rendement capitalisé	4,4 %	5,3 %	6,2 %
Prêt moyen (€)	82 239	88 329	74 156
Rendement <i>flat</i> au 5 ^{ème} percentile	3,3 %	4,4 %	5,9 %
Probabilité NNEG	33,8 %	35,0 %	36,1 %
LTV moyenne	44,0 %	47,2 %	39,6 %
Prop. du ptf. contrainte par le rdt. <i>flat</i>	21,5 %	13,2 %	1,2 %
Prop. du ptf. contrainte par le rdt. <i>flat</i> p5	16,3 %	6,6 %	2,9 %
Prop. du ptf. contrainte par la proba. NNEG	23,9 %	46,2 %	52,5 %
Prop. du ptf. contrainte par la valeur NNEG	20,9 %	33,5 %	43,5 %
Prop. du ptf. refusée	17,3 %	0,4 %	–

Contraintes isolées L'analyse des contraintes isolées met en évidence des effets très différenciés sur les indicateurs. En appliquant uniquement la contrainte de rendement, on observe un niveau de prêt moyen élevé : cela s'explique par l'effet de levier permis par les emprunteurs 'jeunes', dont la longue espérance de vie maximise les intérêts composés. Cependant, cela entraîne un fort risque de déclenchement de la clause NNEG, illustré par une probabilité de 72,7%. À l'inverse, la contrainte sur la valeur NNEG conduit à un encours plus modéré : en prêtant moins (LTV moyenne à 49%), on vise à réduire drastiquement le risque d'activation de la garantie, ce qui permet un rendement *flat* relativement élevé sur la durée, mais sur un capital prêté moindre. Enfin, la contrainte lié au rendement au 5^{ème} percentile agit comme un filtre conservateur sur les cas extrêmes, limitant les situations à très faible rendement : on prête davantage qu'en central, mais en acceptant une probabilité NNEG plus forte, ce qui donne un rendement *flat* proche de celui du cas central.

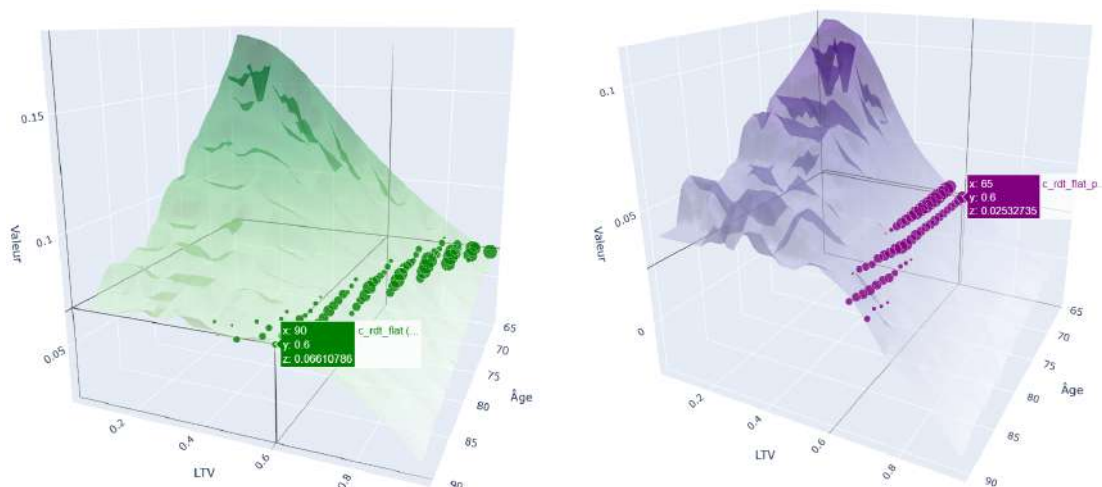
TABLE 3.11 – Indicateurs globaux du Portefeuille PVH - contraintes isolées

Indicateur	Central	Proba. NNEG	Valeur NNEG	Rendement	Rendement p5
Rendement <i>flat</i>	9,8 %	9,0 %	9,4 %	6,7 %	8,8 %
Rendement capitalisé	5,3 %	5,0 %	5,1 %	4,0 %	4,9 %
Prêt moyen (€)	88 329	95 416	92 440	142 860	109 249
Rendement <i>flat</i> au 5 ^{ème} percentile	4,4 %	4,0 %	3,9 %	1,1 %	2,9 %
Probabilité NNEG	35,0 %	41,0 %	38,7 %	72,7 %	51,1 %
LTV moyenne	47,2 %	51,0 %	49,4 %	76,4 %	58,4 %

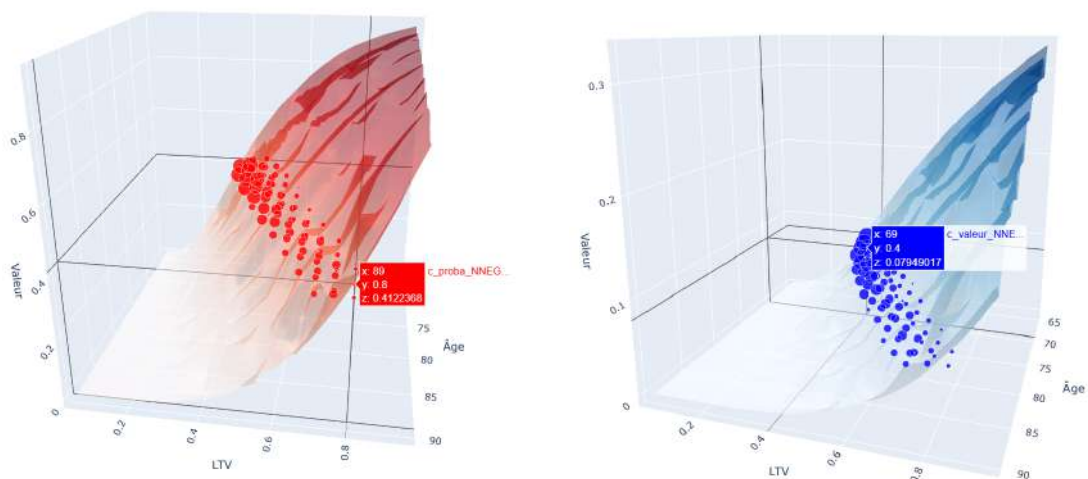
Les figures¹¹ ci-dessous permettent de mettre en évidence de manière visuelle les comportements différents impliqués par les contraintes, mais aussi les âges :

11. Les figures de type 3.6 reposent sur une interpolation linéaire permettant de passer outre le caractère discrétionnaire des LTV testées et des âges.

FIGURE 3.6 – LTV optimale - contraintes isolées



(a) Rendement moyen minimal = 6,5%

(b) 5^{ème} percentile de rdt minimal = 2,5%

(c) Probabilité de NNEG maximal = 45%

(d) Valeur de NNEG maximale = 8,5% du prêt

Note : Sur ces figures, les points représentent les individus effectivement limités par la contrainte en question, au croisement Age-LTV optimale. Les surfaces représentent les dynamiques des indicateurs étudiés en absence de contrainte.

3.6.2 Limites

L'approche retenue comporte plusieurs limites, présentées ci-après :

Modélisation simplifiée du contrat

Dans cette étude, le produit PVH est modélisé comme un prêt in fine versé en une seule fois (*lump sum*) et remboursé au décès, sans flux intermédiaire excepté les frais, ni tirage supplémentaire. Cette modélisation ne reflète pas la diversité des contrats pouvant être commercialisés. En pratique :

- des options de remboursement anticipé ou de sortie peuvent être incluses,
- la durée de détention dépend de comportements endogènes (revente, changement de résidence, décès du conjoint, etc.),

- certains contrats proposent des avances fractionnées ou des retraits échelonnés (notamment en Australie ou aux Etats-Unis).

Ici, la seule sortie du contrat modélisée est la mort de l'emprunteur ou du groupe au dernier décès assimilé en cas de couple. Cet élément représente une limite.

Ces simplifications influencent la durée effective du prêt, les flux de trésorerie réels, ainsi que la valorisation de la NNEG. Ne pas les intégrer dans la modélisation revient à sous-estimer la variabilité des résultats économiques. Cependant, cette limite n'entrave pas fortement l'objectif final de l'étude, à savoir une comparaison de méthode liée à la modélisation segmentée ou non des rendements immobiliers. Affiner davantage la flexibilité du produit nécessiterait d'introduire une série d'hypothèses supplémentaires en l'absence de données détaillées sur les comportements observés et les caractéristiques contractuelles spécifiques.

Duration, gestion actif-passif, coût du capital

L'étude se focalise sur l'équilibre du contrat individuel sans intégrer les contraintes de durée ou de liquidité à l'échelle du portefeuille. Or, pour le prêteur, l'*Asset Liability Management* est essentiel pour garantir la soutenabilité du produit dans le temps. Un prêteur peut avoir des contraintes en terme de durée de blocage des fonds prêtés, ce qui n'est pas représenté dans le modèle construit ici (ou du moins seulement par l'âge minimal de souscription fixé à 65 ans).

Un autre aspect non modélisé concerne le dénouement opérationnel du contrat au décès de l'emprunteur. Dans la réalité, le recouvrement du capital prêté n'est pas instantané : il dépend du délai de mise en vente du bien immobilier, de la complexité de la succession, des éventuels litiges, et des frais de gestion ou de notaire. Ces éléments peuvent entraîner un délai entre le décès et la récupération effective des fonds, générant un besoin de refinancement temporaire ou un coût d'opportunité pour le prêteur. De plus, l'utilisation des tables de mortalité TGH05 et TGF05 induit un biais dans les durées de vie simulées, le profil de mortalité du portefeuille s'écartant de celui décrit par ces tables.

Ici, la capitalisation des intérêts, la revalorisation du bien immobilier ainsi que les frais s'arrêtent au décès, cela représente un risque supplémentaire, car le prêteur continue de supporter des charges (fiscales, administratives, de refinancement) pendant la période de vente sans perception d'intérêts, et la valeur du bien varie en continue.

Une contrainte explicite sur la durée moyenne pourrait être intégrée dans une future extension. Par ailleurs, aucune hypothèse explicite n'est faite ici sur le coût du capital réglementaire ou économique. En pratique, un produit à forte consommation de capital peut être moins attractif même s'il est rentable.

Couple LTV - Taux

Taux d'emprunt unique L'étude repose sur l'hypothèse d'un taux d'intérêt contractuel unique fixé à 6%. Ce choix vise à représenter un contexte de marché réaliste tout en facilitant les comparaisons entre profils emprunteurs. Cependant, cela ne permet pas d'explorer l'ensemble des configurations possibles, notamment celles correspondant à des taux plus faibles, qui pourraient rendre le produit plus compétitif pour certaines catégories d'emprunteurs. La mise en place d'une politique tarifaire différenciée, ajustant le taux en fonction du profil de l'emprunteur, constituerait une amélioration pertinente de l'outil actuel.

A noter que des calculs ont été réalisés pour les taux 5% et 7% de manière à chiffrer l'impact d'un changement sur cette hypothèse (c.f. 3.6.1). Un choix de taux dépendant de l'âge, et po-

tentiellement à terme de la situation géographique, pourrait constituer une piste intéressante, mais introduirait une complexité supplémentaire au modèle.

Granularité de la grille de LTV La grille de recherche des LTV optimales repose sur un pas fixe de 5 points de pourcentage entre 5% et 95%. Si cela permet un calcul rapide et robuste, cette granularité peut manquer de précision, notamment dans les zones frontalières où une LTV admissible pourrait être juste au-dessus ou en dessous d'un seuil. L'impact de cette limite semble tout de même relativement limité.

Capter l'asymétrie du *payoff* prêteur, sous contraintes

Approche Monte-Carlo L'utilisation d'une approche Monte-Carlo, via en quelques sortes un double tirage aléatoire (scénarios économiques et scénarios de mortalité individuels), permet de capturer l'asymétrie du *payoff* du prêteur, liée à la longévité extrême ou au faible rendement immobilier. Cette approche génère une distribution empirique des indicateurs, ce qui en fait un outil adapté à l'évaluation du risque dans un cadre incertain. Le nombre total de combinaisons simulées dans cette étude s'élève à 10 000, valeur validée par une analyse de sensibilité sur la convergence des résultats (c.f. Annexe.4). Cette taille d'échantillon génère une variabilité résiduelle (erreur Monte-Carlo), mais reste suffisante pour stabiliser les estimations tout en maintenant un faible temps de calcul.

Contraintes fixées Les seuils de rentabilité ou de risque utilisés (rendement minimal, probabilité de déclenchement et valeur de la NNEG) sont fixés en amont, par avis d'expert. Or, en pratique, ces seuils relèvent de la politique de risque propre à chaque institution de prêt. Une approche alternative consisterait à intégrer ces contraintes dans un cadre d'optimisation global.

Par ailleurs, les contraintes prises en compte dans l'analyse ne dépendent pas directement du montant total prêté, ni du montant effectivement exposé au risque de perte. En effet, la valeur du NNEG est rapportée au montant prêté, de sorte qu'elle s'exprime sous la forme d'un ratio. Cette approche vise à modéliser la situation d'un prêteur disposant d'un capital donné (C) et souhaitant atteindre un objectif de rendement annuel (Y) tout en maîtrisant son niveau de risque.

Ce choix méthodologique implique que la mutualisation du risque entre emprunteurs n'est pas explicitement prise en compte. En effet, dix emprunteurs sollicitant chacun 100 000€ et un emprunteur unique sollicitant 1 000 000€ sont traités de la même manière, pour un profil de risque similaire. Or, la diversification des montants prêtés entre plusieurs individus pourrait permettre de réduire la volatilité des pertes potentielles et d'optimiser la gestion du risque global du portefeuille.

Absence du point de vue emprunteur

Le modèle se place exclusivement du point de vue du prêteur, sans modéliser les préférences, contraintes ou comportements des emprunteurs. Dans le cadre de cette étude, seule la maximisation de la LTV a pour objectif de défendre l'intérêt des emprunteurs. Une extension naturelle serait d'intégrer un module comportemental ou d'évaluer la valeur perçue côté emprunteur. A noter que le mémoire IA [D. Crespine, 2017] *Le prêt viager hypothécaire, une solution pour financer sa retraite ?*, réalisé également au sein de Deloitte, s'attarde sur le besoin client dans le cadre du PVH.

Validité des modèles financiers utilisés

Les modèles financiers utilisés pour projeter l'évolution des taux d'intérêt ou des prix immobiliers (Vasicek) reposent sur des hypothèses structurelles fortes. Or, les marchés immobiliers ne vérifient que très partiellement ces hypothèses (c.f. 1.1.2). Ils sont marqués par une hétérogénéité structurelle des actifs (localisation, qualité, usage) difficilement modélisable avec un processus ici paramétré à une maille agrégée.

Ces limites structurelles remettent en question l'adéquation de modèles financiers traditionnels paramétré de la sorte pour la modélisation des dynamiques immobilières dans le cadre du PVH. Elles justifient le recours à une approche alternative, plus fine.

Dans cette optique, une modélisation des prix immobiliers à une maille géographique fine, tenant compte des caractéristiques intrinsèques des biens (typologie et localisation précise), permet à minima de capter l'hétérogénéité des actifs immobiliers. Elle permet de mieux saisir les déterminants structurels de la valorisation des biens et d'introduire une granularité indispensable à une évaluation plus robuste des risques associés aux PVH. Le développement de deux méthodes de tarification basées sur ce point est fait en section suivante.

Construction de l'outil de tarification

L'objectif de ce chapitre a consisté à présenter le calcul de la LTV optimale en environnement central, considérée comme la LTV maximale sous contraintes de rentabilité et de gestion de risque, pour un taux d'emprunt à 6%, en phase avec le marché actuel.

1. Entrées du modèle :

Hypothèses : taux de prêt fixe à 6%, aucune sortie anticipée, frais tout au long de la vie du prêt (plus élevés à t_0 , lors de la signature et $t_{\text{décès}}$, lors de la clôture), tables de mortalité TGH05/TGF05, remboursement in fine, *lump sum* (sans tirage), remboursement au moment du décès.

Caractéristiques individuelles : âge, sexe, statut marital.

Caractéristiques immobilières : département, type de bien, valeur foncière (calculée à partir de la surface et du prix/m²).

Scénarios simulés : 10 000 scénarios économiques (immobilier, inflation, taux courts, taux longs), 10 000 scénarios de mortalité par individu.

2. Sorties du modèle :

Pour chaque couple LTV \times individu (avec LTV $\in \{5\%, 10\%, \dots, 95\%\}$), le simulateur retourne des indicateurs de rentabilité (dont la moyenne et des quantiles de rendement et la VAN) et de gestion de risque (dont le coût espéré de la NNEG, et sa probabilité de déclenchement).

La LTV optimale est calculée comme la LTV maximale qui respecte :

- un rendement *flat*, assimilable à un coupon, moyen supérieur à 6,5%,
- un 5^{ème} quantile de rendement *flat* supérieur à 2,5%,
- une probabilité de déclenchement de la NNEG inférieure à 45%,
- une valeur de NNEG inférieure à 8,5% du montant prêté.

3. Limites :

L'approche retenue présente certaines limites : elle repose sur une modélisation simplifiée des contrats et ne prend pas en compte les aspects de duration ou de coût du capital. L'analyse reste centrée sur le point de vue du prêteur, sans intégrer celui de l'emprunteur ni la mutualisation des risques à l'échelle portefeuille. Toutefois, ces choix se justifient dans la mesure où l'outil central vise avant tout à comparer différentes approches de tarification.

Modélisation 'centrale' du rendement immobilier

Sous cette 1^{ère} méthodologie, servant de témoin, un unique rendement immobilier est modélisé à l'échelle nationale, via un GSE de type Ahlgrim simplifié. Ce rendement est simulé selon une structure de dépendance avec les taux courts, taux longs et l'inflation, et présente une valeur de long terme autour de 3,33%. La LTV moyenne qui en découle est de l'ordre de 47,2%.

Chapitre 4

Modélisation segmentée du risque immobilier

La valorisation des prêts viagers hypothécaires repose en grande partie sur l'évolution future de la valeur des biens immobiliers sous-jacents. Si une modélisation homogène à l'ensemble du marché offre une première approximation du risque immobilier, elle ne permet pas de capter l'hétérogénéité réelle des rendements de ce type d'actif.

Inspiré par des travaux tels que [A-W. Shao et al., 2014] *Reverse Mortgage Pricing and Risk Analysis Allowing for Idiosyncratic House Price Risk and Longevity Risk*, qui, pour le marché de Sydney, évaluent l'impact sur les PVH d'une modélisation stochastique de la longévité couplée à une segmentation des rendements immobiliers en cinq sous-marchés, cette étude propose une approche basée sur la modélisation segmentée du risque immobilier, appliquée au contexte national français.

Une telle approche nécessite de disposer d'une base historique suffisamment longue et homogène afin de calibrer les modèles et identifier les dynamiques structurelles propres à chaque segment. Or, il n'existe aucun indice immobilier segmenté *open-source*, couvrant l'ensemble du territoire, avec un historique suffisamment long et homogène ; il est donc nécessaire de reconstituer des indices segmentés de prix immobiliers afin de refléter au mieux la diversité des trajectoires locales. La segmentation appliquée repose sur le département et le type de bien. Cette différenciation pourrait, en théorie, être enrichie par d'autres dimensions telles que l'environnement immédiat du bien ou l'état du logement. Ces informations n'étant pas publiques, les méthodes retenues ont été conçues de façon modulaire afin de pouvoir être complétées ou affinées ultérieurement si des données plus détaillées deviennent disponibles.

4.1 Étape préliminaire : Construction d'indices segmentés

Cette section décrit la méthode employée pour construire des indices immobiliers segmentés à la maille département \times type de bien, en s'appuyant sur un modèle hybride combinant une approche hédonique et une approche par ventes répétées. Idéalement, si des indices segmentés et historiquement profonds existaient déjà pour le marché français, ils auraient pu être utilisés directement lors du paramétrage et de l'entraînement des modèles.

En France, les données DVF recensent l'intégralité des transactions immobilières et permettent ainsi de construire des indices segmentés à la maille souhaitée. Toutefois, la profondeur historique de cette base reste limitée : de 2014 à 2024. Cette fenêtre temporelle, trop courte, ne permet pas à elle seule de modéliser des dynamiques immobilières de long terme. Pour pallier cette limitation, une démarche de *backcast* est mise en œuvre, permettant de reconstituer un historique étendu (1980–2024) à partir des dynamiques observées localement et nationalement.

Ce *backcast* ne compromet pas la validité des analyses, mais peut apporter un biais. Plusieurs vérifications empiriques ont été réalisées pour en garantir la cohérence : d'une part, les rendements obtenus sont comparés aux rendements réels observés sur la période 2014–2024, à

partir d'une source indépendante, afin d'en contrôler l'alignement ; d'autre part, les trajectoires reconstruites sont recentrées vers la moyenne de rendement national connu (INSEE), assurant ainsi un ancrage macroéconomique. À titre d'illustration, les indices reconstruits pour Paris reproduisent fidèlement les rendements annuels observés ; un écart modéré peut apparaître en cumul sur le long terme (de l'ordre de 50 à 100% après 30 ans), ce qui s'explique par les incertitudes agrégées sur longue période. Ces écarts restent compatibles avec les objectifs de modélisation de cycles immobiliers à long terme.

Les indices retraités sont utilisés pour les deux méthodes de projection affinées :

- Méthode 1 : Génération de scénarios économiques stochastiques segmentés (c.f. 4.2),
- Méthode 2 : Réseau de neurones récurrent *Long Short-Term Memory* (LSTM, c.f. 4.3).

Les principes théoriques régissant la construction d'indices immobiliers sont présentés ci-après.

4.1.1 Principes théoriques

Les régressions hédoniques sont souvent utilisées dans le monde immobilier pour construire des indices, et consistent à se baser sur les caractéristiques d'un bien pour lui donner une valeur. De nombreux travaux scientifiques relatent de ce sujet, on peut notamment citer [J. Hansen, 2006], [C. Jones, 2009], [S-C. Bourassa et al., 2010], [A-W. Shao et al., 2013] ou encore [INSEE Méthodes n° 132, 2019].

Régression hédonique simple

Les modèles hédoniques reposent sur l'hypothèse fondamentale que le prix d'un bien immobilier est déterminé par les caractéristiques observables du bien et de son environnement. Cette hypothèse permet de modéliser les log-prix comme une fonction des attributs structurels, temporels et géographiques¹.

La forme mathématique classique du modèle hédonique est :

$$\ln(P_{it}) = \alpha + X'_{it}\beta_t + T'_t\gamma_t + \varepsilon_{it} \quad (4.1)$$

où :

- P_{it} est le prix du bien i observé à la date t ,
- α est l'intercept,
- X est le vecteur de caractéristiques observables du bien (surface, nombre de pièces, type de bien, etc.),
- β est le vecteur de coefficients associés aux caractéristiques,
- T est le vecteur d'indicateurs temporels,
- γ est le vecteur de coefficients associés aux effets fixes temporels,
- les erreurs ε_{it} sont supposées i.i.d., centrées. Elles capturent l'hétérogénéité non observée : caractéristiques du bien ou de son environnement non mesurées (qualité de construction, orientation, vue, bruit, état intérieur, etc.) ainsi que les chocs idiosyncratiques de marché liés à une transaction particulière (pression de négociation, urgence de vente, etc.).

1. [A-W. Shao et al., 2013].

Ce modèle doit se baser sur un ensemble de ventes uniques, chaque bien ne figurant qu'une seule fois dans l'échantillon considéré.

Bien que le modèle hédonique soit intuitif et interprétable, il présente certaines limites. En effet, il repose sur l'hypothèse que la qualité des biens est parfaitement captée par les variables explicatives, ce qui peut introduire un biais si des caractéristiques importantes ne sont pas observées. De plus, il suppose que les préférences des acheteurs et les valeurs marginales sont constantes dans le temps, ce qui peut ne pas être vérifié en pratique. Enfin, il n'exploite pas les informations portées par les biens vendus plusieurs fois. Pour remédier à ces limites, la littérature propose des modèles alternatifs tels que les modèles à ventes répétées (*repeat-sales*) ou encore hybrides.

Méthode des ventes répétées

Les modèles à ventes répétées exploitent l'information contenue dans les biens ayant fait l'objet de plusieurs transactions au cours du temps². L'idée centrale du modèle est d'éliminer l'hétérogénéité inobservable entre biens en comparant uniquement les prix successifs d'un même bien, supposé inchangé entre deux transactions. Cela permet d'éviter explicitement la modélisation des caractéristiques individuelles du bien. Si un bien i est vendu à deux dates distinctes t_1 et t_2 avec $t_1 < t_2$, on modélise la différence des logarithmes de prix telle que :

$$\ln(P_{i,t_2}) - \ln(P_{i,t_1}) = (T_2 - T_1)\gamma_t + \varepsilon_{i,t_2} - \varepsilon_{i,t_1} \quad (4.2)$$

où :

- $P_{i,t}$ est le prix du bien i à la date t ,
- γ est un effet fixe temporel (indicatrice d'année de mutation) capturant la tendance du marché dans le temps,
- ε_{i,t_1} et ε_{i,t_2} sont supposés i.i.d et suivre les mêmes caractéristiques que les termes d'erreur de la régression hédonique simple.

Cependant, le modèle à ventes répétées ne repose que sur un sous-ensemble des données disponibles, ce qui introduit un biais d'échantillonnage. Ces biens peuvent ne pas être représentatifs du marché global. De plus, il suppose implicitement que les caractéristiques physiques du bien n'ont pas changé entre les deux ventes, ce qui peut être discutable en pratique. Le modèle à ventes répétées est rarement utilisé seul mais plutôt au sein de modèles hybrides.

Modèle hybride

Afin de tirer parti des avantages respectifs du modèle hédonique et du modèle à ventes répétées, le modèle hybride est souvent utilisé. [A-W. Shao et al., 2013] met d'ailleurs en évidence la qualité de ce modèle en le comparant à 8 autres méthodes. L'approche repose sur trois équations complémentaires :

- Une régression hédonique classique sur les ventes uniques, modélisant le logarithme du prix en fonction des attributs observables et du temps :

$$\ln(P_{it}) = \alpha + X'_{it}\beta_t + T'_t\gamma_t + \delta_{r(i)} + \eta_i + \varepsilon_{it} \quad (4.3)$$

où η_i capte les erreurs intrinsèques (effets non observés mais systématiques) et ε_i traduit la variabilité aléatoire résiduelle (aléa pur autour de la relation modélisée)

2. [C. Jones, 2009].

- Une régression hédonique sur les ventes répétées, en excluant la dernière vente de chaque bien :

$$\ln(P_{it}) = \alpha + X'_{it}\beta_t + T'_t\gamma_t + \delta_{r(i)} + \eta_i + \varepsilon_{it} \quad (\text{pour } t < t_{\max}) \quad (4.4)$$

- Une régression à ventes répétées (où X_i est supposé constant entre deux ventes, soit des caractéristiques inchangées, et β ne dépendant pas du temps) permettant d'estimer l'évolution temporelle de l'indice :

$$\ln(P_{i,t_2}) - \ln(P_{i,t_1}) = T_2\gamma_{t_2} - T_1\gamma_{t_1} + \varepsilon_{i,t_2} - \varepsilon_{i,t_1} \quad (4.5)$$

où pour deux biens différents i et j , on suppose :

$$\begin{aligned} \mathbb{E}(\varepsilon_{it}) &= 0, \\ \mathbb{E}(\eta_i) &= 0, \\ \mathbb{E}(\varepsilon_{it_2}^2) &= \sigma_\varepsilon^2, \\ \mathbb{E}(\eta_i^2) &= \sigma_\eta^2 \\ \mathbb{E}(\varepsilon_{it_2}\varepsilon_{jt_1}) &= 0 \quad \text{si } (i-j)^2 + (t_2-t_1)^2 \neq 0, \\ \mathbb{E}(\eta_i\eta_j) &= 0 \quad \text{si } i \neq j, \\ \mathbb{E}(\eta_i\varepsilon_{it_2}) &= 0. \end{aligned}$$

Avec, sous ces hypothèses, la matrice de covariance suivante :

$$\text{Cov} = \begin{pmatrix} (\sigma_\varepsilon^2 + \sigma_\eta^2) \cdot I_M & 0 & 0 \\ 0 & (\sigma_\varepsilon^2 + \sigma_\eta^2) \cdot I_N & -\sigma_\varepsilon^2 I_N \\ 0 & -\sigma_\varepsilon^2 I_N & 2\sigma_\varepsilon^2 I_N \end{pmatrix} \quad (4.6)$$

où I_M et I_N représentent les matrices identités de dimensions M et N respectivement, avec M le nombre de biens vendus une seule fois et N le nombre de paires de ventes répétées.

La matrice de covariance définie ci-dessus découle directement des hypothèses faites sur la structure des erreurs ε_{it} et des effets spécifiques aux biens η_i . Elle reflète la dépendance particulière entre les erreurs dans le cadre d'un modèle combinant ventes uniques et ventes répétées.

- Le premier élément de la diagonale $(\sigma_\varepsilon^2 + \sigma_\eta^2) \cdot I_M$ correspond à la variance totale des observations issues des ventes uniques, incluant à la fois l'erreur idiosyncratique ε_{it} et l'effet fixe individuel η_i .
- Le deuxième élément de la diagonale $(\sigma_\varepsilon^2 + \sigma_\eta^2) \cdot I_N$ concerne la première observation de chaque paire de ventes répétées. Chacune de ces ventes est affectée par la même structure de variance que les ventes uniques.
- Le troisième terme de la diagonale $2\sigma_\varepsilon^2 \cdot I_N$ correspond à la variance des différences de log-prix dans le modèle à ventes répétées. Elle reflète que l'effet fixe η_i s'annule dans la soustraction, laissant une variance doublée de l'erreur aléatoire.
- Les termes hors diagonale $-\sigma_\varepsilon^2 \cdot I_N$ capturent la covariance négative entre deux observations consécutives pour un même bien dans le cadre des ventes répétées : l'erreur ε_{it} est partagée dans les deux observations, ce qui crée une dépendance négative dans leur différence.

Cette structure particulière permet de prendre en compte à la fois l'effet fixe lié au bien et les erreurs de mesure indépendantes, tout en conservant une forme exploitable pour une estimation par moindres carrés généralisés (GLS).

Calcul d'un indice

À partir des coefficients estimés $\hat{\gamma}$ et $\hat{\beta}$, on peut construire un indice global des prix des logements tel que :

$$\ln(I_t) = \hat{\gamma}_t + \bar{X}_0' \hat{\beta}_t, \text{ donc } I_t = \exp(\hat{\gamma}_t + \bar{X}_0' \hat{\beta}_t)$$

où \bar{X}_0 correspond au vecteur des caractéristiques moyennes pour l'année de référence. Pour déterminer l'évolution du prix d'un indice désagrégé, désigné par k , on utilise la formule :

$$I_t^k = I_t \exp\left[(X^k - \bar{X}_0)' \hat{\beta}_t\right], \text{ rebasé ensuite en base 100 en } s \text{ tel que } I_t^k(s) = \frac{I_t^k}{I_s^k} \cdot 100$$

dans laquelle X^k représente le vecteur des caractéristiques propres au segment k de logements.

4.1.2 Construction des indices département \times type de bien

Pour chaque segment défini par le croisement département \times type de bien, un indice de prix annuel en base 100 (2014) est construit en suivant la méthode hybride décrite ci-avant. Cette approche combine trois sources d'information issues des données DVF :

1. Les ventes simples,
2. Les ventes répétées,
3. Les différences de log-prix entre deux ventes successives.

L'analyse est restreinte aux ventes de biens résidentiels clairement identifiables, en conservant uniquement les ventes d'appartements uniques et de maisons individuelles, afin de limiter l'ambiguïté liée aux mutations multiples ou aux biens composites. Les données DVF accessibles ne contiennent pas d'adresse précise ou de coordonnée géographique ; il est nécessaire de déterminer à partir des données disponibles une méthode d'identification des ventes répétées. Les ventes répétées sont ici identifiées par une clef construite sur la base du numéro de parcelle, de la surface, du type de bien et du nombre de pièces. De cette clef découle de nombreuses ventes répétées ; la surface étant arrondie à l'unité, plusieurs ventes de biens au sein de la même parcelle, avec la même surface et le même nombre de pièces ont lieu la même année. Pour palier cela, ces ventes identifiées comme identiques sont moyennées pour être considérées comme une unique vente. L'hétérogénéité au sein de ces biens de même caractéristiques est donc perdue ; mais elle n'aurait pas pu être captée au vu de la précision des données et des variables disponibles.

Une fois les bases de ventes simples et répétées identifiées, deux régressions hédoniques indépendantes sont réalisées sur les ventes simples et répétées, selon les variables de surface et de nombre de pièces. En parallèle, un modèle *repeat-sales* est ajusté sur les différences de log-prix calculés entre deux mutations successives d'un même bien où les variables explicatives correspondent aux indicatrices temporelles associées aux années des mutations. À partir des résidus issus de ces régressions, trois variances sont estimées :

- $\hat{\sigma}_\varepsilon^2$, la variance totale des résidus des modèles 1., calculée comme la moyenne des carrés des résidus issus des régressions sur ventes simples et répétées, soit :

$$\hat{\sigma}_\varepsilon^2 = \text{MSE}_{\text{simples} + \text{répétées}}$$

- $\hat{\sigma}_\xi^2$, la variance résiduelle du modèle 2., rapportée à 2 car elle porte sur la différence entre deux observations,

- $\hat{\sigma}_\eta^2 = \hat{\sigma}_\varepsilon^2 - \hat{\sigma}_\xi^2$, la composante d'hétérogénéité inobservable individuelle, bornée inférieurement à une petite constante pour garantir une covariance positive.

Dans un second temps, ces trois systèmes d'équations sont combinés dans une estimation unique par moindres carrés généralisés (GLS), implémentée sous forme d'un empilement pondéré. Concrètement, chaque sous-échantillon (ventes simples, ventes répétées, et différences de ventes répétées) est pré-multiplié par l'inverse de la racine carré de l'erreur qui lui est associée.

La solution GLS agrège ainsi les effets d'année, les effets hédoniques et les écarts liés aux ventes répétées. Les effets annuels $\hat{\gamma}_t$ extraits de cette estimation, combinés aux caractéristiques moyennes observées sur l'année de base \bar{X}_{2014} , permettent ensuite de construire un indice hédonique hybride annuel rebasé à 100 en 2014 :

$$\text{Indice}_t = \exp\left(\gamma_t + \bar{X}'_{2014}\hat{\beta}\right),$$

Cette construction est répétée pour chaque segment, produisant ainsi un ensemble d'indices géographiquement et typologiquement différenciés. Le lecteur intéressé pourra trouver en Annexe 9 les coefficients et métriques de validation des modèles obtenus. Les statistiques sont concluantes, avec un bon ajustement, des erreurs de prédiction faibles, des coefficients temporels cohérents avec les dynamiques de prix connues et des variables explicatives avec des coefficients significatifs. Cette méthodologie permet d'obtenir les indices suivants :

FIGURE 4.1 – Indices immobiliers segmentés



Ces indices de prix présentent une dynamique proche des indices nationaux publiés par l'INSEE ou J. Friggitt sur la période 2014-2024. Cela témoigne de la cohérence des estimations hédoniques segmentées, qui parviennent à reproduire les tendances nationales de manière endogène, à partir des données de transactions.

Certaines disparités territoriales sont également bien captées. Par exemple, la Creuse affiche des rendements immobiliers faibles, tandis que Paris, en particulier sur le segment des appartements, présente une croissance plus soutenue jusqu'à 2021, puis un ralentissement post-covid.

Cependant, les données DVF ne permettent d'observer les dynamiques locales qu'à partir de 2014. Pour simuler des trajectoires longues nécessaires à l'évaluation de PVH à long terme, il est nécessaire de prolonger les séries vers le passé. Cette opération de *backcast* s'appuie sur des variables macroéconomiques connues historiquement (inflation, taux d'intérêt, population), ce qui permet de reconstituer un historique plausible, ancré dans la réalité économique. Ce *backcast* permet d'étendre l'historique des indices bien au-delà des dix dernières années, offrant ainsi une base de simulation plus robuste et mieux adaptée aux analyses de long terme. La projection vers le passé des indices créés est réalisée jusqu'à 1980.

4.1.3 Projection passée des indices

Afin de reconstruire l'historique des prix immobiliers par segment, une procédure de *backcast* est mise en place. Elle consiste à estimer la dynamique passée des indices à l'aide d'un modèle économétrique basé notamment sur les données réelles observées sur cette période. Ce modèle relie les indices hédoniques log-transformés à un ensemble de variables macroéconomiques pertinentes et disponibles sur la période. Ce prolongement rendra possible une analyse réaliste de la rentabilité des prêts à longue durée que sont les PVH.

Mise en application

Une fois les indices estimés pour chaque segment, ceux-ci sont prolongés vers le passé afin d'obtenir une série historique cohérente sur une période plus longue (1980–2014). Cette opération de *backcast* repose sur une régression linéaire pénalisée de type Elastic Net, où la variable cible est le logarithme de l'indice immobilier local.

Le choix de l'Elastic Net se justifie par la nature du problème : il s'agit d'un modèle reposant sur des variables explicatives corrélées entre elles. L'Elastic Net combine les avantages du Lasso³ (sélection de variables et régularisation) et du Ridge⁴ (stabilité en présence de colinéarité), ce qui le rend particulièrement adapté dans ce contexte. Contrairement à une régression par une régression par les moindres carrés ordinaires, l'Elastic Net permet :

- une meilleure robustesse de la prévision,
- un compromis optimal entre parcimonie et stabilité des coefficients,
- et une résistance accrue au sur-apprentissage, particulièrement importante dans un contexte de projection hors échantillon.

Ce choix est également justifié empiriquement par des tests comparatifs sur la période observée, qui ont montré une meilleure performance de l'Elastic Net (bien que mauvaise) en termes de stabilité des indices simulés et de reproduction des dynamiques nationales.

Pour chaque segment, la régression est effectuée à partir de variables macroéconomiques disponibles au niveau national et départemental : inflation, taux d'intérêt réels (court et long terme), population nationale, population du département concerné, ainsi que le rendement immobilier national. L'ensemble des variables explicatives est standardisé avant estimation. La régression Elastic Net⁵ combine deux types de pénalisations :

$$\mathcal{L}(\beta) = \|y - X\beta\|^2 + \alpha \left((1 - \rho)\|\beta\|^2 + \rho\|\beta\|_1 \right),$$

où :

- α contrôle l'intensité globale de la pénalisation,
- $\rho \in [0, 1]$ contrôle la balance entre ℓ_1 et ℓ_2 : $\rho = 1$ pour un Lasso pur et $\rho = 0$ pour un Ridge pur. La relation $\lambda_1 = \alpha \cdot \rho$ et $\lambda_2 = \alpha \cdot (1 - \rho)$ permet de retrouver l'équation (.7). Ces hyperparamètres sont sélectionnés automatiquement via validation croisée sur les années observées (2014–2024). Le lecteur intéressé pourra trouver la liste des paramètres finaux en Annexe 10.

3. Le lecteur intéressé pourra trouver en Annexe.7 une description de la régression Lasso.

4. Le lecteur intéressé pourra trouver en Annexe.7 une description de la régression Ridge.

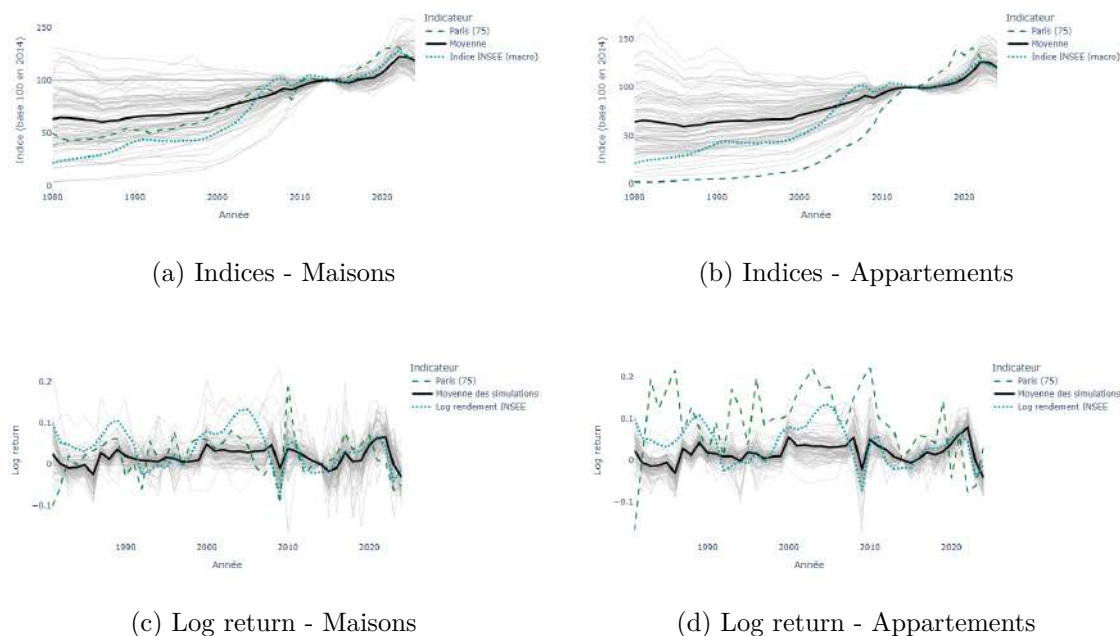
5. Reparamétrée ici comme dans Elastic Net de *scikit-learn* python ; Les principes théoriques de cette méthode, ainsi que de la régression Lasso et Ridge sont décrits en Annexe.7.

Premiers résultats et limites

La capacité prédictive du modèle, entraîné sur les données historiques de 2016 à 2024 et testé sur les années 2014-2015, s'est révélée être mauvaise, puisqu'elle est inférieure à celle d'un modèle prédisant seulement la moyenne. Les résultats détaillés en annexe (c.f. Annexe 10) montrent des performances faibles en termes de R^2 (contre une RMSE faible) et mettent en évidence une mauvaise aptitude du modèle à généraliser en dehors de l'échantillon d'entraînement. Cette contre-performance s'explique notamment par la taille réduite du jeu de données : seules 10 années sont disponibles pour l'apprentissage et le test, ce qui rend le calibrage robuste extrêmement difficile. Par ailleurs, les variables explicatives disponibles à la maille départementale en *open-source* sont limitées. Cette configuration induit un fort biais dans les résultats présentés ci-après. En effet, la moyenne nationale des projections s'écarte de la moyenne observée, traduisant un décalage systématique dans le niveau simulé.

Pour tenter d'améliorer les performances de prédiction, plusieurs modèles alternatifs ont été envisagés, tels que la régression linéaire simple ou encore des séries temporelles avec variables exogènes. Ces approches offraient une certaine simplicité ou des avantages théoriques, comme une solide prise en compte de la dynamique temporelle ou de l'influence explicite des facteurs macroéconomiques. Toutefois, elles n'ont pas été approfondies dans ce travail, leur pertinence étant limitée par la forte corrélation entre les rares variables explicatives disponibles pour la régression, et par les hypothèses structurelles très restrictives imposées par les modèles de séries temporelles, peu compatibles avec la complexité du marché immobilier.

FIGURE 4.2 – Indices *backcastés* et rendements associés



Malgré la faiblesse des métriques de validation du modèle, les dynamiques construites restent intéressantes : les grandes tendances historiques et structurelles sont reproduites, avec une phase haussière spéculative sur 1985-90, un ralentissement autour de la crise de 2008 et un rendement plus fort sur les zones tendues. Ce type de lecture qualitative importe ici, dans le cadre de futurs GSEs. Habituellement, leur calibrage est fait à partir d'un historique réel observé ; ici, du fait du manque de données, cet historique est simulé, considéré alors comme

le référent réaliste de la projection.

Compte tenu de la mauvaise capacité prédictive (passée) du modèle, l'usage de ce dernier constitue l'une des principales limites de cette étude. Les résultats présentés par la suite doivent être interprétés avec prudence. L'obtention de performances plus solides nécessiterait un historique de données plus long à une maille aussi fine, ainsi que l'accès à des variables macroéconomiques complémentaires à l'échelle départementale. La faible profondeur des données immobilières françaises en *open-source* demeure un frein majeur à la mise en œuvre rigoureuse de ce type de modélisation.

On constate que les indices projetés dans le passé tendent à être surestimés dans l'ensemble ; les rendements construits sont donc sous-estimés. Il est nécessaire de retraiter les chroniques obtenues, de manière à assurer la comparabilité des indicateurs calculés sur le portefeuille de prêts viagers hypothécaires (PVH) selon différentes méthodes. L'approche utilisée consiste à ajuster les dynamiques pour qu'elles soient crédibles économiquement, tout en corrigeant les biais de niveau pour rester aligné avec les indices INSEE observés. Cette méthodologie permet ainsi de concilier rigueur de modélisation et contraintes opérationnelles liées à la rareté des données fines.

Alignement avec l'Indice INSEE et réduction des extrêmes

Une fois les indices immobiliers segmentés prolongés dans le passé à l'aide du *backcast* tel que présenté en figure 4.2, une étape de correction complémentaire est introduite afin d'assurer la cohérence agrégée avec les données nationales, tout en atténuant les segments extrêmes. L'objectif est notamment d'éviter une surestimation persistante des indices déjà élevés — même lorsque leur dynamique est potentiellement justifiée — et inversement, afin de préserver une structure de rendement économiquement plausible.

Cette correction prend d'autant plus de sens étant donné le lien avec les GSEs futurs. Compte tenu du *payoff* asymétrique du PVH, les scénarios extrêmes sur l'immobilier ont un effet déséquilibré sur la LTV optimale calculée. À ce titre, il est préférable de privilégier une légère convergence vers la moyenne nationale observée, plutôt que de conserver des écarts trop marqués susceptibles d'amplifier artificiellement la dispersion des scénarios simulés. Ce retraitement s'inscrit d'ailleurs dans une démarche de prudence puisque le rendement moyen est alors diminué.

Concrètement, la correction appliquée introduit une phase intermédiaire de lissage partiel, permettant de rapprocher les indices extrêmes de la moyenne nationale, sans détruire la dynamique relative des séries. Pour chaque année t , la moyenne pondérée par la population départementale (proxy en l'absence de données publiques sur les volumes de ventes par département sur l'historique pertinent) \bar{I}_t est calculée telle que :

$$\bar{I}_t = \sum_d w_{d,t} \cdot I_{d,t}, \quad \text{avec } w_{d,t} = \frac{\text{Pop}_{d,t}}{\sum_d \text{Pop}_{d,t}}.$$

Le choix d'une pondération par la population départementale permet de refléter le poids économique relatif de chaque territoire dans l'indice national. Un lissage est ensuite appliqué de manière à faire tendre chaque indice désagrégé vers cette moyenne, via un paramètre α compris entre 0 et 1 :

$$I_{d,t}^{\text{lissé}} = \alpha \cdot I_{d,t} + (1 - \alpha) \cdot \bar{I}_t.$$

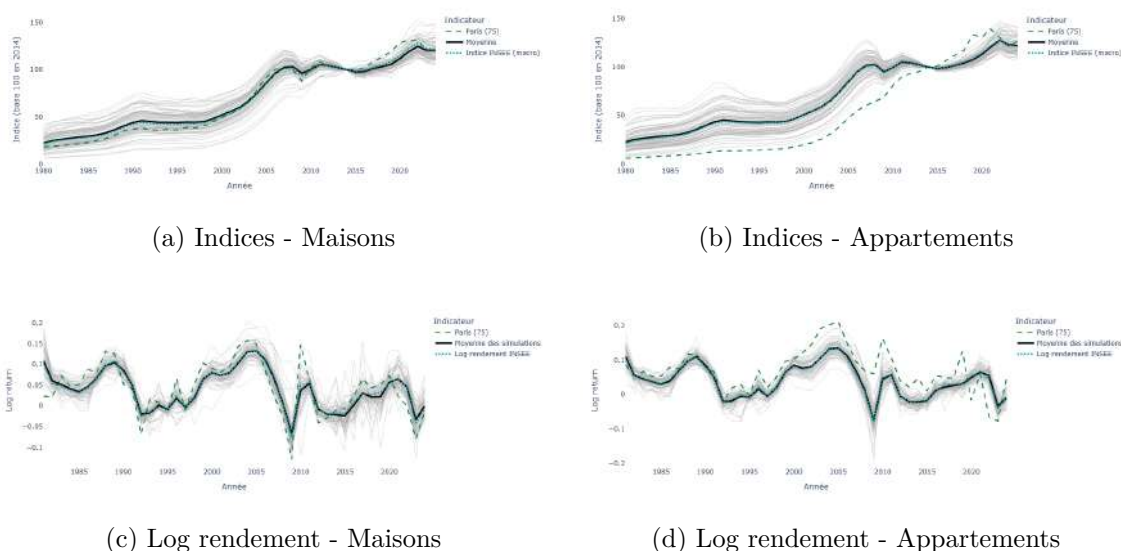
Ce lissage permet de réduire la dispersion des indices tout en conservant leur dynamique relative. Un α proche de 1 conserve fortement les dynamiques locales ; un α plus faible rapproche davantage les segments du profil moyen national. Après plusieurs tests, allant de 25 à 75%, ce paramètre est fixé à 50%, permettant de conserver des écarts conséquents en terme de rendement sans impliquer de moyennes long terme trop élevées lors du calibrage des modèles de projection.

Enfin, pour garantir la cohérence agrégée avec l'indice INSEE, un facteur d'alignement final est utilisé :

$$\rho_t = \frac{I_{INSEE,t}}{\sum_d w_{d,t} \cdot I_{d,t}^{lissé}}, \quad \Rightarrow \quad I_{d,t}^{corrigé} = \rho_t \cdot I_{d,t}^{lissé}.$$

Cette méthode est appliquée uniquement aux années du *backcast* (ante-2014) en conservant une base 100 en 2014. Elle permet de concilier deux exigences complémentaires : la fidélité aux dynamiques locales issues du *backcast* et la cohérence globale avec les statistiques nationales.

FIGURE 4.3 – Indices *backcastés* et rendements associés - Alignés avec l'indice INSEE

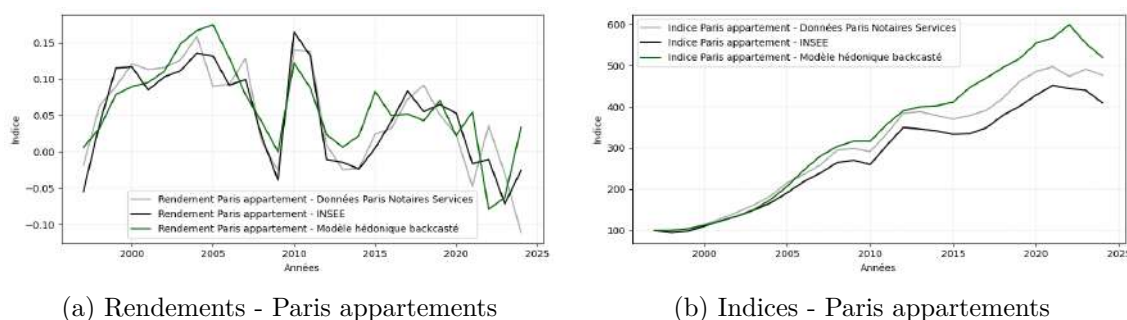


Focus Paris - appartement

Après avoir rapproché les indices construits de l'indice national global, il est important de s'arrêter à la maille segment. L'INSEE publie certains indices segmentés ; certains sont très récents (2010 pour l'Auvergne-Rhône-Alpes par exemple) d'autres plus profonds, comme l'indice Paris - Appartements, publié depuis 1992. Cet indice, ainsi que les données Paris Notaires Services de prix/m² à Paris⁶, de 1996 à 2024, servent de comparateurs :

6. [Paris Notaires Services, 2024].

FIGURE 4.4 – Comparaisons des rendements et indices



Ces graphiques montrent que le modèle capture correctement la tendance des rendements annuels, pour le segment parisien. En revanche, lorsque ces rendements sont cumulés sur une longue période (30 ans), un écart important apparaît entre le modèle et les données observées, de l'ordre de 50 à 100 points de rendement sur 30 ans. Cela suggère que de faibles écarts annuels s'accumulent significativement dans le temps. En fin d'étude, et en guise de marge de prudence, les rendements de certains segments (dont Paris fait partie) ont fait l'objet d'un retraitement, dont le rationnel est décrit en section 4.5.

Il convient de garder à l'esprit que la reconstruction d'indices à une maille fine peut engendrer un biais non négligeable. Néanmoins, ces indices constituent une base précieuse pour le calibrage des modèles de projection présentés ci-après, sous réserve d'une attention particulière portée à leur qualité et à leur cohérence.

4.2 Modélisation via GSEs segmentés

Cette section décrit la mise en place de la projection des rendements immobiliers à une maille fine à partir de GSEs de type Ahlgrim simplifiés (sans action). La section 3.4.1 présente en détails la construction théorique du modèle d'Ahlgrim, dont la méthodologie est reprise pour cette modélisation fine.

L'utilisation d'un ensemble de GSEs segmentés se justifie par sa capacité à produire des scénarios macroéconomiques stochastiques cohérents sur le long terme, tout en préservant une structure économétrique reliant explicitement les variables exogènes (taux d'intérêt, inflation) aux rendements immobiliers. Cette approche s'inscrit dans la continuité de son utilisation à l'échelle nationale dans le cadre central.

4.2.1 Projection des indices, branchement aux GSEs

Pour garantir la cohérence des projections d'indices immobiliers, l'intégration de variables macroéconomiques est nécessaire. Elle permet de prendre en compte les liens entre conjoncture nationale et rendements immobiliers segmentés, en combinant tendances globales et dynamiques locales propres à chaque segment.

Cette section décrit l'approche retenue pour modéliser les variables économiques et établir les liens de dépendance entre facteurs économiques et rendements immobiliers à un niveau fin.

Approche de modélisation des variables économiques

Dans le cadre de la génération de scénarios économiques stochastiques pour les segments département \times type de bien (182 segments⁷), le schéma de dépendance entre les facteurs

7. Etant donné la non prise en compte de l'Alsace (67 et 68), la Moselle (57) et de la Corse (20 codé comme 2A et 2B) dans cette étude, 4 départements sont retirés des 95 départements métropolitains.

simulés est central.

Plusieurs structures ont été envisagées pour modéliser la dépendance entre les facteurs macroéconomiques (inflation, taux courts, taux longs) et les rendements immobiliers simulés par segments :

- Approche 1 : utiliser un unique GSE incluant les 182 segments immobiliers et les 3 variables économiques,
- Approche 2 : construire 182 générateurs de scénarios, chacun incluant les 3 variables macroéconomiques et le rendement immobilier d'un segment donné,
- Approche 3 : utiliser un autre type de modèle qu'Ahlgrim, de type VAR ou VARX.

La 1^{ère} méthode impose une dépendance globale entre tous les facteurs. Sous cette approche, la simulation des résidus issus des calibrages des processus de Vasicek s'appuie sur une unique matrice de corrélation, de taille 185×185 . Ces résidus sont alors projetés via une décomposition de Cholesky, de la même manière que décrit pour la modélisation homogène (c.f. 3.4.6). Cependant, la structure de la décomposition de Cholesky induit une hiérarchisation implicite dans la génération des variables simulées : les premières variables (partie haute de la matrice triangulaire inférieure) conservent une pleine variance aléatoire, tandis que les dernières héritent d'une partie de leur dynamique via des combinaisons linéaires fortement contraintes par les précédentes. Sur une matrice de très grande dimension, cette structure conduit à une atténuation progressive de la variance résiduelle. Concrètement, cela se traduit par une dynamique irréaliste des trajectoires simulées pour les dernières variables : la dispersion intra-segment s'effondre, rendant les scénarios simulés non cohérents.

La 2^{ème} approche permet de contourner ce point bloquant. Chaque GSE inclut uniquement les trois variables macroéconomiques et un unique segment immobilier. En forçant les résidus macroéconomiques à être identiques pour chaque bloc de simulation, l'inflation, les taux courts et les taux longs restent cohérents inter-segments. Ainsi, une dépendance cohérente des trajectoires immobilières vis-à-vis des chocs macroéconomiques est garantie. Cette approche permet de concilier la cohérence des simulations à la maille nationale avec la granularité requise pour une analyse fine des risques immobiliers, en assurant que les taux d'intérêt et l'inflation évoluent de manière homogène sur l'ensemble du territoire français.

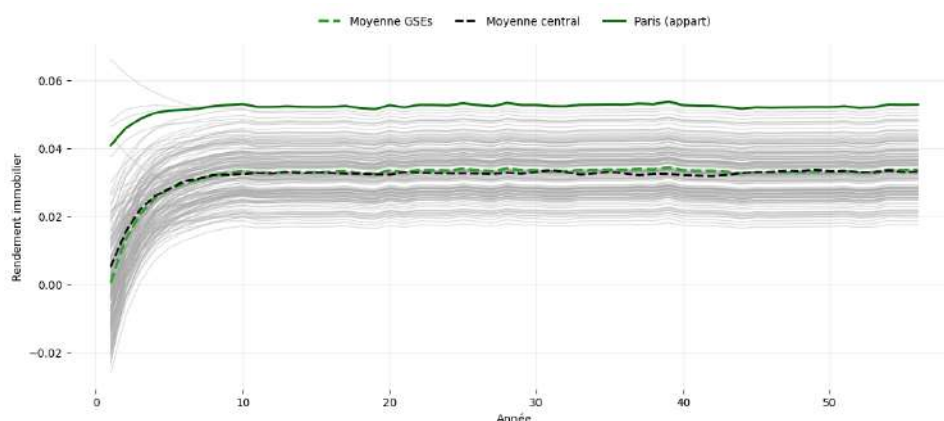
La 3^{ème} méthode consiste à utiliser un modèle multivarié dynamique, tel qu'un VARX (*Vector AutoRegressive model with exogenous variables*), pour modéliser directement les dépendances temporelles et les interactions entre facteurs macroéconomiques et rendements immobiliers. Cette approche permet de prendre en compte de manière explicite les effets de rétroaction et les relations croisées entre variables. Toutefois, son calibrage devient rapidement complexe et instable lorsque l'horizon historique est long et que le nombre de segments à modéliser est élevé, ce qui en limite la faisabilité dans le contexte de cette étude.

Après avoir testé la 1^{ère} approche, et poussé par la volonté de garder la même structure de modèle qu'en central, la 2^{ème} approche a été sélectionnée. Ainsi, dans le cadre d'application de cette étude, la dynamique immobilière de chaque segment est évaluée sur la base de scénarios économiques cohérents, mais indépendants entre segments, ce qui permet de capter l'asymétrie de la valorisation, liée à la NNEG.

Paramètres, corrélation et cohérence

La figure 4.5 met en évidence les différences présentes au sein des moyennes des rendements simulés. La moyenne nationale reconstruite est de l'ordre de 3,30%, et est atteinte autour de la 10^{ème} année de projection, en cohérence avec les rendements obtenus en central (c.f. 3).

FIGURE 4.5 – Rendement immobilier moyen, par segment



Les variables macroéconomiques de taux et d'inflation sont simulées à partir des mêmes paramètres de calibrage que ceux définis dans l'environnement central. Par ailleurs, les résidus aléatoires utilisés dans la génération stochastique sont identiques entre segments pour ces variables. Cela garantit une comparabilité directe des trajectoires simulées pour ces variables, pour chaque segment. La structure de corrélation influe légèrement sur les trois variables de taux, du fait de l'application de la décomposition de Cholesky. Toutefois, cette influence reste marginale, et les distributions (moyennes et quantiles) observées en sortie demeurent très proches de celles de l'environnement central.

Pour l'immobilier, bien que les moments globaux (moyenne, variance, quantiles) restent comparables à ceux de l'environnement central lorsqu'on agrège tous les segments, cette méthode permet de faire apparaître des disparités de rendements à l'échelle segmentée, ce qui est conforme aux objectifs. Ces disparités reflètent les hétérogénéités structurelles souhaitées. A noter que les distributions de rendements immobiliers de certains segments sont présentés lors des analyses de résultats (c.f. 4.18).

Les scénarios obtenus sont ensuite intégrés dans le moteur de tarification afin de produire une LTV pour chaque individu. Les indicateurs du portefeuille de PVH obtenus par cette approche seront par la suite confrontés à ceux issus des autres méthodes, dont l'approche par réseaux de neurones, détaillée dans la section suivante.

4.3 Modélisation via réseau de neurones récurrent

Après avoir exploré la modélisation des rendements immobiliers via des générateurs de scénarios économiques segmentés, une approche alternative basée sur les réseaux de neurones est proposée. Ces modèles, et notamment les réseaux de neurones récurrents (RNN, pour *Recurrent Neural Network*), sont particulièrement adaptés à la modélisation de séries temporelles complexes grâce à leur capacité à mémoriser l'information sur plusieurs pas de temps.

Parmi les différentes architectures de RNN, celle des *Long Short-Term Memory* (LSTM) est largement reconnue pour son efficacité à capturer des dépendances à long terme dans les données séquentielles⁸. L'utilisation des LSTM pour projeter les rendements immobiliers segmentés permet de modéliser non seulement les tendances historiques, mais aussi l'impact dynamique des variables macroéconomiques sur ces rendements. Cette méthode offre ainsi

8. [S. Hochreiter, 1997] et [K. Greff et al., 2015].

une approche flexible et puissante pour capturer la complexité temporelle et segmentée des données immobilières.

4.3.1 Concepts généraux et application à l'immobilier

Inspirés du fonctionnement biologique du cerveau humain, ces modèles ont démontré une capacité à capturer des structures latentes, non observables, dans des données massives, bruitées ou non linéaires. En actuariat, les réseaux de neurones, et particulièrement les RNN, connaissent une importance croissante, comme en témoignent plusieurs travaux récents réalisés dans le cadre de mémoires IA :

- [P. Ottou, 2017] *Méthodes d'apprentissage automatique appliquées au provisionnement ligne à ligne en assurance non-vie*,
- [T-S. Goma, 2020] *Apport des réseaux neuronaux au provisionnement IARD : Application d'architectures récurrentes à la prédiction d'évolution temporelle de montants*,
- [A. Toure, 2021] *Modélisation des obligations catastrophes paramétriques dans un marché incomplet par une approche deep learning et une approche de modélisation par loi*,
- [A. Morisse, 2022] *Application des réseaux de neurones récurrents à l'estimation des calculs réglementaires*,
- [N. Richard, 2022] *Cash flow prediction in reinsurance : short and long term projection for cash requirement and FX risk management*,
- [T. Rouer, 2023] *Estimation de la densité de provisions non-vie ligne à ligne par réseaux de neurones*.

Ces travaux ont largement nourri la rédaction de la section théorique présentée ci-dessous. Ils démontrent aussi l'intérêt croissant pour ces méthodes, d'autant plus que de nombreux autres travaux et mémoires se basent sur cet outil, pour des sujets actuariels mais aussi tout autres, tels que la médecine, l'analyse de texte ou le domaine musical. Plus largement encore, les réseaux de neurones constituent aujourd'hui un pilier central du champ de l'apprentissage statistique, comme l'illustre le cours de l'EURIA [F. Vermet, 2025] *Apprentissage statistique supervisé* ou du CNAM [M. Crucianu et al., 2023] *Apprentissage, réseaux de neurones et modèles graphiques*, utilisés comme références ici.

Les réseaux de neurones artificiels sont des modèles d'apprentissage supervisé puissants et largement documentés dans la littérature scientifique. Leur application au domaine immobilier se heurte toutefois à des défis spécifiques : la forte hétérogénéité des marchés, la disponibilité limitée des données et la sensibilité aux chocs macroéconomiques. [J. Meszaros, 2024], une revue de littérature récente, souligne d'ailleurs ces difficultés et met en évidence la complexité intrinsèque de la prévision des prix immobiliers : les dynamiques immobilières combinent à la fois une forte inertie à long terme et une volatilité marquée à court terme, ce qui rend difficile l'obtention de modèles robustes et prédictifs. Les méthodes traditionnelles, qu'elles soient économétriques (ARIMA, VAR, GARCH) ou hédoniques, captent partiellement les déterminants des prix mais échouent souvent à reproduire l'hétérogénéité spatiale, les effets de cycles et la sensibilité aux chocs macroéconomiques. De plus, la rareté des données à haute fréquence et la spécificité locale des marchés accentuent ces difficultés. Ainsi, même les modèles les plus sophistiqués atteignent rarement des niveaux élevés de précision prédictive, ce qui explique l'émergence d'approches hybrides et de techniques d'apprentissage automatique visant davantage à générer des trajectoires cohérentes et prospectives qu'à produire des prévisions exactes.

Face aux limites des modèles classiques, la littérature récente s’est orientée vers les approches de *machine learning*, et en particulier les réseaux de neurones, afin d’améliorer la prévision des prix immobiliers. Ces méthodes permettent de capturer des relations non linéaires complexes, d’intégrer des variables exogènes et de tirer parti de la dimension séquentielle des données. Les réseaux de type LSTM (*Long Short-Term Memory*), spécialement conçus pour le traitement des séries temporelles, se distinguent par leur capacité à modéliser les dépendances de long terme, ce qui constitue un atout majeur dans le contexte de cette étude. Leur caractère itératif natif, consistant à générer des prévisions de proche en proche, est particulièrement adapté dans un cadre de projection pluriannuelle, même si cette approche présente la limite d’une propagation potentielle des erreurs.

Plusieurs publications documentent l’utilisation de modèle LSTM pour la projection d’indices financiers ou immobiliers. À titre d’exemple, [H. DiStefano, 2022] présente un modèle LSTM appliqué aux prix résidentiels aux États-Unis, montrant que, si la précision prédictive demeure limitée, cette approche permet néanmoins de reproduire de manière cohérente les cycles immobiliers et leur sensibilité aux variables macroéconomiques. De son côté, [K-C. Chiu, 2024] propose une application au marché immobilier taïwanais dans le contexte post-pandémique. Ces travaux confirment l’intérêt des architectures LSTM pour intégrer l’effet de chocs exogènes et générer des trajectoires prospectives réalistes.

L’objectif de la section suivant est de présenter la structure des réseaux de neurones et en particulier les *Long Short-Term Memory*.

4.3.2 Fondements théoriques

Réseaux de neurones classiques

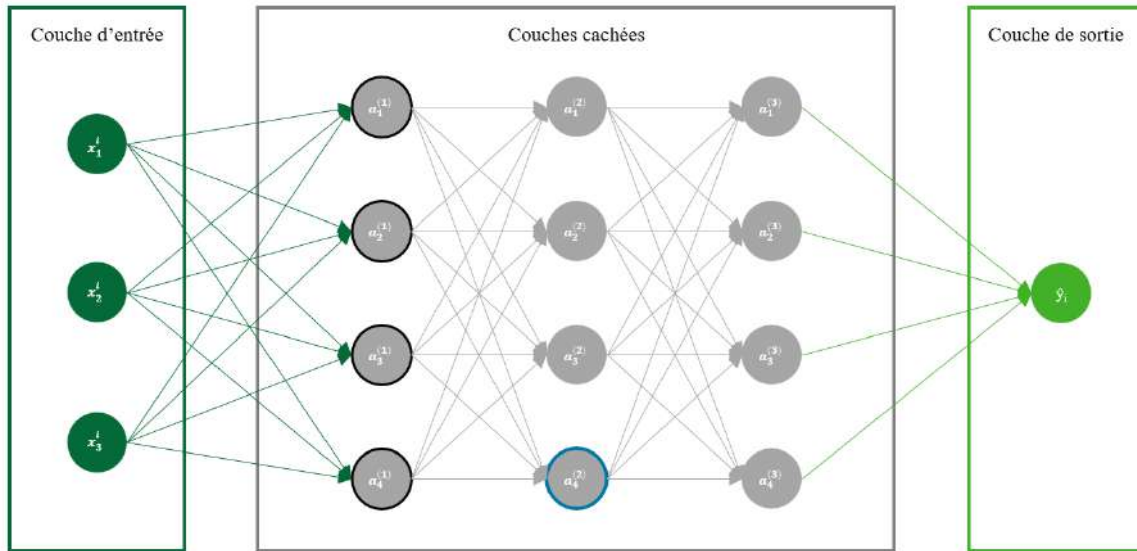
L’architecture d’un réseau de neurones⁹ dit *feed-forward* est définie par l’agencement de couches et par les connexions pondérées entre les neurones. Une couche est un niveau de traitement qui reçoit un vecteur d’activations \mathbf{x} , applique à chacun de ses n neurones la même opération linéaire $w_i^\top \mathbf{x} + b_i$, où w_i est le vecteur de poids et b_i le biais du neurone i , suivie d’une fonction d’activation φ , et renvoie le vecteur transformé $\hat{\mathbf{Y}} = (\varphi(w_1^\top \mathbf{x} + b_1), \dots, \varphi(w_n^\top \mathbf{x} + b_n))$ à la couche suivante. Un neurone, correspond précisément à cette fonction scalaire paramétrable $x \mapsto \varphi(w^\top x + b)$. Il combine de façon pondérée les sorties précédentes, ajoute un biais, puis introduit la non-linéarité par φ . Ainsi, l’architecture d’un réseau se définit par l’enchaînement ordonné de ces couches et par l’ensemble des éléments w et b qui relie chaque couche à la suivante. Un réseau de neurones est composé de trois types de couches :

- une couche d’entrée, correspondant aux variables explicatives du problème,
- une ou plusieurs couches cachées, qui construisent des représentations intermédiaires non observables,
- une couche de sortie, qui produit la prédiction du modèle.

Les figures 4.6 et 4.7 illustrent cette structure, pour un réseau à 3 variables explicatives, 3 couches cachées et 4 neurones par couches cachées :

9. Les notations mathématiques liées aux réseaux de neurones utilisées au long de cette section sont répertoriées en Annexe.8)

FIGURE 4.6 – Structure d'un réseau de neurones

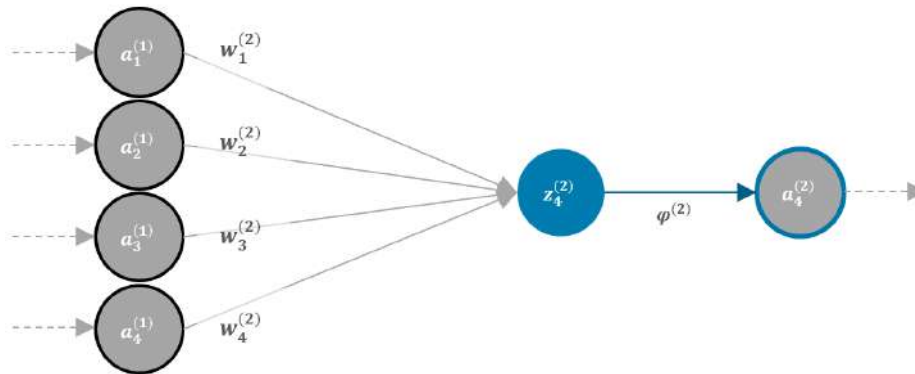


Où :

- x_k^s est la $k^{\text{ième}}$ variable d'entrée utilisée pour la prédiction de la sortie \hat{y}_s ,
- $a_k^{(i)}$ est le $k^{\text{ième}}$ neurone de la $i^{\text{ième}}$ couche.

Le fonctionnement entre deux couches suit l'architecture suivante :

FIGURE 4.7 – Focus sur le fonctionnement entre deux couches



Où :

- $w_{j,k}^{(i+1)}$ est le $j^{\text{ième}}$ poids du neurone $a_k^{(i+1)}$, soit le poids reliant la sortie du neurone $a_j^{(i)}$ au neurone $a_k^{(i+1)}$,
- $z_k^{(i+1)}$ est la pondération des sorties précédentes, sommées de biais $b_k^{(i)}$, tel que :

$$z_k^{(i+1)} = \sum_{j=1}^{n_i} w_{j,k}^{(i+1)} \cdot a_j^{(i)} + b_k^{(i)},$$

- $\varphi^{(i)}$ est la fonction d'activation de la $i^{\text{ième}}$ couche, telle que :

$$a_k^{(i+1)} = \varphi^{(i+1)}(z_k^{(i+1)}).$$

Les fonctions d'activation classiques sont :

- ReLU : $\varphi(x) = \max(0, x)$,
- Sigmoidale : $\varphi(x) = \frac{1}{1+e^{-x}}$,
- Swish : $\varphi(x) = x \cdot \text{sigmoid}(x)$,
- Tanh : $\varphi(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$,

Lors de l'apprentissage, le modèle ajuste ces paramètres, notamment les poids et les biais, de manière à minimiser une fonction de coût qui mesure l'écart entre les prédictions du réseau \hat{y} et les observations réelles y , tel que :

- $\mathcal{L}(Y, \hat{Y})$: fonction de perte (souvent MSE en régression),
- $J(\theta) = \frac{1}{m} \sum_{i=1}^m \mathcal{L}(y^{(i)}, \hat{y}^{(i)})$: fonction objectif à minimiser, où θ représente l'ensemble des paramètres du réseau, soit $\theta = \{W^{[l]}, b^{[l]}\}_{l=1}^L$ avec L le nombre de couches.

Structure globale du réseau Après avoir détaillé le fonctionnement d'un réseau de neurones, il est utile de revenir à une vue d'ensemble. Le réseau complet se comporte comme une fonction de la forme :

$$\hat{f}(X; \theta) = \varphi^{[L]}(\dots \varphi^{[2]}(\varphi^{[1]}(XW^{[1]} + b^{[1]}))) = A^{[L]},$$

où $A^{[l]}$ désigne l'activation de la $l^{\text{ème}}$ couche, et $A^{[L]}$ l'activation de la couche de sortie, soit \hat{y} . L'apprentissage consiste à minimiser une fonction de perte moyenne, en ajustant les paramètres θ :

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m \mathcal{L}(y^{(i)}, \hat{f}(x^{(i)}; \theta)),$$

généralement par descente de gradient « vanille » ou par une variante plus sophistiquée (Adam, RMSProp, *etc.*). La section suivante s'attarde sur les principes théoriques liés à la descente de gradient.

Descente de gradient et mise à jour des poids

L'apprentissage d'un réseau de neurones repose sur l'ajustement des poids $\theta = \{W^{[l]}, b^{[l]}\}$ de manière à minimiser une fonction de perte $J(\theta)$, qui mesure l'écart entre les prédictions du modèle \hat{y} et les observations réelles y . Ce processus repose sur un algorithme d'optimisation appelé descente de gradient consistant à calculer, pour chaque paramètre θ_j , la dérivée partielle de la fonction de coût :

$$\frac{\partial J(\theta)}{\partial \theta_j}$$

et à mettre à jour ce paramètre dans la direction opposée au gradient, selon la règle :

$$\theta_j \leftarrow \theta_j - \alpha \cdot \frac{\partial J(\theta)}{\partial \theta_j},$$

où $\alpha > 0$ est le taux d'apprentissage (*learning rate*). Le calcul des gradients repose sur deux étapes fondamentales :

- Propagation directe : à partir des entrées $x^{(i)}$, on calcule les activations successives $A^{[1]}, \dots, A^{[L]}$ jusqu'à obtenir la prédiction $\hat{y}^{(i)}$,

- Rétropropagation de l'erreur : l'erreur est propagée en sens inverse à travers le réseau en appliquant la règle de la chaîne, ce qui permet de calculer $\frac{\partial J}{\partial W^{[\ell]}}$ et $\frac{\partial J}{\partial b^{[\ell]}}$ pour chaque couche ℓ . Une présentation détaillée de l'algorithme de rétropropagation du gradient, incluant ses fondements mathématiques et ses implications numériques, est présentée dans le cours [F. Vermet, 2025] *Apprentissage statistique : une approche connexionniste*, dispensé en Master 1 à l'EURIA.

De nombreuses variantes du schéma de descente de gradient classique ont été développées pour améliorer la vitesse de convergence et la stabilité de l'apprentissage :

- Momentum : accélère la descente dans les directions stables en tenant compte de l'historique des gradients,
- RMSProp : adapte dynamiquement le taux d'apprentissage à chaque paramètre en normalisant par la moyenne quadratique des gradients récents,
- Adam¹⁰ : combine les avantages du momentum et de RMSProp.

La descente de gradient constitue donc le cœur de l'apprentissage d'un réseau de neurones, quel que soit son type d'architecture. Toutefois, la structure du réseau influence profondément la nature des dépendances capturées dans les données. Les modèles évoqués jusqu'ici, dits *feed-forward*, supposent que chaque observation est indépendante et traitée de façon isolée, sans mémoire des précédentes. Cette hypothèse est souvent restrictive, notamment lorsque les données présentent une forte dimension temporelle ou séquentielle, comme c'est le cas des rendements immobiliers.

Les réseaux de neurones récurrents (RNN) remédient à cette limite en ajoutant un état qui s'actualise à chaque pas de temps, donnant ainsi au modèle une mémoire dynamique des observations passées. Les difficultés d'optimisation qui en découlent ont conduit à l'apparition de cellules spécialisées, dont la plus répandue est la *Long Short-Term Memory* (LSTM). De nombreux hyperparamètres rentrent en jeu dans le cadre de ces modèles.

Hyperparamètres En effet, la mise en œuvre d'un réseau de neurones dépend d'un ensemble d'hyperparamètres, qui ne figurent pas dans les équations internes mais pilotent la dynamique d'apprentissage. Cette section énumère ces hyperparamètres essentiels et présente leur influence ainsi que les stratégies usuelles pour les sélectionner :

- Taux d'apprentissage α : le *learning rate* contrôle l'amplitude de la mise à jour des poids lors de la descente de gradient. Un α trop grand provoque des oscillations ou un échec de convergence ; trop petit, il ralentit l'entraînement et risque le blocage dans des minima locaux,
- Coefficient de régularisation λ : Ce paramètre pénalise la norme des poids pour éviter le surapprentissage,
- Nombre de couches cachées L : La profondeur détermine le niveau d'abstraction et la complexité,
- Taille de la couche ℓ , t_ℓ : Nombre de neurones de la couche cachée ℓ . Une décroissance ou plateau ($t_1 \geq t_2 \geq \dots \geq t_c$) sont généralement respectés pour compresser progressivement l'information,
- Taux de *dropout* p : Probabilité de mettre à zéro l'activation d'un neurone à l'entraînement ; sert à réduire la co-adaptation¹¹ des poids,

10. Optimisateur utilisé lors de la mise en application de cette méthode.

11. La coadaptation représente le fait que certains neurones apprennent à "travailler ensemble" de façon trop spécifique, en s'appuyant les uns sur les autres pour corriger les erreurs, au lieu d'apprendre des représentations individuellement robustes. Cela nuit à la généralisation et certains neurones deviennent inutiles seuls.

- Taille des *batches* B : Nombre d'exemples traités avant chaque mise à jour des poids,
- Nombre d'*epoch* E : Nombre de passages complets sur le jeu d'apprentissage. E doit être assez grand pour atteindre la convergence, mais un arrêt tôt (*early stopping*) est nécessaire, dès que la perte de validation cesse de diminuer afin de prévenir le sur-apprentissage.

Réseaux de Neurones Récurrents

Les réseaux de neurones classiques (*feed-forward*) sont peu adaptés aux données séquentielles. En effet, dans ce type d'architecture, chaque observation est traitée de manière indépendante, sans tenir compte de l'ordre ou des dépendances temporelles qui peuvent exister entre elles. Or, dans de nombreuses applications l'information pertinente ne réside pas seulement dans la valeur instantanée d'une variable, mais aussi dans son évolution et ses interactions avec le passé. Ainsi, ce type de réseau, appliqué à une série temporelle, se limite à considérer un instant donné comme une photographie isolée, perdant la richesse de la dynamique historique et se révélant incapable de capturer les dépendances à court ou long terme. Pour pallier cette limite, une famille particulière de réseaux a été développée : les réseaux de neurones récurrents (RNN).

RNN 'classiques' Contrairement aux architectures classiques, les RNN conservent en mémoire une trace des étapes précédentes. À chaque instant t , le réseau ne se contente pas de traiter l'entrée courante x_t , mais tient également compte d'un état caché h_{t-1} , hérité de l'instant précédent. Ce mécanisme permet au modèle d'intégrer une dynamique temporelle et de mieux représenter les relations successives entre observations. Mathématiquement, ce fonctionnement se décrit par :

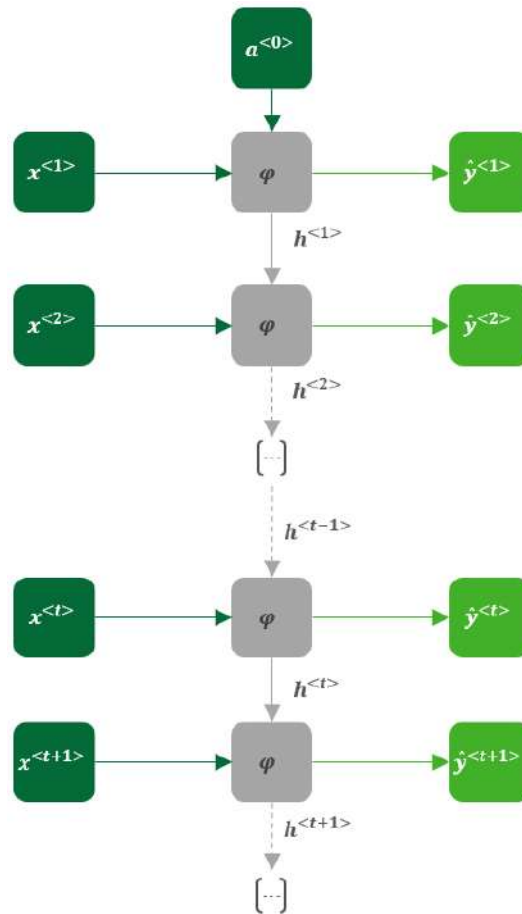
$$h_t = \phi(W_h h_{t-1} + W_x x_t + b), \quad \hat{y}_t = \varphi(W_y h_t + b'),$$

où ϕ est la fonction d'activation de l'état caché et φ est la fonction d'activation de sortie, W_h , W_x , W_y les matrices de poids, b et b' les vecteurs de biais, et \hat{y}_t la sortie en t .

Ce type de modèle peut être *one-to-one*, *many-to-one*, *one-to-many* ou *many-to-many* ; ces termes décrivent la cardinalité entrée-sortie¹². Schématiquement, un RNN *many-to-many* peut se représenter de manière simplifiée par :

12. [P. Ottou, 2017].

FIGURE 4.8 – Structure d'un réseau de neurones récurrent



Pour n couches, un RNN *many-to-one* verrait les cellules $\hat{y}^{<1>} \dots \hat{y}^{<n-1>}$ effacées, laissant seule $\hat{y}^{<n>}$.

La conservation d'informations passées permet au réseau d'emmagasiner une représentation latente de l'historique. Néanmoins, l'apprentissage de dépendances lointaines dans ces RNN classiques se heurte à des phénomènes de *vanishing* et *exploding gradients* lors de la rétropropagation à travers le temps, rendant la convergence hasardeuse et limitant la prise en compte d'événements éloignés (c.f. Annexe.8). Ces difficultés ont conduit au développement d'architectures plus sophistiquées, telles que les LSTM.

LSTM Il est courant d'observer un effet de latence de l'information dans les séries temporelles, traduisant l'existence de dépendances à long terme. Les RNN classiques peinent à les retenir, leurs gradients ayant tendance à s'annuler ou à exploser au fil des rétropropagations. Les *Long Short-Term Memory* (LSTM) ont justement été conçus pour résoudre ce problème (c.f. [A. Graves, 2012]). Ils intègrent une cellule mémoire persistante, associée à trois portes qui dosent dynamiquement l'information qui entre, persiste ou sort de cette mémoire :

- *Forget gate* f_t : efface sélectivement des informations de c_{t-1} ,
- *Input gate* i_t : décide des nouvelles informations (\tilde{c}_t) à stocker,
- *Output gate* o_t : filtre la part de c_t transmise à la sortie h_t .

Ce dispositif permet de contrôler le passage des signaux tout en maintenant des gradients stables durant l'apprentissage. À chaque étape temporelle t , l'unité LSTM met à jour deux états :

$$\begin{aligned}
\text{porte d'entrée : } & i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i), \\
\text{porte d'oubli : } & f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f), \\
\text{candidat mémoire : } & \tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c), \\
\text{mémoire : } & \boxed{c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t}, \\
\text{porte de sortie : } & o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o), \\
\text{état caché : } & h_t = o_t \odot \tanh(c_t).
\end{aligned}$$

La mémoire c_t est donc une combinaison éléments par éléments de f_t et c_{t-1} , qui retient ou oublie l'information passée selon la porte d'oubli f_t et i_t et \tilde{c}_t , qui injecte les nouvelles informations pertinentes issues de l'entrée courante via la porte d'entrée i_t . La combinaison de ces deux flux, pondérés entre 0 et 1 par des sigmoïdes, prévient la disparition ou l'explosion des gradients et permet au réseau de conserver une trace stable des dépendances à long terme.

Dans un cadre génératif, un réseau récurrent de type LSTM *many-to-one* peut produire une séquence complète en procédant de manière autorégressive¹³. Concrètement, à chaque pas de temps, \hat{y}_t est réinjectée en entrée x_{t+1} afin de calculer \hat{y}_{t+1} . Ce mécanisme récursif, dit de génération autorégressive, permet de dérouler des séquences cohérentes de longueur arbitraire, le modèle « se nourrissant » de ses propres sorties pour prolonger la dynamique temporelle. Au-delà de l'aspect pratique, ce décodage autorégressif contourne plusieurs limites des approches *many-to-many* : il évite de définir une perte multi-horizon aux pondérations arbitraires, ramène l'apprentissage à une tâche 1-pas mieux conditionnée, augmente fortement le nombre d'exemples supervisés (*sliding-window*) et stabilise l'entraînement avec peu de données.

L'architecture LSTM constitue une base robuste pour la modélisation de séries temporelles économiques¹⁴. Après avoir décrit la mécanique d'un réseau de neurones, il convient désormais de préciser comment cette architecture abstraite est rendue opérationnelle sur des données réelles.

4.3.3 Mise en application

Après avoir présenté les fondements théoriques des réseaux de neurones récurrents et les motivations ayant conduit à l'utilisation des LSTM, cette section s'attarde sur sa mise en application dans le cadre précis de cette étude.

Données et architecture

L'ensemble d'indices construit en amont couvre la période 1980 à 2024, soit 45 années d'observation, sur 182 segments. Les variables utilisées en entrée sont les taux courts et longs, l'inflation et les rendements immobiliers nationaux et segmentés, qui constituent un socle macro-financier commun à l'ensemble des segments. Pour la période historique, les valeurs utilisées correspondent aux données observées, tandis que pour la période de projection, les valeurs injectées dans le modèle proviennent des trajectoires simulées en central (10 000 scénarios différents). Pour chaque segment, les observations annuelles sont transformées en séquences temporelles (*unrolling*) de taille $k = 5$ (après optimisation, valeurs testées :

13. [A. Graves, 2013].

14. Comme le montre [S-S. Namini et al., 2019] ou encore [D. Hopp, 2022].

$k \in \{5, 7, 10, 12\}$). Chaque séquence de k années (de l'année $t - (k - 1)$ à t) sert à prédire le rendement à l'année $t + 1$; c'est un réseau *many-to-one*. Cette architecture exploite efficacement la profondeur temporelle des données historiques et renforce la capacité d'apprentissage du modèle.

Pour traduire les disparités inter-segment, chaque segment est représenté par un vecteur d'*embedding* de taille $e = 9$ (après optimisation), afin de permettre l'apprentissage de dynamiques spécifiques. Cet *embedding* permet la représentation vectorielle dense des segments, où chaque coordonnée du vecteur est apprise automatiquement au cours de l'entraînement du réseau. L'*embedding* apprend ainsi une représentation continue, optimisée conjointement avec les autres paramètres du modèle. Cette approche permet d'enrichir le signal d'entrée du modèle tout en limitant le nombre de paramètres, et contribue à une meilleure généralisation.

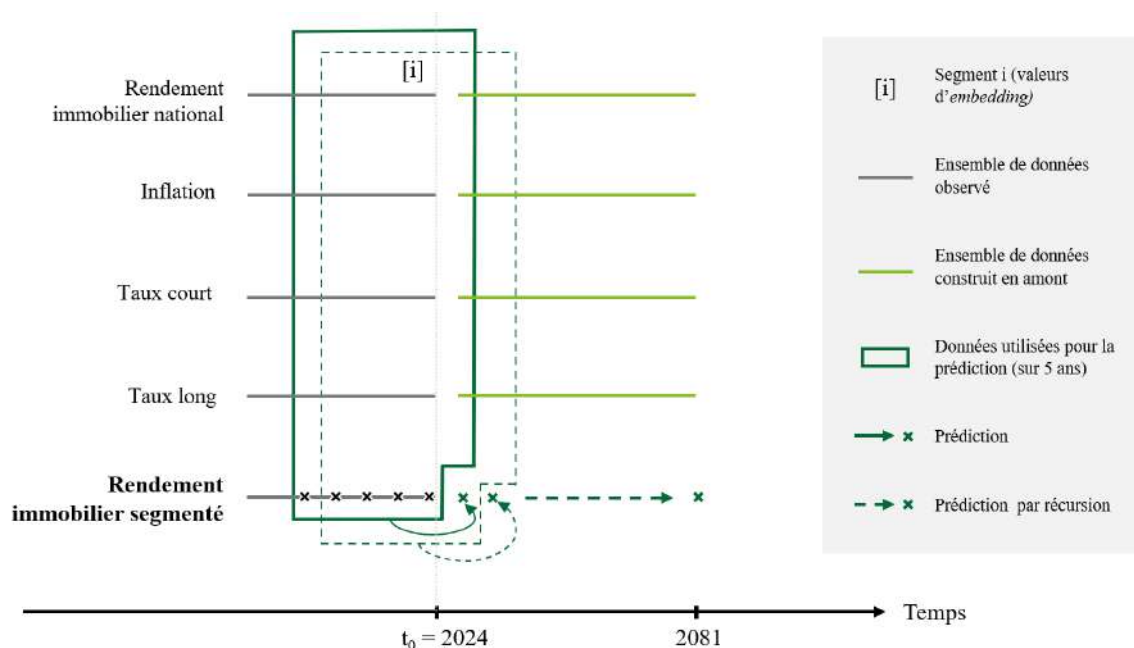
Les données sont séparées chronologiquement en deux ensembles : un jeu d'apprentissage (1980–2015) et un jeu de test (2016–2024), afin d'éviter toute fuite d'information. L'ensemble d'apprentissage est scindé en sous-ensemble de validation dans une procédure de validation croisée glissante (*sliding cross validation*). À chaque fenêtre, la standardisation est ajustée uniquement sur le jeu d'apprentissage, puis appliquée telle quelle aux jeux de validation et de test ; pour le modèle final, le *scaler* est ajusté sur 1980–2015 puis appliqué à 2016–2024. Les variables catégorielles (*embeddings*) ne sont pas standardisées.

L'architecture du LSTM développé s'organise en deux branches distinctes :

- Une branche séquentielle, recevant en entrée des séquences temporelles de 5 années de variables explicatives (k optimisé manuellement parmi les valeurs énumérées ci-avant). Ces séquences sont traitées par deux couches LSTM empilées, permettant d'extraire des représentations dynamiques de l'évolution temporelle. Un mécanisme de régularisation par *dropout* est appliqué à chaque couche pour limiter le sur-apprentissage,
- Une branche catégorielle, recevant un identifiant unique par segment, projeté dans un espace latent de dimension réduite via un *embedding*. Cette représentation dense encode les spécificités propres à chaque segment.

Les sorties des deux branches sont concaténées avant d'être transmises à une couche du réseau, qui produit la prédiction du rendement immobilier à l'année suivante.

Afin d'exploiter le modèle entraîné pour la projection à long terme, une procédure de prédiction récursive est mise en œuvre, comme schématisé en figure 4.9. Pour chaque segment, les 5 dernières années connues servent de point de départ. À chaque horizon futur, la prédiction réalisée est injectée comme entrée pour la période suivante. Les trajectoires issues du GSE central sont utilisées en entrée. Le modèle est entraîné en minimisant l'erreur quadratique moyenne, à l'aide de l'optimiseur Adam (c.f. 4.3.2). Les performances sont ensuite évaluées à l'aide des indicateurs MSE, RMSE et R^2 présentés dans la section suivante.

FIGURE 4.9 – Projection du rendement immobilier du segment i , par récursion

Optimisation du modèle, évaluation et projections

Pour améliorer l'architecture et les hyperparamètres, une sélection manuelle de la taille d'*unrolling* ainsi qu'une optimisation automatique avec la librairie *Optuna* ont été réalisées. Cette méthode de recherche adaptative sélectionne les configurations les plus prometteuses en minimisant la perte de validation sur un jeu chronologique glissant, évitant ainsi une exploration exhaustive coûteuse. Les paramètres testés sont :

- Dimension de l'*embedding* (représentation latente des segments),
- Nombre de couches,
- Nombre d'unités dans la $i^{\text{ème}}$ couche,
- Taux de *dropout* appliqué après chaque couche LSTM pour limiter le sur-apprentissage,
- Dimension de l'*unrolling* (taille des fenêtres de données en entrée) - manuellement.

L'optimisation des hyperparamètres via *Optuna* a conduit à la configuration finale utilisée, offrant un compromis équilibré entre biais et variance. Dans ce contexte, la *Loss* correspond à la fonction de coût calculée sur l'échantillon d'entraînement, mesurant l'écart moyen entre les prédictions du modèle et les valeurs réelles. La *Val_Loss* est la même mesure, mais appliquée à l'échantillon de validation, qui n'intervient pas dans la mise à jour des poids. Elle reflète la capacité du modèle à généraliser à de nouvelles données.

TABLE 4.1 – Résultats de l'optimisation par *Optuna*

Paramètre	Valeur optimale	Epoch	Loss	Val_Loss
Dimension d' <i>embedding</i>	9	1	0,0028	0,0027
Nombre de couches	2	2	0,0018	0,0023
Unités LSTM (couche 1)	32	3	0,0014	0,0020
Unités LSTM (couche 2)	32	4	0,0012	0,0018
<i>Dropout</i>	0,47	5	0,0010	0,0016
<i>Unrolling</i> (sélection manuelle)	5	6	0,0009	0,0016
		7	0,0008	0,0017
		8	0,0007	0,0017
		9	0,0007	0,0018

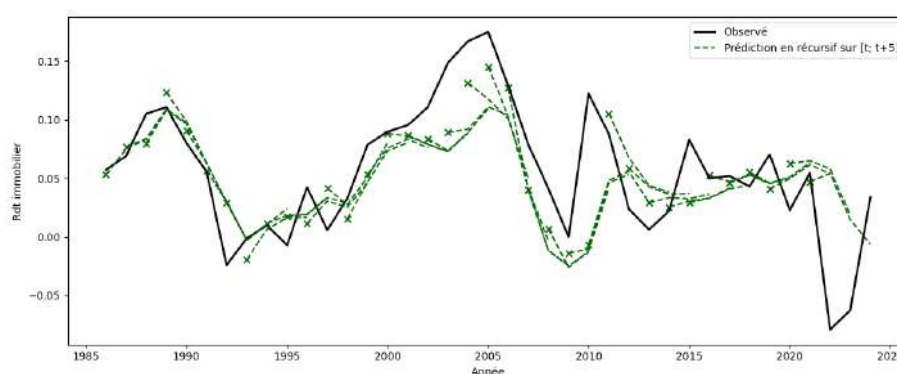
L'évaluation du modèle repose sur trois indicateurs standards, calculés en mode *pooled*, après inverse-transformation : MSE, RMSE et R^2 . Deux cadres d'évaluation complémentaires sont considérés : $t+1$ et récursif (où chaque prédiction est réinjectée pour simuler l'année suivante tout en utilisant les exogènes réelles).

TABLE 4.2 – Métriques de performance — $t+1$ *vs.* récursif

	Validation	Test ($t + 1$)	Test (récursif)
MSE	0,0006	0,0009	0,0018
RMSE	0,0253	0,0304	0,0421
R^2	0,7703	0,4234	0,3714

Ces niveaux reflètent la forte volatilité et le bruit des rendements immobiliers locaux, la courte profondeur historique par segment et le faible nombre de covariables observables. Les performances en $t+1$ sont, comme attendu, meilleures que dans le cadre récursif où les erreurs se propagent à l'horizon. L'écart entre apprentissage et test demeure contenu, signe d'une généralisation correcte et d'une régularisation efficace. L'objectif n'est pas une prévision point-à-point parfaite, mais la génération de trajectoires cohérentes par segment conditionnellement aux variables exogènes, utilisables pour la tarification future. Dans cette optique, la priorité est donnée à la robustesse et à la cohérence économique plutôt qu'au seul niveau de précision instantané. La figure suivante présente les trajectoires prédictives issues d'une simulation récursive sur l'ensemble des fenêtres de temps, pour le segment parisien :

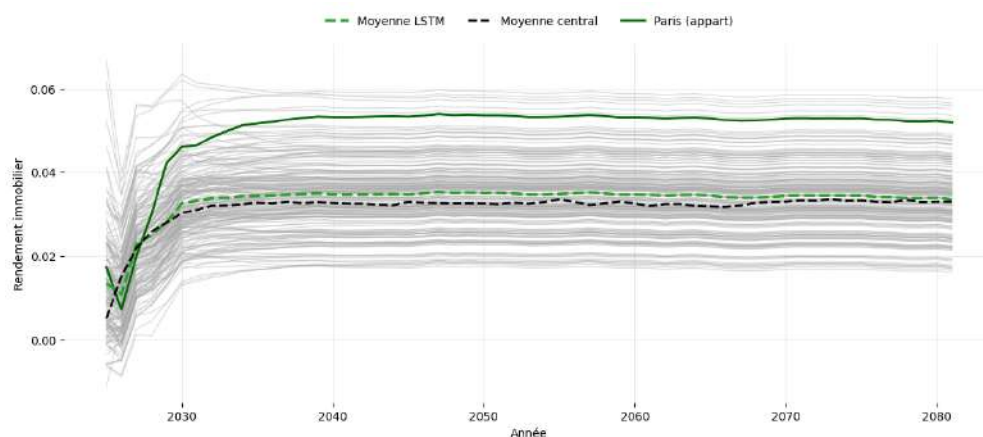
FIGURE 4.10 – Projection du rendement immobilier par récursion, Paris-Appartement



Ce modèle, via l'architecture décrite ci-avant, permet d'obtenir les courbes de rendements présentées en figure 4.11. Il est essentiel de comparer ces projections avec celles de la méthode centrale et la méthode par GSEs segmentés, dans la mesure où les rendements immobiliers conditionnent directement le calcul des LTV, toutes choses égales par ailleurs.

Les projections du modèle LSTM aboutissent à un rendement moyen de long terme d'environ 3,40% (courbe verte en pointillés), contre 3,33% pour la méthode via un unique GSE national (courbe noire en pointillés) et 3,30% pour la méthode via GSEs segmentés. Ces écarts faibles confirment la cohérence globale entre les différentes approches.

FIGURE 4.11 – Rendement immobilier moyen, par segment



Malgré cette cohérence d'ensemble, des divergences de variance persistent entre les deux méthodes segmentées. Leur étude et leur analyse sont présentées dans la section suivante. Ces écarts, qui conditionnent directement le déclenchement des NNEG, constituent un élément central pour la compréhension du mécanisme de tarification. La section suivante présente ainsi les résultats issus de ces approches de projection, et en propose une analyse comparative une fois intégrées dans l'outil de tarification.

4.4 Résultats et analyses

Les trajectoires issues des modèles présentés en sections 4.2 et 4.3 sont intégrées au modèle de projection de flux PVH afin d'en analyser les effets sur la tarification. Les résultats issus de ces modélisations segmentées sont comparés à ceux du modèle central. L'analyse qui suit s'attache à comprendre les écarts observés, à identifier les apports et à mesurer les limites d'une telle segmentation. Pour ce faire, une comparaison systématique est effectuée entre les trois méthodes mises en place. Par la suite, la modélisation des rendements immobiliers à l'aide de GSEs segmentés est appelée Méthode 1 et la Méthode 2 correspond à la modélisation de ces rendements au moyen d'un LSTM.

4.4.1 Analyses graphiques

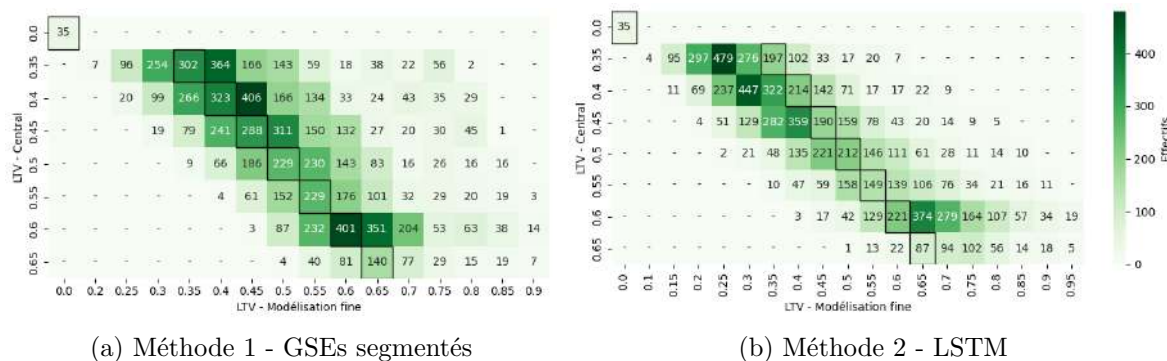
Les méthodes de tarification sont identiques pour les trois approches ; la seule différence réside dans les scénarios de rendements immobiliers utilisés pour valoriser les flux. Plusieurs analyses visuelles permettent de mettre en évidence l'impact du changement de méthode de modélisation de ces scénarios.

Écarts de LTV

En premier lieu, il est intéressant de comprendre dans quelle mesure les LTV ont évolué lors du passage d'une tarification homogène à une approche segmentée en fonction du bien. Les matrices de fréquence ci-dessous mettent en évidence deux dynamiques distinctes :

- Méthode 1 : tendance haussière des LTV, comparée au modèle central,

- Méthode 2 : accentuation de la dynamique présente en central, conduisant à prêter respectivement moins (plus) aux individus pour lesquels la LTV optimale était inférieure (supérieure) à 50% en central. Cette dynamique s’inscrit dans la lignée de résultats déjà publiés sur le sujet ¹⁵.

FIGURE 4.12 – Matrice de fréquence — LTV centrale \times LTV segmentée

(a) Méthode 1 - GSEs segmentés

(b) Méthode 2 - LSTM

Note : Ces figures se lisent ainsi : respectivement 9 (48) individus ayant 50% de LTV en central sont passés à 35% en GSEs segmentés (LSTM) - croisement de 0.35 en abscisse et 0.5 en ordonnée.

Sur les figures 4.12a et 4.12b, les cellules encadrées correspondent aux individus dont la LTV reste inchangée par rapport au central, représentant respectivement 23% et 15% du portefeuille. La plage des LTV optimales est élargie par les deux approches segmentées, passant de 35–65% en central à 10–95%. En effet, la prise en compte de la typologie et de la localisation dans la tarification fait perdre de l’importance à l’âge dans le calcul de LTV. Cependant, les écarts peuvent être jugés trop forts : autour de 1 500 individus (soit plus de 15% du portefeuille) se retrouvent avec une LTV en dehors de l’intervalle de référence [35% ; 65%]. Ce constat pose question, car une LTV supérieure à 65% peut être considérée comme excessivement risquée, tandis qu’un prêt de 10% ou 20% de la valeur du bien est souvent insuffisant pour répondre aux attentes de l’emprunteur. Dans ce sens une analyse critique des résultats est présentée en section 4.5.

Analyse départementale

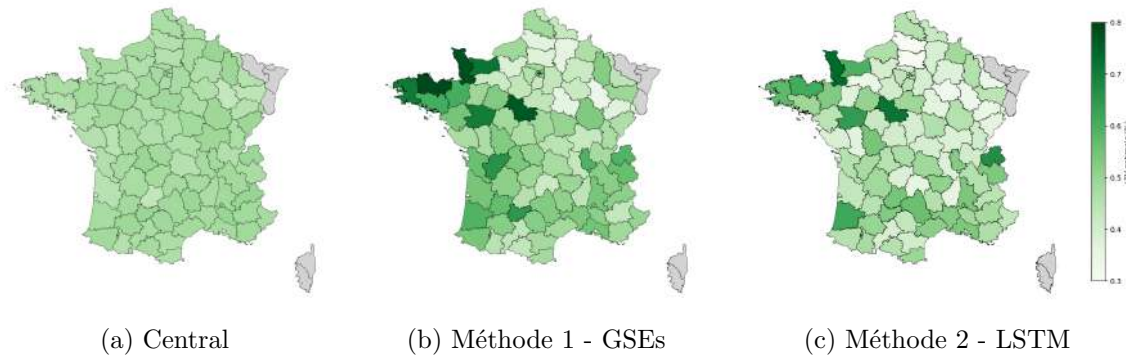
La lecture des cartes 4.13 met en évidence une nette différence de dynamique de tarification. En central, les moyennes départementales de LTV sont homogènes, ne révélant aucune différenciation faite sur cet axe ; ce sont les variables liées à la longévité qui portent une grande importance dans le calcul de LTV optimale. Les deux méthodes segmentées font apparaître une variabilité spatiale des LTV, illustrant l’apport de l’utilisation de scénarios immobiliers segmentés.

La Bretagne élargie se distingue nettement du reste du territoire, tout comme Paris ¹⁶. La base DVF met en évidence une forte hausse des prix en Bretagne entre 2014 et 2023 (cf. annexe 5). Cette dynamique a été prise en compte lors de la construction des indices, puis intégrée par les modèles.

15. [A-W. Shao et al., 2014] et [H. Alai et al., 2014], relatifs au secteur restreint de Sydney, sur 5 segments.

16. Ces deux zones géographiques font l’objet d’une analyse critique, présentée en section 4.5.

FIGURE 4.13 – Cartes des LTV optimales

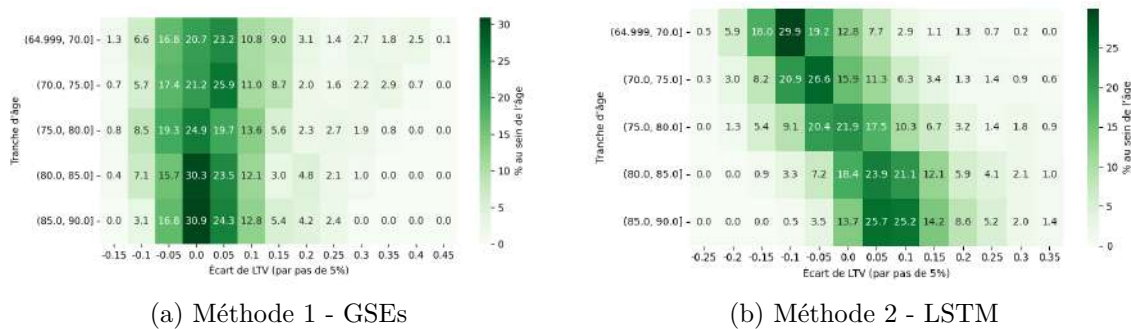


L'étendue des moyennes départementales de LTV optimales devient beaucoup plus large, allant de 35% à 75%, contre une plage autour de 45% pour le central. À l'échelle départementale, les écarts entre les deux méthodes segmentées demeurent limités, ce qui est rassurant quant à la cohérence des dynamiques produites par chacun des modèles. Les cartes régionales des écarts de LTV sont lisibles en Annexe, c.f. 8.

Effet de l'âge

L'analyse par tranches d'âge montre que, pour la Méthode 1, tous les profils d'âge sont impactés de la même manière par le changement de méthode de projection immobilière. En revanche, pour la Méthode 2, les profils 'jeunes' voient leur LTV baisser au passage à la tarification segmentée par LSTM, tandis que les individus plus âgés voient leur LTV augmenter. Cette dynamique est illustrée par la figure 4.14 :

FIGURE 4.14 – Heatmap des écarts de LTV par Age



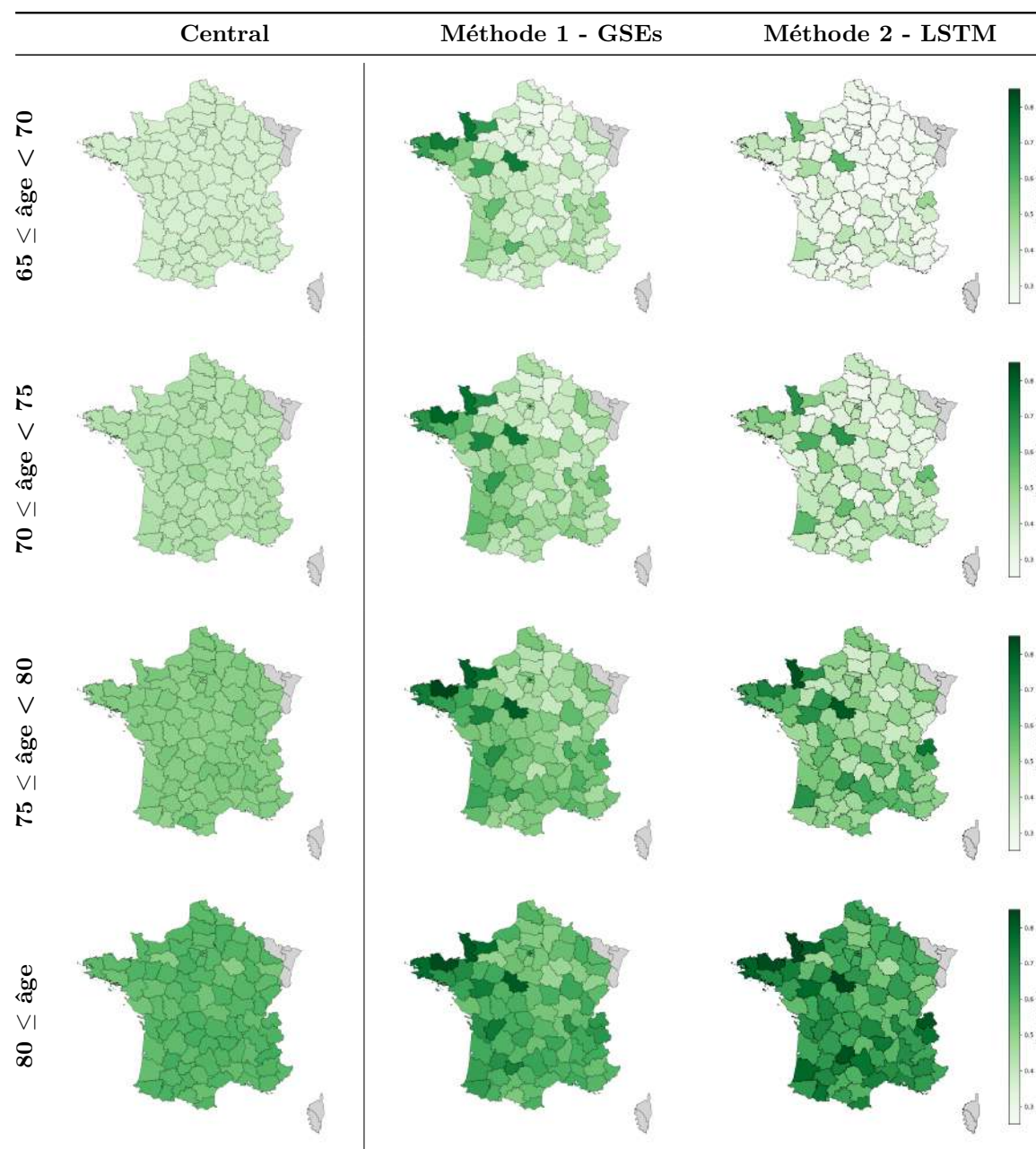
Note : Ces figures se lisent tel que respectivement 30,9% (13,7%) des 85-90 ans ne voient pas leur LTV être modifiée en Méthode 1 (Méthode 2) par rapport au central - croisement de Tranche d'âge [85,0, 90,0] en ordonnée avec Ecart de LTV 0.0 en abscisse. En agrégeant par sens d'écart, ces graphiques montrent que 25% (74%) des 65-70 ans voient leur LTV diminuer pour la Méthode 1 (Méthode 2) et 49% (82,3%) des 85-90 ans voient leur LTV augmenter.

Cette dynamique s'explique par le rôle différencié des contraintes selon les profils : chez les emprunteurs 'jeunes', une hausse des rendement immobiliers tend la contrainte de NNEG tant en valeur qu'en probabilité, alors qu'elle assouplit la contrainte de rendement chez les plus âgés. Cependant, une étendue accrue des rendements n'impacte pas la probabilité de déclenchement de NNEG mais sa valeur (phénomène expliqué en section 4.4.3). Sur les figures 4.15a et 4.15b, les deux groupes de contraintes sont bien distincts : les mêmes familles de contraintes (rendement vs. gestion du risque) se déclenchent pour des profils d'âges similaires dans les deux méthodes.

Effets croisés de l'âge et de la localisation

Les cartes ci-dessous permettent de lire l'effet combiné de l'âge de l'emprunteur et de la localisation géographique du bien sur la LTV optimale. On observe notamment un impact particulièrement marqué pour les profils les plus 'jeunes' situés en Bretagne, confirmant l'influence significative des dynamiques immobilières locales. Ces cartes confirment également la dynamique précédemment décrite : les LTV augmentent chez les profils âgés et diminuent chez les profils jeunes lors du passage à la modélisation fine de l'immobilier.

FIGURE 4.17 – LTV moyenne par département et tranche d'âge – Central *vs.* Segmenté



Si l'âge demeure un facteur explicatif important, son effet semble complété par celui de la localisation, en lien direct avec les rendements immobiliers différenciés. Cette observation motive naturellement la transition vers l'analyse de la variance, afin de quantifier précisément la part de l'âge, du département et du type de bien dans l'explication de ces écarts.

4.4.2 Analyses de l'explication de la variance de la *Loan-To-Value*

Pour mesurer l'importance relative de l'âge, du département et du type de bien dans la variabilité de la LTV optimale, une série de régressions linéaires ainsi qu'une ANOVA ont été réalisées à partir des données individuelles.

Lors de l'analyse par régression, chacun des trois facteurs (âge, département, type de bien) a été introduit successivement dans des modèles explicatifs, seul ou en combinaison, afin d'en isoler la contribution propre. Cette approche permet d'identifier la part de variance expliquée par chaque variable indépendamment des autres. À partir des coefficients de détermination R^2 associés à chaque modèle, on peut calculer la part marginale expliquée par chaque variable de la manière suivante :

$$\text{Part marginale de } X_i = R^2_{\{X_1, X_2, \dots, X_p\}} - R^2_{\{X_j\}_{j \neq i}}$$

Les résultats présentés en table 4.3 confirment la très forte dépendance du modèle central à la seule variable d'âge : à elle seule, elle explique plus de 65% de la variance de la LTV optimale, tandis que la contribution marginale du département et du type de bien sont quasi nulles. Ce comportement est cohérent avec une hypothèse de projection immobilière unique, constante sur tout le territoire.

TABLE 4.3 – Décomposition de la variance expliquée de la LTV optimale

Source de variation	Modèle central	Méthode 1 - GSEs	Méthode 2 - LSTM
R^2 – âge seul	65,3 %	28,3 %	60,3 %
R^2 – département seul	1,6 %	46,4 %	20,9 %
R^2 – type seul	0,0 %	0,6 %	0,2 %
R^2 – âge + département	65,7 %	73,1 %	79,3 %
R^2 – âge + type	65,3 %	28,9 %	60,5 %
R^2 – département + type	1,6 %	47,0 %	21,1 %
R^2 – âge + département + type	65,7 %	73,8 %	79,5 %
Contribution marginale de l'âge	64,1 %	26,8 %	58,4 %
Contribution marginale du département	0,4 %	44,9 %	19,0 %
Contribution marginale du type de bien	0,0 %	0,7 %	0,0 %

À l'inverse de la configuration centrale, l'introduction d'une modélisation segmentée des rendements immobiliers modifie profondément la hiérarchie des facteurs explicatifs.

- Pour la Méthode 1, le département devient la variable dominante ; le type de bien reste très secondaire, mais conserve un effet non nul¹⁷. Cette configuration reflète la capacité de cette approche à capter les disparités territoriales dans les rendements immobiliers et à ajuster finement les LTV optimales selon les zones géographiques,
- Pour la Méthode 2, l'âge conserve son pouvoir explicatif prédominant, tandis que le département prend un poids important, bien supérieur à celui en central. Cette structure traduit un compromis : les spécificités géographiques restent capturées, mais l'information portée par l'âge conserve un rôle prépondérant, en lien avec la dynamique décrite précédemment entre contraintes de rendement et contraintes de NNEG. Aussi, l'*embedding* mis en œuvre dans le cadre de la Méthode 2 induit, par définition, une perte d'information, dans la mesure où la dimension a été réduite de 182 à 9 composantes.

Afin de compléter cette analyse sous un angle purement statistique, une analyse de la variance (ANOVA) a été réalisée. Contrairement à la seule lecture des R^2 , l'ANOVA fournit une

17. Voir Annexe .9 pour une représentation cartographique de la segmentation par type de bien.

décomposition formelle de la variance ainsi qu'un test d'hypothèse associé à chaque facteur (statistique de Fisher et p -value). Cette approche est donc complémentaire pour juger de la robustesse et de la significativité statistique des effets identifiés précédemment.

TABLE 4.4 – Analyse de la variance — Effets explicatifs sur la LTV optimale

(a) Modèle central				(b) Méthode 1 - GSEs			
	Som. carrés	stat. Fisher	p -value		Som. carrés	stat. Fisher	p -value
dép.	0.308	0.98	0.523	dép.	67.962	153.82	$< 2.2 \times 10^{-16}$
type_bien	0.0002	0.045	0.832	type_bien	1.012	206.22	$< 2.2 \times 10^{-16}$
âge	52.529	15121.71	$< 2.2 \times 10^{-16}$	âge	40.545	8258.90	$< 2.2 \times 10^{-16}$
Résidu	28.116	–	–	Résidu	39.735	–	–

(c) Méthode 2 - LSTM			
	Som. carrés	stat. Fisher	p -value
dép.	50.4415	83.3082	$< 2.2 \times 10^{-16}$
type_bien	0.5160	76.6962	$< 2.2 \times 10^{-16}$
âge	154.951	23032.263	$< 2.2 \times 10^{-16}$
Résidu	54.4529	–	–

Dans le modèle central, seul l'effet de l'âge ressort significatif ; les effets du département et du type de bien ne sont pas significatifs au seuil de 5%. L'introduction de rendements immobiliers segmentés par département et type de bien modifie profondément la hiérarchie des facteurs explicatifs : les trois effets (âge, département, type) deviennent hautement significatifs pour les méthodes segmentées. Ce résultat est en cohérence avec les analyses faites à partir des calculs de contributions marginales ci-dessus.

Après avoir quantifiés les effets des différentes variables prises en compte dans la tarification, les écarts de tarification sont objectivés par une étude quantitative fondée sur des métriques agrégées à l'échelle du portefeuille et sur une décomposition de la variance des LTV. Pour cela, dans un premier temps, les distributions de rendements immobiliers sont étudiées.

4.4.3 Comparaison des scénarios économiques projetés

Les différences de méthodologies résident dans la modélisation des rendements immobiliers. Le tableau 4.5 compare les distributions des scénarios immobiliers à partir d'indicateurs homogènes. Pour un segment et une méthode donnés, on dispose de N trajectoires sur T années, notées $\{R_{n,t}\}_{t=1..T}$ pour $n = 1..N$. Pour chaque trajectoire n , on calcule (i) la moyenne temporelle $\mu_n = \frac{1}{T} \sum_{t=1}^T R_{n,t}$ et (ii) la variance temporelle $v_n = \text{Var}_t(R_{n,t})$:

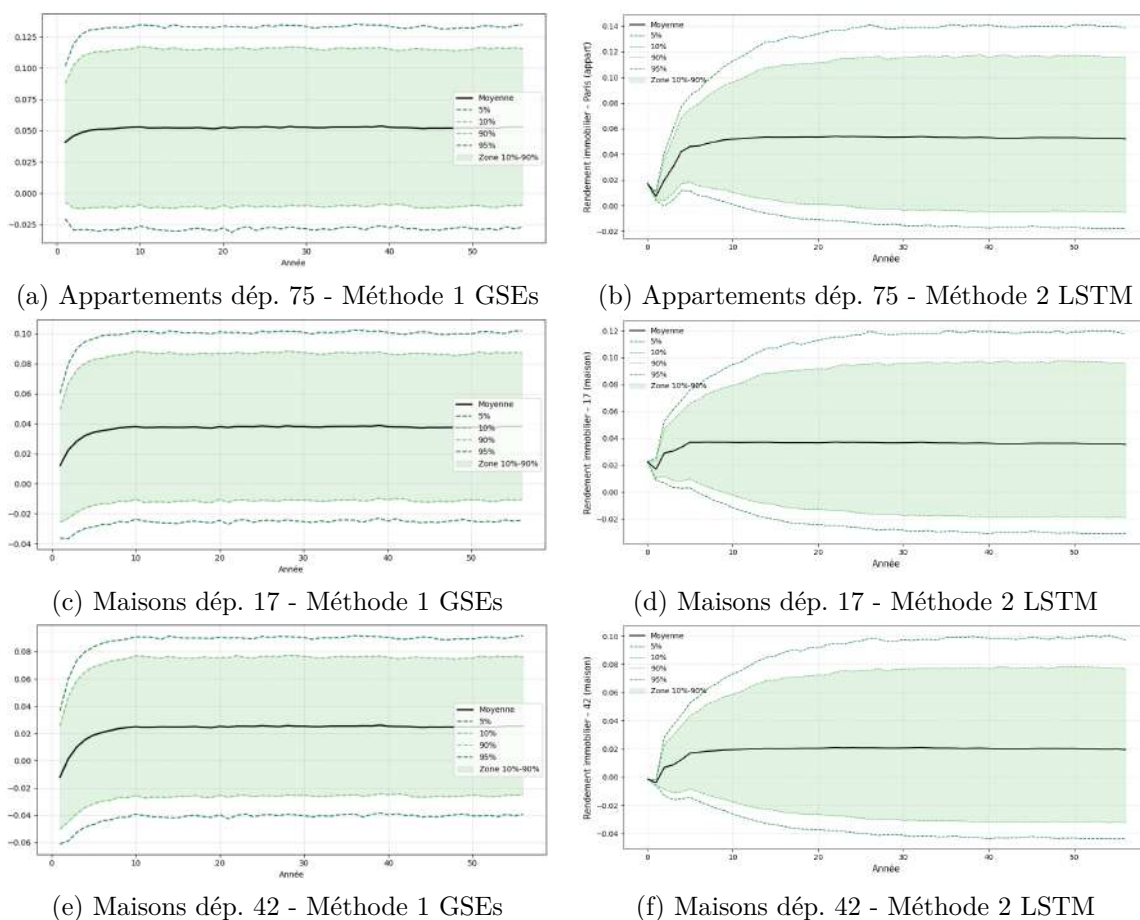
TABLE 4.5 – Comparaison des indicateurs statistiques des scénarios

	GSEs	LSTM
Variance - $\bar{v} = \frac{1}{N} \sum_n v_n$	0,00158	0,00167
Variance - $\min(v_n)$	0,00115	0,00149
Variance - $\max(v_n)$	0,00253	0,00194
Moyenne - $\bar{\mu} = \frac{1}{N} \sum_n \mu_n$	3,08 %	3,32 %
Moyenne - $\min(\mu_n)$	1,39 %	1,61 %
Moyenne - $\max(\mu_n)$	5,20 %	5,74 %

Ces indicateurs permettent de comparer simultanément le niveau moyen des rendements

simulés et l'hétérogénéité des trajectoires, ainsi que leur dispersion inter-scénarios, entre les deux méthodes. L'analyse des quantiles met aussi en évidence des divergences marquées :

FIGURE 4.18 – Exemples de quantiles et moyennes de rendements immobiliers



Les profils temporels diffèrent sensiblement : les trajectoires issues du LSTM présentent un début de projection plus modéré que celles construites via les GSEs. Ce phénomène s'explique par la structure même des deux méthodes. Dans les GSEs segmentés, les aléas simulés sont injectés dès la première année. À l'inverse, l'approche LSTM génère ses trajectoires de manière autorégressive : les conditions initiales étant similaires pour chaque scénario d'un même segment (*modulo* les données exogènes obtenues via le GSE central et utilisées en entrée), les prédictions restent proches en début de projection et la variabilité entre scénarios n'apparaît que progressivement. Au-delà de 20 ans (régime établi), l'écart inter-quantiles des scénarios produits par LSTM s'élargit, révélant une dispersion plus marquée qu'en Méthode 1.

Concrètement, cette combinaison rendements de long terme supérieurs, variabilité accrue et départ plus prudent aboutit à une hausse de la valeur NNEG espérée en LSTM *vs.* scénarios fondés sur les GSEs. Le coût de la garantie de non-dépassement capte l'asymétrie du PVH et se calcule comme l'espérance du manque-à-gagner actualisé à la date de liquidation T :

$$\text{Valeur}_{\text{NNEG}} = \mathbb{E} \left[\frac{\max(CRD_T - V_T, 0)}{DF_T} \right]$$

Par construction, la NNEG est positive ou nulle; les scénarios favorables n'en réduisent pas la valeur, tandis que les scénarios défavorables l'augmentent. La queue basse des rendements étant plus dégradée sous LSTM (c.f. figure 4.18), la valeur de NNEG augmente mécaniquement. Ces éléments expliquent pourquoi l'âge joue un rôle plus déterminant dans la variance de la LTV sous LSTM analysée en section 4.4.2 : le calcul de la LTV optimale devient davantage sensible à cette variable lorsque les rendements présentent une dispersion accrue.

Ainsi, la valeur de la NNEG est plus forte en Méthode 2, et d'autant plus pour les profils 'jeunes', pour lesquels la valeur de la dette est gonflée par les intérêts composés à 6% sur une plus longue période. Cette dynamique se lit clairement sur la figure 4.15 : 100% des individus de moins de 72 ans voient leur LTV plafonnée à cause de la contrainte de valeur NNEG. Cela conduit à une plus forte relation âge-LTV en LSTM qu'en GSEs.

4.4.4 Analyses d'indicateurs quantitatifs

Comme déjà mentionné, la Méthode 2 génère des rendements immobiliers en moyenne plus élevés, ce qui se traduit directement sur les valeurs immobilières en fin de prêt. La capitalisation des rendements, combinée à une répartition du portefeuille non uniforme entre départements amplifie les faibles écarts de moyennes long terme à la sortie des modèles. Pour rappel, le patrimoine moyen initial est de 187k€ et l'espérance de vie d'environ 21 ans.

TABLE 4.6 – Statistiques agrégées – Portefeuille immobilier

	Central	GSEs	LSTM
Patrimoine moyen à $t_{\text{décès}}$ (€)	412 872	428 790	473 698
Évolution patrimoine total	+121 %	+129 %	+153 %
Rendement <i>flat</i>	5,64 %	6,04 %	7,16 %
Rendement capitalisé	3,77 %	3,95 %	4,44 %

Malgré un rendement immobilier plus fort via LSTM, la table 4.7 montre que les LTV globales issues de cette méthode apparaissent plus conservatrices que celles obtenues via le central et les GSEs. Ce constat s'explique par la hausse de la valeur de NNEG en Méthode 2. Alors que la Méthode 1 déclenche majoritairement la contrainte liée à la probabilité de NNEG, le LSTM est très largement contraint par la valeur de cette même garantie. Si la fréquence de déclenchement est globalement comparable, la sévérité des pertes sous LSTM est plus élevée, ce qui élève le coût de la garantie et bloque plus tôt la maximisation de la LTV. Cet effet neutralise le léger avantage de rendement moyen du LSTM, et aboutit à une LTV moyenne comparable au central et nettement inférieure à celle obtenue avec les scénarios GSEs.

TABLE 4.7 – Statistiques agrégées – Portefeuille de PVH

	Central	GSEs	LSTM
Rendement <i>flat</i>	9,8 %	9,9 %	9,8 %
Rendement capitalisé	5,3 %	5,3 %	5,4 %
Prêt moyen (€)	88 329	97 223	87 171
VAN moyenne (€)	104 175	117 967	91 365
Probabilité NNEG	35,0 %	38,0 %	28,4 %
LTV moyenne	47,2 %	52,0 %	46,6 %
Prop. du ptf. contrainte par le rdt. <i>flat</i>	13,2 %	15,2 %	13,9 %
Prop. du ptf. contrainte par le rdt. <i>flat</i> p5	6,6 %	1,3 %	–
Prop. du ptf. contrainte par le proba. NNEG	46,3 %	80,2 %	5,9 %
Prop. du ptf. contrainte par la valeur NNEG	33,5 %	2,9 %	79,8 %

En synthèse, ces analyses quantitatives et graphiques permettent de mettre en évidence les dynamiques spatiales et les effets différenciés induits par la segmentation immobilière. Dans l'ensemble, les LTV obtenues apparaissent plus précises et mieux adaptées au profil de risque, grâce à une meilleure mesure du risque immobilier. Toutefois, certaines projections aboutissent à des niveaux de LTV particulièrement élevés, notamment à Paris et dans la Bretagne élargie. L'ampleur des rendements anticipés sur le marché parisien, en particulier, conduit mécaniquement à des LTV très importantes (cf. 4.12b), ce qui appelle à une lecture critique. Afin de préserver la cohérence économique des résultats, plusieurs méthodes de retraitement sont proposées dans la section suivante.

4.5 Regard critique

Les niveaux de LTV calculés pour certains individus peuvent s'avérer très élevés, atteignant 90% pour certains profils. Bien que ces résultats soient cohérents d'un point de vue modèle, il convient de rappeler que l'outil développé constitue une aide à la décision, et non un mécanisme prescriptif devant être suivi sans discernement.

De plus, un biais important a été identifié et doit être gardé en tête tout au long des analyses. La construction des indices immobiliers, utilisés ensuite pour paramétrer les modèles de Vasicek et entraîner le modèle LSTM développé, présente des limitations méthodologiques pouvant induire une surestimation des rendements. Ce biais pourrait expliquer en partie des rendements immobiliers projetés excessivement élevés en sortie de modèle, et, par ricochet, les niveaux de LTV optimales parfois trop élevés proposés par l'outil de tarification.

En connaissance de ces éléments, une réflexion méthodologique ayant conduit à réaliser des retraitements sur ces valeurs extrêmes est présentée ci-après. Cette démarche se concentre sur la Méthode 1, bien qu'une analyse similaire sur les LTV issues de la Méthode 2 serait également pertinente. L'application de cette sensibilité à la Méthode 2 conduirait vraisemblablement à des conclusions similaires.

4.5.1 Rationnel sur les *Loan-To-Value* jugées trop élevées

Deux cas géographiques spécifiques méritent une attention particulière, et ce à propos des deux méthodes :

- Paris affiche des LTV extrêmement élevées, directement imputables à la performance historique exceptionnelle de son marché immobilier, notamment sur la décennie 2010–2020,

- La Bretagne présente également des LTV optimales élevées, reflet probable d'un double effet : une dynamique de prix déjà haussière sur une longue période, accentuée récemment par un regain d'attractivité post-covid.

Les LTV extrêmes dans ces zones posent la question de leur plausibilité opérationnelle. Même si elles sont justifiables dans le cadre des modèles (rendements élevés en entrée), elles peuvent être difficilement soutenables dans une politique de crédit prudente.

Il est légitime de questionner la persistance des rendements immobiliers élevés observés sur certains territoires. Dans le cas de Paris, plusieurs indicateurs montrent un possible ralentissement à moyen-long terme : la population intra-muros est en baisse depuis 2015, chutant de près de 73 000 habitants entre 2015 et 2021, selon l'INSEE. De plus, l'effet post-covid et l'essor du télétravail ont impacté le marché immobilier et certains modes de vie de la population française : en 2021, 21,7% des salariés télétravaillaient au moins un jour par semaine en France, ce qui a favorisé l'installation dans des zones moins denses. L'INSEE note une migration accrue depuis les grandes métropoles, avec un ratio entrées/départs augmentant de 1,87 en 2019 à 2,15 en 2022 pour l'aire de Paris.

Ces mouvements migratoires, stimulés par la recherche d'espace et la flexibilité offerte par le télétravail, peuvent avoir un rôle dans l'essor des prix dans certaines zones littorales bretonnes. Cependant, ils relèvent d'un phénomène conjoncturel post-crise, et risquent de s'atténuer avec le retour à une vie professionnelle plus classique. Le fort rendement immobilier dans la région bretonne peut également être expliqué par l'attrait grandissant des acquéreurs immobiliers pour la Bretagne, en raison de ses températures estivales modérées, un avantage devenu plus recherché dans le contexte de hausse des températures.

4.5.2 Rationnel de modification des perspectives immobilières

Outre les dynamiques de migration et de marché, plusieurs risques structurels ou conjoncturels pourraient modifier les perspectives immobilières :

- Submersion des côtes : le risque croissant de la montée des eaux et d'érosion côtière peut rendre certains biens situés sur les littoraux moins attractifs et moins valorisés à long terme, voir même les détruire.
- Évolution démographique : le vieillissement de la population, la baisse des naissances, ou la dynamique des migrations internes peuvent transformer la demande structurelle dans certaines zones,
- Transition écologique : par le futur, les normes énergétiques pourront se voir renforcées (obligation de rénovation ou tarification carbone) : certains biens verront leur valeur potentiellement diminuer.
- Réchauffement climatique : certaines régions du Sud, historiquement prisées, pourraient perdre en attractivité à cause de la hausse des températures estivales, des vagues de chaleur répétées ou des épisodes de sécheresse.

Ces éléments, bien que structurellement importants, ne sont pas pris en compte dans la modélisation actuelle, en raison de leur caractère incertain ou difficilement quantifiable. Ils rappellent toutefois que les rendements immobiliers historiques ne peuvent être considérés comme des repères immuables. Cela renforce l'intérêt de tester l'impact de retraitements prudentiels, afin de garantir la robustesse et la plausibilité des résultats produits par le modèle.

Plusieurs approches peuvent être appliquées, à différents niveaux de la chaîne de modélisation :

1. en amont du GSE : retraiter les rendements historiques utilisés pour calibrer les indices immobiliers, par exemple en réduisant l'impact des hausses exceptionnelles survenues à Paris ou en Bretagne. Cette option permettrait de construire un scénario plus conservateur dès la génération des trajectoires futures.
2. en sortie du GSE : plafonner ou lisser les rendements projetés par le modèle avant de réaliser la tarification. Cela permet de limiter l'amplitude des scénarios extrêmes en conservant une modélisation segmentée.
3. en sortie du modèle de tarification : appliquer un plafonnement sur les LTV finales issues du modèle, ce qui est simple opérationnellement et transparent.

Pour intégrer cette logique prudentielle tout en conservant la richesse des projections immobilières, deux méthodes de retraitement ont été mises en place séparément :

- Un retraitement en sortie du modèle de tarification : les LTV issues du modèle de pricing sont plafonnées à 70%. Ce scénario s'apparente à un stress test opérationnel visant à rapprocher les résultats du modèle d'une réalité de marché.
- Un second retraitement a été appliqué à certains segments territoriaux, principalement situés à Paris, ou en Bretagne et dans des départements limitrophes¹⁸, pour lesquels les indices issus du modèle hybride présentaient une dynamique très élevée. Pour le segment parisien, un indice construit par l'INSEE et des données publiques de prix/m² parisiens (Paris Notaires Services) sont utilisés pour remplacer l'indice branché en amont du modèle GSE du segment Paris × Appartement (c.f. 4.1.3). La moyenne des log-rendements INSEE et de l'évolution des prix/m² remplace donc les log-rendements sortis par le modèle hybride. Pour les autres segments retraités, les trajectoires entre 2021 et 2024 (post-covid) ont été lissées ; chaque valeur entre 2021 et 2024 a été recalculée comme une moyenne pondérée entre la moyenne sur 2014–2020 et la valeur initialement sortie par le modèle. Plus précisément, une pondération $\alpha = 0,8$ a été appliquée à la moyenne historique (2014–2020), de façon à atténuer les hausses excessives de l'indice observées post-covid sur ces segments. Le niveau de α a été déterminé empiriquement, de manière à conserver les tendances de ces dernières années tout en se rapprochant du niveau moyen historique. Ce lissage vise ainsi à stabiliser artificiellement les rendements jugés anormaux tout en conservant une part de la dynamique issue du modèle.

4.5.3 Analyses après retraitements

LTV plafonnée à 70%, post-tarification

L'impact de ce plafonnement reste limité à l'échelle du portefeuille. En moyenne, la LTV globale du portefeuille diminue de 0,7%, et le montant moyen prêté baisse d'environ 1 000€. Certaines catégories sont plus affectées : les segments parisiens voient leur LTV moyenne diminuer de 8%, contre 1% pour les maisons en province. À la maille régionale, la LTV moyenne en Bretagne diminue de 3%. Le lecteur intéressé pourra trouver en Annexe un ensemble de cartes départementales, par âge, post-retraitement (c.f. Annexe 10).

Chez les individus pour lesquels la LTV a été baissée, le risque de dépassement est nettement réduit : la probabilité d'activation de la NNEG chute de 12%. Ce gain en sécurité s'accompagne d'une baisse du montant prêté. Par ailleurs, une LTV de 70% reste un niveau très attractif en comparaison des standards du marché traditionnel, confirmant la compétitivité du prêt même après ajustement.

18. Les départements concernés sont : 14 Calvados ; 16 Charente ; 22 Côtes-d'Armor ; 29 Finistère ; 41 Loir-et-Cher ; 49 Maine-et-Loire ; 50 Manche et 75 Paris

TABLE 4.8 – Évolution des indicateurs pour les individus dont la LTV a diminué – Méthode de plafond

Indicateur	Avant	Après	Évolution
LTV moyenne	79 %	70 %	-9 %
Probabilité NNEG	39 %	27 %	-12 %
Rendement capitalisé	5,2 %	5,3 %	+0,1 %
Nombre d'indiv. concernés		565	
Nombre de dép. concernés		18	

Lissage de certains indices, pré-GSE

L'ajustement appliqué en amont du paramétrage des GSEs a eu un impact mesurable sur les scénarios de rendement immobilier. Cette baisse de rendement se répercute naturellement sur les indicateurs de performance et de gestion de risque calculés.

En effet, la LTV moyenne du portefeuille chute de 52,0% à 50,0% traduisant une gestion du risque plus prudente induite par les projections plus réalistes et prudentes des prix futurs. L'effet est particulièrement marqué sur le segment "Maisons en Province". D'un point de vue régional, la Bretagne enregistre une baisse moyenne de LTV de l'ordre de 10%, conséquence directe du retraitement ciblé de plusieurs départements bretons. De plus, le rendement immobilier parisien est fortement réduit, passant de 277% en moyenne sur la durée de vie du portefeuille à 161%. La LTV moyenne sur Paris passe alors de 78,2% à 53,3%, ce qui semble plus cohérent avec le marché, et est plus prudent.

TABLE 4.9 – Évolution des indicateurs pour les individus dont la LTV a diminué – Méthode de lissage des indices

Indicateur	Avant	Après	Évolution
LTV moyenne	74 %	50 %	-24 %
Probabilité NNEG	39 %	36 %	-3 %
Rendement capitalisé	5,3 %	5,3 %	- %
Nombre d'indiv. concernés		780	
Nombre de dép. concernés		7	

En termes d'indicateurs de rendement ou de risque à l'échelle agrégée, l'impact reste marginal. Toutefois, ce retraitement permet de renforcer la robustesse et la crédibilité opérationnelle des montants proposés, en les ramenant dans des ordres de grandeur compatibles avec la pratique du marché.

L'analyse critique menée précédemment a mis en évidence certaines limites, notamment la surévaluation des rendements immobiliers et des LTV associées dans des zones spécifiques. Les pistes de retraitement envisagées permettent de corriger partiellement ces biais et d'améliorer la cohérence économique des projections. Dans cette continuité, la section suivante propose une approche de tarification prudente. Celle-ci repose sur la consolidation des différentes méthodes de projection, en retenant pour chaque individu la LTV minimale issue des modèles développés, et corrigés.

4.6 Tarification prudente

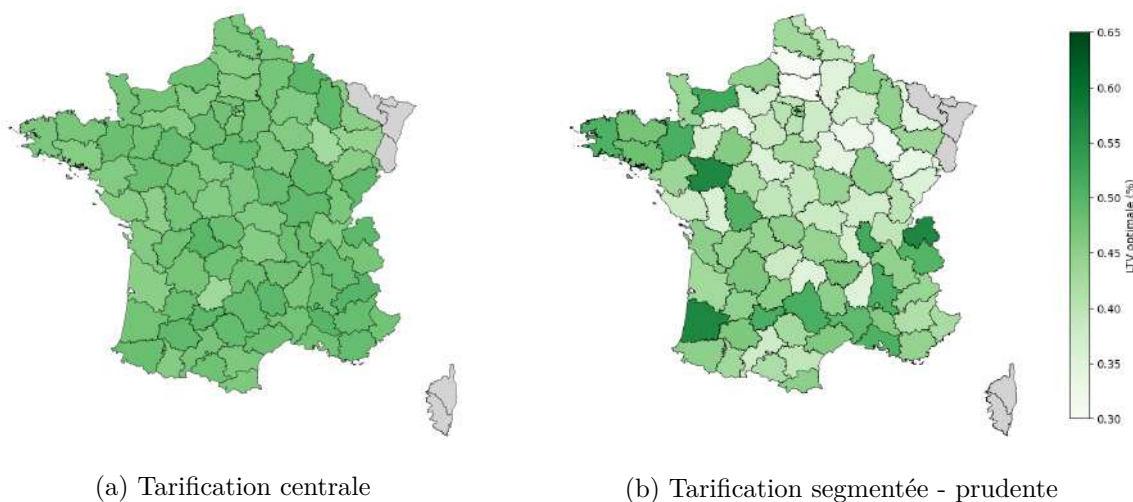
Les approches segmentées permettent de mieux gérer le risque de dépréciation immobilière en adaptant le montant prêté à la localisation et au type du bien sous-jacent. Si les différentes méthodes testées dégagent des tendances globalement cohérentes, elles présentent néanmoins certains écarts de LTV. Dans cette perspective, et afin de renforcer la robustesse de la tarification proposée, les résultats issus des différentes modélisations sont combinés pour définir une valeur unique de référence.

Cette valeur retenue se veut volontairement prudente : les analyses développées en section 4.5 mènent à garder une marge de risque. Ainsi, la LTV optimale utilisée dans la tarification finale correspond à la valeur minimale obtenue parmi les modèles segmentés disponibles, traduisant une approche conservatrice. En cas de conflit entre plusieurs modèles, un ordre de priorité est appliqué afin de trancher de manière cohérente.

- LSTM,
- GSEs segmentés,
- GSEs segmentés, retraitement en amont des GSEs (c.f. 4.5.3, colonne 1),
- GSEs segmentés, LTV plafonnée à 70% (c.f. 4.5.3, colonne 2).

Ce choix méthodologique revient à proposer systématiquement le montant de prêt le plus faible parmi les différentes estimations possibles, de manière à réduire son risque. La comparaison cartographique met en évidence que, même sous ce prisme conservateur, la tarification demeure cohérente avec les réalités territoriales mises en lumière par la modélisation.

FIGURE 4.19 – Cartes des écarts de LTV par région



Par construction, la LTV est en baisse, signe du gain de prudence introduit, tandis que les indicateurs de rendement et de risque restent globalement stables. Autrement dit, le prêteur gagne en prudence sans dégradation notable de la performance, mais au prix d'une moindre compétitivité. Une LTV moyenne autour de 43% semble toutefois en ligne avec les conditions de marché actuelles. Les résultats détaillés par tranches d'âge ou zones géographiques, présentés en annexe (c.f. Annexes 12 et 13), mettent en évidence une différenciation accrue selon l'âge par rapport au modèle central, ainsi que l'apparition de disparités liées aux zones géographiques.

TABLE 4.10 – Indicateurs agrégées – Comparaison Central *vs.* Prudent

	Central	Prudent
Rendement <i>flat</i>	9,8 %	9,9 %
Rendement capitalisé	5,3 %	5,4 %
Prêt moyen (€)	88 329	81 831
VAN moyenne (€)	104 175	88 549
Rendement <i>flat</i> au 5ème percentile	4,4 %	4,9 %
Probabilité NNEG	35,0 %	28,5 %
LTV moyenne	47,2 %	43,7 %

En passant d'une tarification fondée sur une modélisation homogène du risque immobilier à l'échelle nationale à une approche prudente segmentée, environ un cinquième des emprunteurs conservent un montant de prêt inchangé, un autre cinquième bénéficie d'une hausse (pouvant atteindre +20%), tandis que la majorité voit leur montant diminuer, parfois jusqu'à -25% (c.f. Annexe 11).

Ces analyses ont mis en évidence l'impact du passage d'une modélisation immobilière agrégée à une approche segmentée. Bien que cette étude présente plusieurs limites, les enseignements généraux restent valables. Une modélisation segmentée de l'immobilier permet en effet de mieux anticiper et gérer ce risque. De plus, prêter en cohérence avec le profil de risque contribue à une gestion plus optimale de la vie du contrat. Ces constats ouvrent la voie au chapitre de conclusion, qui synthétise les apports et les perspectives de ce travail.

Modélisation segmentée des rendements immobiliers

La tarification des PVH repose sur l'évolution future des prix immobiliers. L'approche classique de modélisation de cet actif ne capte pas l'hétérogénéité territoriale de ce marché. Ce chapitre présente les deux méthodes de modélisations segmentées par département \times type de bien (182 segments) utilisées.

1. Construction des indices segmentés

Aucun ensemble d'indices segmentés concernant le marché français, profond, et à maille fine, n'est publique. Un jeu d'indice a été construit, à partir de la base DVF (2014–2024), selon un modèle hybride. Un historique étendu (1980–2024) est ensuite reconstitué via une régression Elastic Net sur variables macroéconomiques nationales (taux, inflation, population) et segmentées (population). Un double retraitement est appliqué pour assurer un alignement national avec l'indice INSEE ainsi qu'une réduction des rendements extrêmes. La reconstruction de ces indices constitue l'une des principales limites de cette étude et impose de conserver un regard critique sur l'ensemble des résultats.

2. Méthode de modélisation 1 : GSEs segmentés

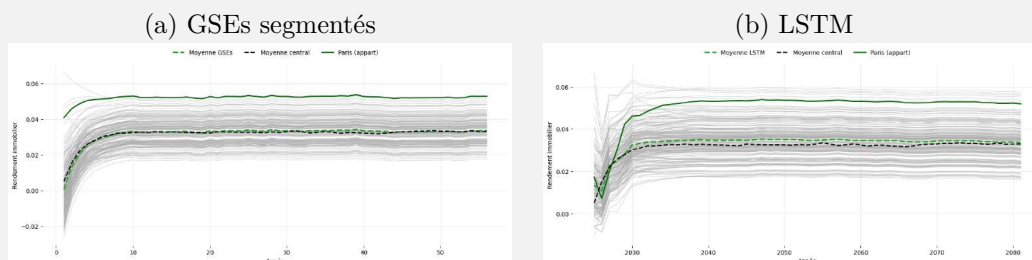
Les rendements découlant de chaque indice segmenté servent à paramétrer un modèle de Vasicek destiné à modéliser un rendement immobilier. Ainsi, chaque indice est relié à un GSE local simulant 4 variables (taux court, taux long, inflation, rendement immobilier). Le portefeuille de PVH est ensuite valorisé à partir de ces scénarios, en maintenant une structure identique à celle du modèle central.

3. Méthode de modélisation 2 : Réseau de neurones (LSTM)

Un modèle LSTM, intégrant un *embedding* des segments, est utilisé pour projeter les rendements immobiliers à long terme à partir de variables macroéconomiques simulées précédemment, en central. L'encodage des segments via *embedding* permet au réseau d'apprendre des dynamiques spécifiques tout en mutualisant l'information entre segments similaires. Cette approche permet ainsi une modélisation fine et dynamique des trajectoires immobilières locales, comparable à la modélisation via GSEs.

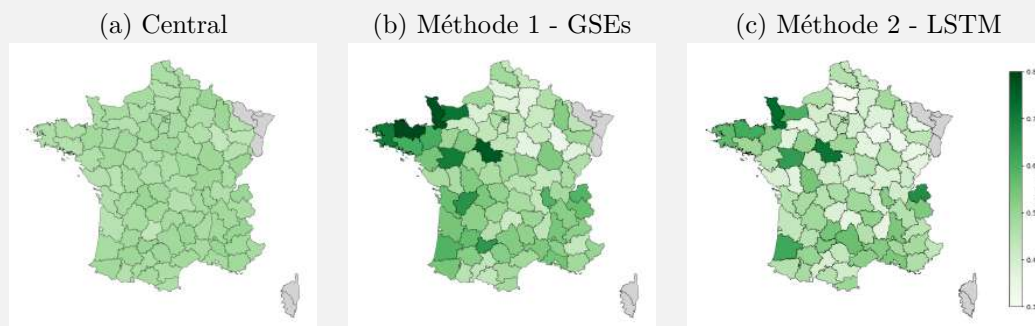
4. Scénarios obtenus

Les scénarios en sortie des GSEs et du LSTM sont comparables entre eux ainsi qu'avec le rendement en environnement central mais présentent des écarts de dispersion. Les rendements en début de projection du LSTM sont moins volatiles et en moyenne plus faibles que ceux des GSEs, ce qui s'explique par le caractère stochastique des GSEs et l'architecture même du LSTM.



5. Résultats et analyses

La lecture graphique des cartes ci-dessous met en évidence l'apport de l'utilisation de rendements segmentés. Ces cartes, couplées à une analyse départementale par tranches d'âge montrent clairement l'impact de la situation géographique sur le montant prêté.



Par ailleurs, une ANOVA ainsi qu'une analyse des R^2 issus de différentes régressions confirment le rôle explicatif du département dans la variation de la LTV. De plus, l'étude d'indicateurs quantitatifs met en évidence des différences dans la gestion du risque selon la méthode utilisée, notamment à travers la répartition des contraintes limitantes.

	Central	GSEs	LSTM
Rendement <i>flat</i>	9,8 %	9,9 %	9,8 %
Rendement capitalisé	5,3 %	5,3 %	5,4 %
Prêt moyen (€)	88 329	97 223	87 171
Rendement <i>flat</i> au 5 ^{ème} percentile	4,4 %	4,8 %	4,9 %
Probabilité NNEG	35,0 %	38,0 %	28,4 %
LTV moyenne	47,2 %	52,0 %	46,6 %
Proportion du portefeuille contrainte par le rdt. <i>flat</i>	13,2 %	15,2 %	13,9 %
Prop. du ptf. contrainte par le rdt. <i>flat</i> p5	6,6 %	1,3 %	–
Prop. du ptf. contrainte par la proba. NNEG	46,2 %	80,1 %	5,9 %
Prop. du ptf. contrainte par la valeur NNEG	33,5 %	2,9 %	79,8 %
Prop. du ptf. refusée	0,4 %	0,4 %	0,4 %

L'hétérogénéité des rendements utilisés dans l'outil de tarification implique différents phénomènes : de manière générale la contrainte de rendement se détend pour les emprunteurs âgés tandis que la contrainte de NNEG devient plus forte pour les plus 'jeunes'. La méthode basée sur les GSEs conduit à une hausse généralisée des LTV, tandis que celle basée sur les LSTM tend à réduire les LTV déjà faibles en environnement central et à accroître celles déjà élevées. Ce phénomène, déjà documenté dans la littérature, suggère que les profils de risque en deçà d'une LTV de 50% en central seraient sous-estimés, et inversement. Cela est notamment dû à l'asymétrie du *payoff* du PVH, et dans l'outil de tarification développé, à la contrainte de valeur NNEG.

Enfin, une tarification prudente est retenue, à l'issue de l'analyse critique des LTV issues de la modélisation. Elle consiste à conserver le minimum des LTV fournies par les modèles, brutes ou retraitées après justification des cas jugés trop élevés. Cette approche conduit à une LTV optimale moyenne de 43%, cohérente avec les niveaux observés sur le marché, tout en prenant compte de l'hétérogénéité selon la situation géographique. Ce résultat met en évidence l'apport concret des méthodes développées, en conciliant prudence et meilleure adéquation au profil de risque réel.

Chapitre 5

Conclusion et limites

5.1 Conclusion

Les travaux réalisés lors de cette étude ont porté sur la modélisation fine du risque immobilier des PVH. Plusieurs méthodologies ont été mises en œuvre pour projeter les rendements immobiliers et analyser leur impact sur les indicateurs clés du produit. L'étude mobilise à la fois une approche économétrique classique et une méthode issue de l'apprentissage automatique. L'analyse comparative des résultats met en évidence une certaine cohérence entre les différentes méthodes, ainsi qu'un gain notable apporté par la prise en compte de la segmentation géographique des rendements immobiliers. Ces différences se traduisent en termes de disparités de niveaux de prêt accordés selon les départements.

La démarche s'est structurée en plusieurs étapes : la construction d'un portefeuille de référence, la conception d'un outil de tarification intégrant des contraintes prudentielles et de rentabilité, puis la génération de scénarios de rendements immobiliers selon différentes méthodologies. Ces scénarios ont ensuite été utilisés en entrée dans l'outil de tarification afin de mesurer leur impact sur les indicateurs financiers et de comparer les approches testées.

En modélisant un rendement immobilier commun à l'ensemble du portefeuille, la *Loan-To-Value* moyenne du portefeuille construit est légèrement supérieure à 47%, ce qui constitue la référence globale. Ce niveau apparaît relativement supérieur aux pratiques observées sur le marché, mais est couplé à un rendement attractif. Lorsque la modélisation est affinée par une segmentation départementale, des disparités apparaissent et sont directement lisibles par carte. Avec la méthode GSEs, la LTV moyenne s'élève à environ 52%, traduisant une propension accrue à prêter dans certains territoires ; à l'inverse, l'approche par réseau de neurones ramène la LTV légèrement en dessous du niveau central, reflétant une prudence relative. Ces écarts, bien qu'apparemment modestes, demeurent significatifs à l'échelle d'un portefeuille et illustrent directement l'influence du choix méthodologique sur les volumes de prêts octroyés.

Au-delà de ces différences globales, l'examen département par département révèle des disparités bien plus marquées. Certains territoires, caractérisés par une dynamique immobilière soutenue, présentent des LTV en moyenne plus élevées, traduisant une plus grande capacité de financement pour les emprunteurs. Toutefois, une hausse des rendements simulés ne se traduit pas mécaniquement par une augmentation de la LTV finale. En effet, les contraintes de rendement et de gestion du risque réagissent différemment selon les profils de longévité lors du calcul de LTV optimale. On observe ainsi un effet de 'tirage des extrémités' : les LTV déjà faibles en central tendent à être encore réduites, tandis que les LTV élevées sont rehaussées — particulièrement avec la modélisation par LSTM. Ce mécanisme s'explique par des profils de risques différents et traduit une surestimation du risque pour les profils peu risqués, et inversement une sous-estimation pour les profils plus exposés. Ce message est en cohérence avec d'autres études sur le sujet.

Cette hétérogénéité géographique illustre pleinement l'intérêt d'une segmentation fine dans

la tarification des PVH. Elle montre que le risque ne peut être appréhendé uniquement à travers une vision agrégée, mais qu'il dépend fortement des spécificités locales. La capacité à intégrer ces disparités territoriales dans les modèles constitue ainsi un levier essentiel pour la tarification des PVH.

Dans une logique opérationnelle, le prêteur dispose ainsi de plusieurs options, basées sur plusieurs méthodes de tarification. Une stratégie pourrait consister à combiner les résultats en retenant pour chaque individu la LTV minimale des différentes méthodes. Ces résultats doivent toutefois être interprétés comme des indications plutôt que comme des tarifs bruts : le modèle fournit un cadre de cohérence et un ordre de grandeur, mais il revient au prêteur de retraiter les niveaux de LTV jugés trop élevés ou de refuser certains profils et d'exercer son propre jugement en fonction des conditions de marché. Une tarification fondée sur l'ensemble des modèles segmentés, conçue dans une optique prudente, aboutit à une LTV de 43,7%, soit une baisse par rapport au scénario central traduisant la volonté de prudence. Tout de même 20% des individus voient le montant de prêt qui leur est proposé augmenter lors du passage à la tarification segmentée malgré la vision prudente appliquée, signe d'une meilleure prise en compte de leur profil de risque. À noter que les résultats sont très sensibles au modèle de tarification et aux contraintes appliquées lors de la maximisation. D'autres approches auraient pu être retenues et présenter des divergences de résultats.

Du point de vue de l'emprunteur, le recours à ce type de produit peut s'avérer particulièrement intéressant selon la situation individuelle. Il permet en effet de dégager des liquidités tout en conservant l'usage du logement. La segmentation par département joue alors un rôle déterminant, en autorisant une adaptation plus fine du niveau de prêt aux caractéristiques locales des marchés immobiliers. Cette granularité favorise une meilleure adéquation entre l'offre de financement et le profil de risque réel, renforçant simultanément l'équité entre emprunteurs et la maîtrise du risque pour le prêteur. Dans les zones dynamiques sur le plan immobilier, les emprunteurs bénéficient ainsi de conditions de financement plus favorables, avec des montants de prêts plus élevés rendus possibles grâce à l'utilisation de telles méthodes de modélisation.

5.2 Limites

L'ensemble des limites méthodologiques et pratiques propres à chacune des approches testées a été discuté dans les sections correspondantes. Néanmoins, plusieurs pistes d'amélioration transversale peuvent être dégagées et constituent des perspectives de recherche et de développement pertinentes.

Un premier axe concerne l'enrichissement des données individuelles disponibles. Les projections réalisées reposent sur des caractéristiques agrégées ou segmentées, ce qui impose une forme d'homogénéisation des comportements immobiliers. Disposer d'informations plus détaillées à l'échelle de chaque emprunteur permettrait de dépasser cette contrainte et d'envisager des projections immobilières véritablement individualisées. Une telle approche offrirait la possibilité de dissocier plus clairement la dynamique immobilière, liée à la valorisation des biens, de la modélisation du risque emprunteur, qui intègre les déterminants démographiques et comportementaux. L'articulation de ces deux dimensions renforcerait significativement la précision et la robustesse des simulations. Une limite importante réside aussi dans le niveau relativement élevé des rendements simulés. Cette caractéristique découle en grande partie du calibrage sur des séries historiques elles-mêmes marquées par des rendements immobi-

liers soutenus. Il importe toutefois de souligner que ces conditions particulières reflètent un contexte passé et que leur prolongement mécanique dans le futur ne peut être considéré comme acquis. Ce constat ouvre une piste de recherche intéressante : analyser l'influence de différents facteurs susceptibles d'affecter l'évolution future des prix immobiliers, notamment les dynamiques environnementales ou démographiques.

Un deuxième axe porte sur l'intégration explicite du coût du capital et de la notion de duration dans les indicateurs de performance. Les résultats présentés se concentrent principalement sur des mesures de rentabilité interne et de probabilité de déclenchement de la NNEG. Or, du point de vue du prêteur, la soutenabilité du produit repose également sur la mobilisation de fonds propres et sur la gestion des risques de liquidité. La prise en compte du coût du capital, de la structure des passifs et de la duration effective des flux permettrait de relier plus directement les projections de rendement aux contraintes prudentielles et de gestion actif-passif. Une telle intégration rapprocherait la modélisation académique des pratiques opérationnelles, et offrirait des outils plus pertinents pour la décision stratégique et la gestion de portefeuille.

Un troisième axe réside dans l'étude des options contractuelles et du développement de produits plus flexibles. Au-delà des sorties anticipées, qui constituent déjà un élément non modélisé impactant la rentabilité effective, il serait pertinent d'examiner l'introduction de facilités complémentaires, à l'instar des possibilités de tirages échelonnés mises en place sur certains marchés étrangers. De telles évolutions contribueraient à rendre le produit plus attractif pour l'emprunteur tout en ouvrant de nouvelles perspectives de gestion pour le prêteur. Le cadre présenté s'est principalement concentré sur le risque de longévité et la valorisation immobilière dans une perspective de détention jusqu'au terme. Or, dans la pratique, des événements tels que des remboursements anticipés ou des reventes interviennent et modifient substantiellement les trajectoires financières des contrats. Leur prise en compte systématique, élargie à l'ensemble des flexibilités contractuelles envisageables, permettrait d'analyser plus finement la sensibilité du portefeuille et de mieux évaluer la variabilité de la rentabilité effective. Du point de vue du prêteur, il s'agit d'un enjeu crucial, dans la mesure où la fréquence et la nature de ces options affectent directement le besoin de liquidité, la rotation du capital investi et, in fine, la rentabilité du produit.

En définitive, les résultats obtenus illustrent le potentiel de la modélisation fine des rendements immobiliers utilisés dans la tarification des prêts viagers hypothécaires. Au-delà des différences structurelles mises en évidence entre les méthodes testées, l'apport principal de cette étude réside dans les perspectives qu'elle ouvre : approfondir la compréhension du risque immobilier et démographique, améliorer la précision des outils de tarification et renforcer la capacité de gestion des portefeuilles. Pris sous l'angle du prêteur, ces développements futurs apparaissent importants pour consolider la soutenabilité du produit dans un cadre cohérent avec les exigences économiques et réglementaires.

Bibliographie

- [A. Bruchet, 2014] A. BRUCHET (2014). *Détermination du taux et des montants prêtés pour un prêt viager hypothécaire à taux fixe sous contrainte de rentabilité dans le cadre juridique français*. Mémoire d'actuariat - ISUP.
- [A. Graves, 2012] A. GRAVES (2012). *Supervised Sequence Labelling with Recurrent Neural Networks*. Canadian Institute for Advanced Research.
- [A. Graves, 2013] A. GRAVES (2013). *Generating sequences with recurrent neural networks*. Department of Computer Science University of Toronto.
- [A-K. Alexandridis et al., 2017] A-K. ALEXANDRIDIS ET AL. (2017). *Real Estate valuation and forecasting in non-homogeneous markets : A case study in Greece during the financial crisis*. University of Kent, Athens University of Economics and Business.
- [A. Lewden, 2024] A. LEWDEN (2024). *Modélisation d'investissement en viager immobilier*. Mémoire d'actuariat - ISUP.
- [A. Morisse, 2022] A. MORISSE (2022). *Application des réseaux de neurones récurrents à l'estimation des calculs réglementaires*. ENSAE.
- [A. Tealab, 2018] A. TEALAB (2018). *Time series forecasting using artificial neural networks methodologies : A systematic review*. Future University in Egypt.
- [A. Toure, 2021] A. TOURE (2021). *Modélisation des obligations catastrophes paramétriques dans un marché incomplet par une approche deep learning et une approche de modélisation par loi*. Institut du Risk Management.
- [A-W. Shao et al., 2013] A-W. SHAO ET AL. (2013). *Disaggregated House Price Indices*. UNSW Australian School of Business.
- [A-W. Shao et al., 2014] A-W. SHAO ET AL. (2014). *Reverse Mortgage Pricing and Risk Analysis Allowing for Idiosyncratic House Price Risk and Longevity Risk*. UNSW Australian School of Business.
- [Ahlgrim et al., 2005] AHLGRIM ET AL. (2005). *Modeling Financial Scenarios : A Framework for the Actuarial Profession*. Proceedings of the Casualty Actuarial Society 92.
- [B. Ge et al., 2021] B. GE ET AL. (2021). *Research on the Effect of Artificial Intelligence Real Estate Forecasting Using Multiple Regression Analysis and Artificial Neural Network : A Case Study of Ghana*. Anhui University of Science and Technology.
- [Banque de France, 2024] BANQUE DE FRANCE (2024). *Courbe zéro-coupon au 31.12.2024*. Banque de France - Publications et statistiques.
- [C. Jones, 2009] C. JONES (2009). *House Price Measurement : The Hybrid Hedonic Repeat-Sales Method*. The Economic Society of Australia.
- [C. Micciche et al., 2021] C. MICCICHE ET AL. (2021). *A methodology for local housing price index in France*. CY Cergy Paris Université, Laboratoire THEMA.
- [C. Vergnaud, 2018] C. VERGNAUD (2018). *Le viager immobilier est-il une opportunité de développement pour un assureur ?* Mémoire d'actuariat - IRM.
- [D. Crespine, 2017] D. CRESPINE (2017). *Le prêt viager hypothécaire, une solution pour financer la retraite ?* Mémoire d'actuariat - Dauphine.
- [D. Hopp, 2022] D. HOPP (2022). *Economic Nowcasting with Long Short-Term Memory Artificial Neural Networks (LSTM)*. Journal of Official Statistics, Vol. 38, No. 3, 2022.

- [DRESS, 2023] DRESS (2023). *Les retraités et les retraites*. Panoramas de la DREES, Social.
- [Equity Release Council, 2024a] EQUITY RELEASE COUNCIL (2024a). *Equity Release Council - full report Q4 24*.
- [Equity Release Council, 2024b] EQUITY RELEASE COUNCIL (2024b). *The legal guide to equity release*.
- [F. Planchet et al., 2012] F. PLANCHET ET AL. (2012). *Générateur de scénarios économiques en assurance*. WINTER et ASSOCIES.
- [F. Vermet, 2025] F. VERMET (2025). *Apprentissage statistique supervisé*. Cours EURIA, Master 1.
- [G. Parat, 2016] G. PARAT (2016). *Couverture assurantielle du risque de longévité dans le cadre d'un Prêt Viager Hypothécaire*. Mémoire d'actuariat - ISFA.
- [H. Alai et al., 2014] H. ALAI ET AL. (2014). *Developing Equity Release Markets : Risk Analysis for Reverse Mortgages and Home Reversions*. North American Actuarial Journal.
- [H. DiStefano, 2022] H. DiSTEFANO (2022). *Predicting Long-Term U.S. Housing Price Trends Using a Long Short-Term Memory Neural Network*. University of California.
- [I. Assayag, 2022] I. ASSAYAG (2022). *Le risque de longévité dans le cadre d'un viager immobilier*. Mémoire d'actuariat - ISUP.
- [IGF-CGEDD, 2021] IGF-CGEDD (2021). *Prix, loyers et couple rendement-risque du placement en logement*. IGF-CGEDD.
- [INED, 2016] INED (2016). *Tables de mortalité par sexe, âge et niveau de vie*. Institut National d'Etudes Démographiques.
- [INSEE, 2023] INSEE (2023). *Enquête Histoire de vie et Patrimoine*. INSEE.
- [INSEE, 2024] INSEE (T3 2024). *Indices Notaires-Insee des logements anciens*. INSEE.
- [INSEE, IFOP et Sociovision, 2020] INSEE, IFOP ET SOCIOVISION (2020). *EFPS*. Focus n°287 (2020).
- [INSEE Méthodes n° 132, 2019] INSEE MÉTHODES N° 132 (Juin 2019). *Les indices Notaires-Insee des prix des logements anciens Méthodologie*. INSEE.
- [J. Al-Qawasmi, 2022] J. AL-QAWASMI (2022). *Machine Learning Applications in Real Estate : Critical Review of Recent Development*. International Conference on Artificial Intelligence Applications and Innovations (AIAI).
- [J. Friggit, 2021] J. FRIGGIT (2021). *L'élasticité du prix des logements*. CGEDD.
- [J. Friggit, 2025] J. FRIGGIT (2025). *Le prix de l'immobilier d'habitation sur le long terme*. ENPC.
- [J. Hansen, 2006] J. HANSEN (2006). *Australian House Prices : A Comparison of Hedonic and Repeat-sales Measures*. Reserve Bank of Australia.
- [J. Hibbert et al., 2001] J. HIBBERT ET AL. (2001). *A stochastic asset model and calibration for long-term financial planning purposes*. Proceedings of the Casualty Actuarial Society.
- [J. Meszaros, 2024] J. MESZAROS (2024). *A Brief Review of House Price Forecasting Methods*. The Counselors of Real Estate, Volume 48, Number 4 - February 7, 2024.
- [K. Armel, 2025] K. ARMEL (2025). *Générateurs de Scénarios Economiques*. Cours EURIA, Master 2.
- [K-C. Chiu, 2024] K-C. CHIU (2024). *A long short-term memory model for forecasting housing prices in Taiwan in the post-epidemic era through big data analytics*. Asia Pacific Management Review.

- [K. Greff et al., 2015] K. GREFF ET AL. (2015). *LSTM : A search space odyssey*. IEEE Transactions on Neural Networks and Learning Systems - March 2015.
- [Legifrance, 2006] LEGIFRANCE (2006). *Loi n° 2006-872 du 13 juillet 2006 portant engagement national pour le logement*. France.
- [M. Crucianu et al., 2023] M. CRUCIANU ET AL. (2023). *Cours - Réseaux de neurones récurrents*. CNAM - RCP209 - Apprentissage, réseaux de neurones et modèles graphiques.
- [Mirabelle, 2024] MIRABELLE (2024). *Communiqué de presse, Lancement de la fintech Mirabelle*. Site officiel Mirabelle.
- [N. Richard, 2022] N. RICHARD (2022). *Cash flow prediction in reinsurance : short and long term projection for cash requirement and FX risk management*. Mémoire d'actuariat - ENSAE.
- [P. Ottou, 2017] P. OTTOU (2017). *Méthodes d'apprentissage automatique appliqués au provisionnement ligne à ligne en assurance non-vie*. Mémoire d'acturiat - Dauphine.
- [Paris Notaires Services, 2024] PARIS NOTAIRES SERVICES (2024). *Historique des prix au m² standardisés des appartements anciens à Paris par arrondissement*. Département immobilier - bien.
- [R. Pascanu et al., 2013] R. PASCANU ET AL. (2013). *On the difficulty of training Recurrent Neural Networks*. Université de Montreal.
- [Rapport annuel du COR, 2024] RAPPORT ANNUEL DU COR (2024). *Evolutions et perspectives des retraites en France*. COR.
- [Rapport d'évaluation des politiques de Sécurité Sociale, 2023] RAPPORT D'ÉVALUATION DES POLITIQUES DE SÉCURITÉ SOCIALE (2023). *Projet de loi d'approbation des comptes de la sécurité sociale*. Sécurité Sociale.
- [Ressources actuarielles, 2005] RESSOURCES ACTUARIELLES (2005). *Tables de mortalité TGF05 - TGH05*.
- [S-C. Bourassa et al., 2010] S-C. BOURASSA ET AL. (2010). *Land leverage and house prices*. Swiss Finance Institute.
- [S. Hochreiter, 1997] S. HOCHREITER (1997). *Long Short-Term Memory*. Neural Computation - November 1997.
- [S-S. Namini et al., 2019] S-S. NAMINI ET AL. (2019). *A Comparative Analysis of Forecasting Financial Time Series Using ARIMA, LSTM, and BiLSTM*. Texas Tech University and Georgia Institute of Technology.
- [S. Soleiman, 2023] S. SOLEIMAN (2023). *De la complexité spatiale aux prix de l'immobilier : approches statistiques*. Paris 1 - Panthéon-Sorbonne.
- [T. Rouer, 2023] T. ROUER (2023). *Estimation de la densité de provisions non-vie ligne à ligne par réseaux de neurones*. Mémoire d'actuariat - Dauphine.
- [T-S. Goma, 2020] T-S. GOMA (2020). *Apport des réseaux neuronaux au provisionnement IARD : Application d'architectures récurrentes à la prédiction d'évolution temporelle de montants*. Mémoire d'actuariat - ENSAE.
- [T. Tin Tin, 2023] T. TIN TIN (2023). *Real estate price forecasting utilizing recurrent neural networks incorporating genetic algorithms*. Innovative Research Publishing.
- [V-K. Vishwakarma, 2013] V-K. VISHWAKARMA (2013). *Forecasting real estate business : empirical evidence from the Canadian market*. St. Francis Xavier University, Global Journal of Business Research.

Table des figures

1.1	Evolution du ratio de dépendance en France	2
1.2	Indice des logements et des loyers	4
1.3	Flux financiers liés au PVH	7
1.4	Schéma général du PVH	8
1.5	LTV = 60%	9
1.6	LTV = 75%	9
1.7	Décès avant le croisement	9
1.8	Décès après le croisement	9
1.9	Rdt annuel fixe 2% annuel	10
1.10	Rdt annuel fixe 3,5% annuel	10
2.1	Comparaison territoriale du portefeuille et de la population nationale	19
2.2	Cartes des patrimoines immobiliers moyens	26
2.3	Cartes départementales des Prix/m ² et Surface moyenne (hors Paris)	26
2.4	Carte des âges moyens	27
2.5	Pyramide des âges par sexe	27
2.6	Evolution de la répartition de la population active par CSP	28
3.1	Structure du modèle Ahlgrim - sans action	37
3.2	Scénarios économiques simulés	46
3.3	Courbe d'actualisation	46
3.4	Contraintes déclenchées par couple Age - LTV optimale	52
3.5	Cartes des départementales LTV optimales et des âges moyens	53
3.6	LTV optimale - contraintes isolées	57
4.1	Indices immobiliers segmentés	68
4.2	Indices <i>backcastés</i> et rendements associés	70
4.3	Indices <i>backcastés</i> et rendements associés - Allignés avec l'indice INSEE	72
4.4	Comparaisons des rendements et indices	73
4.5	Rendement immobilier moyen, par segment	75
4.6	Structure d'un réseau de neurones	78
4.7	Focus sur le fonctionnement entre deux couches	78
4.8	Structure d'un réseau de neurones récurrent	82
4.9	Projection du rendement immobilier du segment i , par récursion	85
4.10	Projection du rendement immobilier par récursion, Paris-Appartement	86
4.11	Rendement immobilier moyen, par segment	87
4.12	Matrice de fréquence — LTV centrale \times LTV segmentée	88
4.13	Cartes des LTV optimales	89
4.14	Heatmap des écarts de LTV par Age	89
4.15	Contraintes limitantes en fonction de l'âge et la LTV optimale	90
4.16	Matrice de fréquence - LTV optimale LSTM \times Age	90

4.17	LTV moyenne par département et tranche d'âge – Central <i>vs.</i> Segmenté . . .	91
4.18	Exemples de quantiles et moyennes de rendements immobiliers	94
4.19	Cartes des écarts de LTV par région	100
1	Prix/m ² moyen et surface moyenne par région	119
2	Répartition (%) de la population par âge et patrimoine immobilier	119
3	Impact du nombre de simulations - Maille Portefeuille	120
4	Impact du nombre de simulations - Maille Individu	121
5	Évolution du prix moyen d'achat au m ² entre 2014 et 2023	131
6	Carte des LTV optimales par type de bien - Méthode 1 GSEs	132
7	Carte des LTV optimales par type de bien - Méthode 2 LSTM	132
8	Cartes des écarts de LTV par région	132
9	Matrice de fréquence - LTV optimale GSEs × Age	133
10	LTV moyenne par département et tranche d'âge – Central <i>vs.</i> Segmenté . . .	134
11	Matrice de fréquence LTV central - LTV prudente	135

Liste des tableaux

1.1	Evolution du marché de <i>l'Equity Release</i> au Royaume-Uni	11
2.1	Répartition des seniors par sexe et âge	18
2.2	Répartition des seniors par sexe, âge et situation matrimoniale	19
2.3	Coefficient moyen de correction de la surface par type de bien	21
2.4	Statistiques simulées par type de bien	22
2.5	Prix au mètre carré par type de bien et département (simulation)	22
2.6	Montant de patrimoine net début par CSP, début 2021	23
2.7	Montant de patrimoine net par classe d'âge, début 2021	23
2.8	Statistiques patrimoniales	25
2.9	Statistiques des actifs immobiliers par type de bien	25
3.1	Paramètres calibrés des modèles économiques	44
3.2	Comparaison des matrices de corrélation des résidus	45
3.3	Comparaison des moments simulés <i>vs.</i> historiques	47
3.4	Seuils de contrainte	51
3.5	Indicateurs globaux - Portefeuille immobilier	51
3.6	Indicateurs globaux - Portefeuille PVH	52
3.7	Impact du sexe et de la situation matrimoniale sur les indicateurs PVH	54
3.8	Impact de l'âge sur les indicateurs PVH	54
3.9	Impact du niveau de patrimoine sur les indicateurs PVH	55
3.10	Impact du taux d'emprunt sur les indicateurs PVH	56
3.11	Indicateurs globaux du Portefeuille PVH - contraintes isolées	56
4.1	Résultats de l'optimisation par <i>Optuna</i>	86
4.2	Métriques de performance — t+1 <i>vs.</i> récursif	86
4.3	Décomposition de la variance expliquée de la LTV optimale	92
4.4	Analyse de la variance — Effets explicatifs sur la LTV optimale	93
4.5	Comparaison des indicateurs statistiques des scénarios	93
4.6	Statistiques agrégées – Portefeuille immobilier	95
4.7	Statistiques agrégées – Portefeuille de PVH	96
4.8	Évolution des indicateurs pour les individus dont la LTV a diminué – Méthode de plafond	99
4.9	Évolution des indicateurs pour les individus dont la LTV a diminué – Méthode de lissage des indices	99
4.10	Indicateurs agrégées – Comparaison Central <i>vs.</i> Prudent	101
1	Variables de la base DVF+ <i>open-data</i>	117
2	Synthèse des caractéristiques des emprunteurs et des biens immobiliers	118
3	Indicateurs avant optimisation de la LTV - Toutes LTV testées	120
4	Indicateurs après optimisation de la LTV	120

5	Tests statistiques pour l'inflation	121
6	Tests statistiques pour le taux long	121
7	Tests statistiques pour le taux court	122
8	Tests statistiques pour les rendements immobiliers	122
9	Paramètres et métriques du Modèle Hybride	123
10	Paramètres et de l'Elastic Net	127
11	Indicateurs par tranche d'âge et par méthode	133
12	Statistiques agrégées par tranche d'âge — Tarification prudente	134
13	Statistiques agrégées par zone géographique — Tarification prudente	135

Annexes

.1 Outils utilisés

Python

Tout le projet a été réalisé en Python. Ce langage a servi à construire l'outil de tarification, implémenter les générateurs de scénarios économiques et entraîner les modèles LSTM de projection immobilière. L'appui de bibliothèques spécialisées a facilité la préparation des données, la modélisation, l'entraînement des réseaux de neurones et l'optimisation des hyperparamètres.

TensorFlow est une bibliothèque *open-source* développée par *Google* pour la création et l'entraînement de modèles d'apprentissage automatique. Elle offre un environnement flexible pour la construction de réseaux de neurones profonds, notamment des architectures récurrentes comme les LSTM, qui sont particulièrement adaptées aux données séquentielles et temporelles.

Keras est une API haut niveau intégrée à TensorFlow, qui simplifie la conception et l'entraînement des réseaux de neurones grâce à une interface intuitive et modulaire. Elle facilite notamment la création rapide de modèles complexes tout en conservant la puissance et la flexibilité de TensorFlow en arrière-plan.

Optuna est un cadre d'optimisation d'hyperparamètres automatisé. Il permet de définir un espace de recherche pour différents paramètres du modèle (par exemple, nombre d'unités LSTM, taux de *dropout*, dimension des *embeddings*), puis effectue une exploration intelligente basée sur des techniques d'optimisation bayésienne pour trouver la meilleure configuration réduisant la fonction de perte.

Excel et VBA

Excel et VBA ont été utilisés pour développer une maquette répliquante de l'outil de projection des cashflows et de calcul des indicateurs à la maille individu. Cette approche a permis de disposer d'un modèle transparent et facilement vérifiable, servant de référence pour valider les développements plus complexes réalisés ensuite sous Python.

IA générative

ChatGPT et d'autres outils d'Intelligence Artificielle ont constitué une aide précieuse tout au long de ce mémoire. Ils ont contribué aussi bien au développement et à la structuration du code qu'à la rédaction du mémoire.

.2 Base de données DVF

Le Cerema – Climat & Territoires de demain - publie plusieurs bases de données relatives aux transactions immobilières réalisées sur le territoire français. Parmi celles-ci, la base DVF+ *open-data*, élaborée conjointement avec la DGALN, met à disposition des données issues de Demande de Valeurs Foncières (DVF). Elle constitue un outil précieux pour l’observation et l’analyse des marchés fonciers et immobiliers, en offrant un accès libre à des fichiers structurés et directement exploitables, contenant les historiques de transactions de 2014 à 2024. D’autres bases de données plus détaillées sont disponibles sous certaines licences.

Dans le cadre de cette étude, les fichiers datant d’Avril 2025 ont été utilisés, disponibles via la plateforme dédiée. Ces fichiers ont été systématiquement agrégés afin de constituer une base unifiée, couvrant le territoire national et servant de socle à la construction des indices immobiliers segmentés utilisés dans cette étude.

TABLE 1 – Variables de la base DVF+ *open-data*

Variable	Description
idmutation	Identifiant de clef primaire de la table mutation
idmutinvar	Identifiant invariant de la mutation
idopendata	Identifiant <i>open-data</i> de la mutation
idnatmut	Identifiant de clef primaire de la table ann_nature_mutation
codservch	Code du service de conservation des hypothèques
refdoc	Référence d’enregistrement du document (acte de vente)
datemut*	Date de signature du document (acte de vente)
anneemut*	Année de signature du document
moismut*	Mois de signature du document
coddep*	Code du département
libnatmut*	Libellé de la nature de mutation
vefa	<i>Flag</i> VEFA (Vente en l’état futur d’achèvement)
valeurfonc*	Valeur foncière : prix ou évaluation déclarée pour la mutation
nbdispo	Nombre de dispositions associées à la mutation
nblot	Nombre total de lots dans la mutation
nbcomm	Nombre de communes concernées par la mutation
l_codinsee*	Liste des codes INSEE des communes concernées
nbsection	Nombre de sections concernées par la mutation
l_section*	Liste des sections concernées
nbpar	Nombre de parcelles concernées par la mutation
l_idpar*	Liste des identifiants des parcelles concernées (idpar)
nbparmut	Nombre de parcelles ayant muté
l_idparmut*	Liste des identifiants des parcelles ayant muté (idpar)
sterr*	Surface de terrain ayant muté
nbvolmut	Nombre de volumes ayant muté
nblocmut	Nombre de locaux ayant muté
l_idlocmut*	Liste des identifiants des locaux ayant muté (idloc)
nblocmai*	Nombre de maisons ayant muté
nblocapt*	Nombre d’appartements ayant muté
nblocdep	Nombre de dépendances ayant muté
nblocact	Nombre de locaux d’activité ayant muté
nbapt1pp–5pp*	Nb. d’appt. ayant muté selon le nb. de pièces princ. (1 à 5 et +)
nbmai1pp–5pp*	Nb. de maisons ayant muté selon le nb. de pièces princ. (1 à 5 et +)

Variable	Description
sbati*	Surface de l'ensemble du bâti ayant muté
sbatmai*	Surface de l'ensemble des maisons ayant muté
sbatapt*	Surface de l'ensemble des appartements ayant muté
sbatact	Surface de l'ensemble des locaux d'activité ayant muté
sapt1pp-5pp*	Surface des appt. ayant muté selon le nb. de pièces princ. (1 à 5 et +)
smai1pp-5pp*	Surface des maisons ayant muté selon le nb. de pièces princ. (1 à 5 et +)
codtybien*	Code du type de bien
libtybien*	Libellé du type de bien

Source : DVF+ *open-data* – Cerema ; (*) traduit l'utilisation de cette variable

.3 Portefeuille construit

Répartition du portefeuille

TABLE 2 – Synthèse des caractéristiques des emprunteurs et des biens immobiliers

Catégorie	Effectif	Part (%)
Age		
65 ans	2 459	30,0
70 ans	2 125	26,0
75 ans	1 539	18,8
80 ans	1 053	12,9
85 ans	884	10,8
90 ans	127	1,6
Catégorie Socio-Professionnelle		
Retraité ancien autre salarié	4 345	53,1
Retraité ancien cadre ou prof. libérale	3 104	37,9
Retraité ancien agriculteur, artisan, commerçant	738	9,0
Situation matrimoniale		
Couple	4 949	60,4
Seul	3 238	39,6
Région par type de bien		
Région	Appartement (%)	Maison (%)
Auvergne-Rhône-Alpes	46,9	53,1
Bourgogne-Franche-Comté	26,3	73,7
Bretagne	24,2	75,8
Centre-Val de Loire	24,6	75,4
Grand Est	28,6	71,4
Hauts-de-France	27,1	72,9
Normandie	28,0	72,0
Nouvelle-Aquitaine	27,4	72,6
Occitanie	35,2	64,8
Pays de la Loire	26,9	73,1
Provence-Alpes-Côte d'Azur	59,8	40,2
Île-de-France	73,6	26,4

FIGURE 1 – Prix/m² moyen et surface moyenne par région

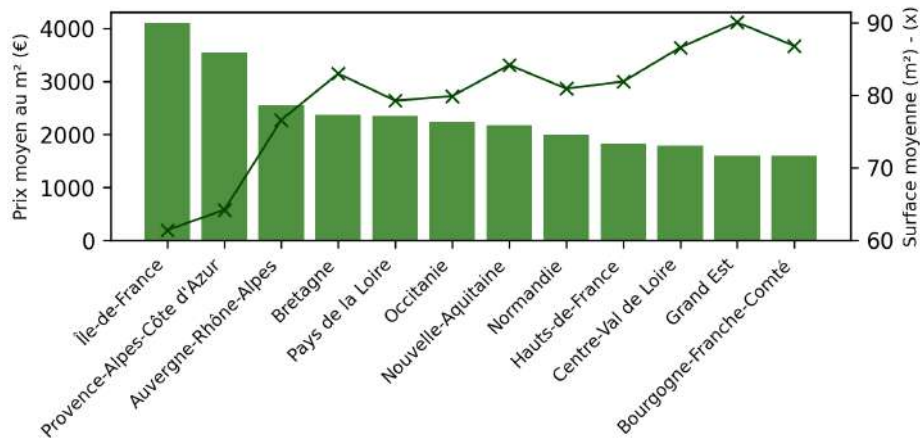


FIGURE 2 – Répartition (%) de la population par âge et patrimoine immobilier



4 Sensibilités

Sensibilité aux frais

Le passage des frais initiaux de 1% à 5% du montant prêté a un impact négatif visible sur les indicateurs économiques lorsque la LTV est maintenue constante. Avant optimisation, et en considérant l'ensemble des LTV testées, la rentabilité diminue : le rendement moyen chute de 64 points de base et la VAN moyenne baisse de plus de 3 700€. Cependant, l'optimisation de la LTV permet de compenser intégralement cette perte, en prêtant moins. En abaissant le taux de prêt moyen (LTV), l'opération devient plus prudente tout en maintenant une rentabilité élevée : la VAN reste stable, le rendement *flat* moyen dépasse 9,9%, et la probabilité de NNEG diminue. Cette approche montre que l'ajustement fin de la LTV est un levier efficace pour absorber l'effet des frais initiaux élevés. L'impact des frais est donc correctement intégré dans le modèle, et la dynamique d'ajustement observée est cohérente au regard des mécanismes économiques sous-jacents.

TABLE 3 – Indicateurs avant optimisation de la LTV - Toutes LTV testées

Indicateur	Central	Sensibilité frais 5%
Rendement <i>flat</i>	9,22 %	8,58 %
Rendement capitalisé	4,88 %	4,57 %
VAN moyenne (€)	84 211	80 470

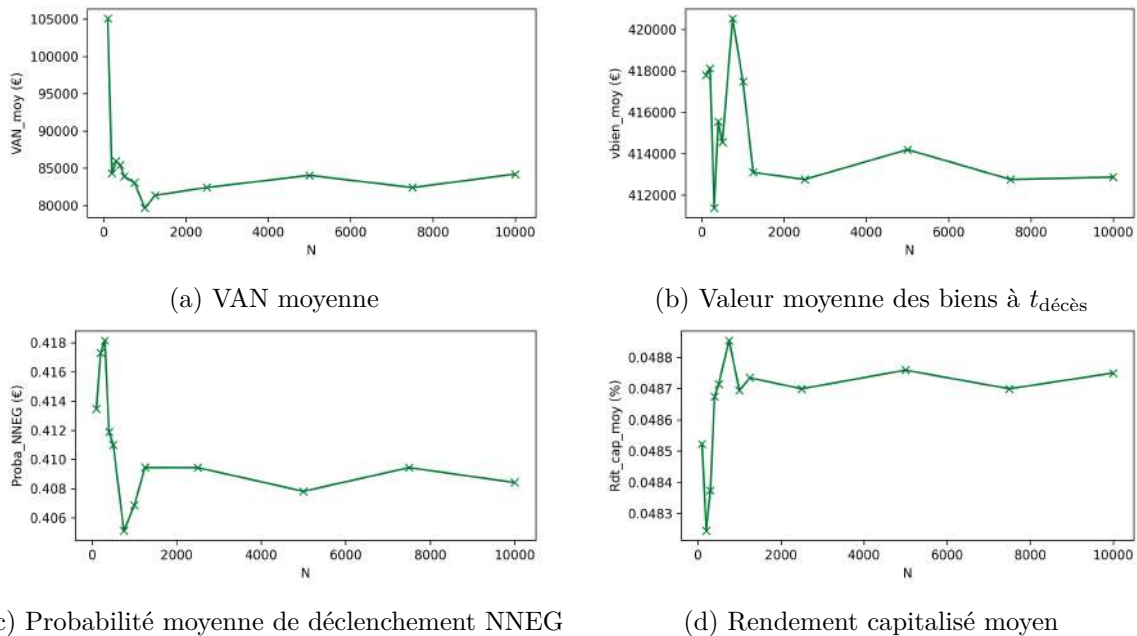
TABLE 4 – Indicateurs après optimisation de la LTV

Indicateur	Central	Sensibilité frais 5%
LTV moyenne	47,27 %	41,68 %
Rendement <i>flat</i>	10,25 %	9,95 %
Rendement capitalisé	5,29 %	5,08 %
Probabilité de NNEG	35,11 %	34,03 %
VAN moyenne (€)	105 174	105 179

Sensibilité au nombre de simulations

On observe une convergence rapide des indicateurs étudiés dès 2 000 simulations, tant pour les valeurs économiques (VAN, valeur du bien, probabilité de NNEG) que pour les rendements moyens, à la maille portefeuille. Cette stabilité des résultats suggère que le nombre de simulations pourrait être réduit sans perte significative de précision.

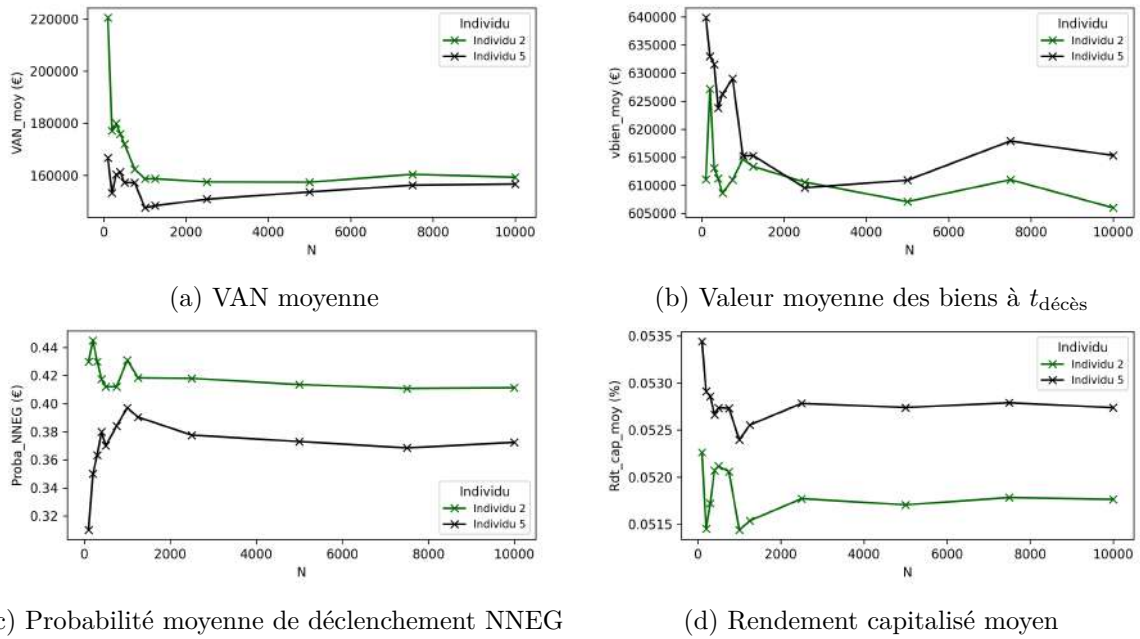
FIGURE 3 – Impact du nombre de simulations - Maille Portefeuille



La figure 4 représente les métriques de deux individus (choisis conjointement car leurs ordres de grandeurs sont similaires, avec LTV fixée à 50%). À l'échelle individuelle, la convergence des résultats est elle aussi rapide. Néanmoins, le recours à 10 000 simulations apporte une marge de sécurité supplémentaire. Grâce à la vectorisation du code développé, l'ajout de plusieurs milliers de simulations n'engendre qu'un surcoût computationnel limité. Une base de 10 000 simulations est donc utilisée dans cette étude, pour garantir la robustesse et la

fiabilité des estimations.

FIGURE 4 – Impact du nombre de simulations - Maille Individu



5.5 Calibrage GSE

TABLE 5 – Tests statistiques pour l'inflation

Inflation	Valeur	Interprétation
R^2 ajusté	0,805798	Fort pouvoir prédictif
Fisher F-stat	223,33 (p=2,052e-18)	Significativité globale du modèle
Moyenne des résidus	-10^{-6}	Erreurs centrées
Breusch-Godfrey	0,0000 (p=0,9976)	Absence d'autocorrélation d'ordre p = 1
ARCH(1)	3,7701 (p=0,0522)	Absence d'hétéroscédasticité d'ordre q = 1 au seuil 5%
Jarque-Bera	9,8216 (p=0,0074)	Hypothèse de normalité des résidus non vérifiée

TABLE 6 – Tests statistiques pour le taux long

Taux long	Valeur	Interprétation
R^2 ajusté	0,8305	Fort pouvoir prédictif
Fisher F-stat	211,68 (p=5,288e-18)	Significativité globale du modèle
Moyenne des résidus	-10^{-6}	Erreurs centrées
Breusch-Godfrey	0,0375 (p=0,8464)	Absence d'autocorrélation d'ordre 1
ARCH(1)	10,0758 (p=0,0015)	Hétéroscédasticité
Jarque-Bera	1,5454 (p=0,461)	Résidus normalement distribués

TABLE 7 – Tests statistiques pour le taux court

Taux court	Valeur	Interprétation
R^2 ajusté	-0,0166	Performance inférieure à la moyenne
Fisher F-stat	0,30 (p=0,5881)	Pas de significativité globale
Moyenne des résidus	$-8,37 \times 10^{-4}$	Erreurs centrées
Breusch-Godfrey	0,9976 (p=0,3179)	Absence d'autocorrélation d'ordre p = 1
ARCH(1)	0,1537 (p=0,6950)	Absence d'hétéroscédasticité d'ordre q = 1
Jarque-Bera	6,6349 (p=0,0362)	Hypothèse de normalité des résidus non vérifiée

TABLE 8 – Tests statistiques pour les rendements immobiliers

Rdts. immo.	Valeur	Interprétation
R^2 ajusté	0,6391	Modèle explicatif à 63%
Fisher F-stat	112,56 (p=1,44e-15)	Significativité globale du modèle
Moyenne des résidus	-10^{-6}	Erreurs centrées
Breusch-Godfrey	2,373 (p=0,124)	Absence d'autocorrélation d'ordre 1
ARCH(1)	2,232 (p=0,135)	Absence d'hétéroscédasticité d'ordre q = 1
Jarque-Bera	85,1005 (p=0,0000)	Hypothèse de normalité des résidus non vérifiée

6.6 Construction d'un indice

TABLE 9 – Paramètres et métriques du Modèle Hybride

Dep	Type	R ² reprod.	RMSE	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	Intercept	log(sbati)	nb_piece
1	111	0.973	1.021	0.007	0.007	0.039	0.062	0.082	0.131	0.222	0.311	0.297	0.229	8.120	0.736	0.148
1	121	0.985	1.021	-0.057	-0.060	-0.023	0.014	0.039	0.036	0.075	0.139	0.117	0.079	8.219	0.910	-0.023
2	111	0.970	1.018	-0.020	-0.021	-0.009	-0.015	-0.003	0.036	0.081	0.132	0.091	0.037	7.775	0.718	0.110
2	121	0.989	1.014	-0.029	-0.088	0.066	0.142	-0.011	-0.004	0.040	0.084	0.070	0.062	8.011	0.771	0.002
3	111	0.959	1,021	-0,033	-0,045	-0,026	-0,020	-0,023	0,013	0,023	0,094	0,073	0,061	7,077	0,843	0,128
3	121	0.986	1,024	-0,077	-0,061	-0,081	-0,054	-0,039	-0,010	0,063	0,136	0,141	0,100	6,942	1,053	-0,087
4	111	0.966	1,013	-0,005	0,018	-0,008	-0,029	-0,013	0,033	0,093	0,153	0,230	0,187	7,905	0,851	0,076
4	121	0.990	1,021	-0,024	0,033	0,037	-0,072	-0,036	0,039	0,030	0,124	0,186	0,141	8,490	0,687	0,037
5	111	0.967	1,013	-0,026	-0,007	0,001	0,013	0,043	0,104	0,120	0,180	0,298	0,226	8,341	0,735	0,110
5	121	0.990	1,020	-0,007	-0,039	-0,044	-0,036	-0,043	-0,013	0,068	0,135	0,187	0,198	8,729	0,647	0,117
6	111	0.976	1,015	0,002	-0,020	-0,005	0,030	0,041	0,059	0,116	0,224	0,225	0,172	7,610	1,154	-0,000
6	121	0.991	1,012	-0,004	-0,011	0,016	0,024	0,050	0,082	0,121	0,189	0,219	0,214	8,255	1,014	-0,050
7	111	0.968	1,018	-0,004	0,003	0,029	0,041	0,036	0,071	0,124	0,234	0,227	0,230	7,702	0,857	0,066
7	121	0.986	1,020	-0,040	-0,060	0,021	-0,044	-0,027	0,031	0,049	0,100	0,143	0,120	7,899	0,818	0,007
8	111	0.960	1,015	-0,008	-0,030	0,014	0,019	-0,004	0,036	0,084	0,128	0,139	0,085	7,689	0,665	0,172
8	121	0.988	1,018	-0,039	-0,074	-0,051	0,030	-0,040	0,036	0,071	0,108	0,169	0,067	7,515	0,807	0,029
9	111	0.963	1,016	-0,026	-0,047	-0,042	-0,037	-0,034	0,015	0,066	0,179	0,186	0,183	8,166	0,631	0,128
9	121	0.991	1,024	-0,005	-0,030	-0,112	-0,118	-0,099	-0,007	0,030	0,081	0,141	0,155	9,076	0,461	0,054
10	111	0.970	1,016	-0,005	0,000	-0,011	-0,023	-0,007	0,019	0,078	0,144	0,137	0,093	8,125	0,635	0,157
10	121	0.993	1,031	-0,051	-0,021	-0,009	-0,009	-0,025	0,037	0,146	0,213	0,179	0,111	8,628	0,618	0,028
11	111	0.973	1,037	-0,025	-0,003	-0,006	0,009	0,026	0,042	0,108	0,175	0,173	0,144	9,223	0,417	0,154
11	121	0.990	1,023	0,006	0,006	-0,014	-0,044	-0,085	-0,070	0,008	0,098	0,140	0,129	9,977	0,236	0,164
12	111	0.947	1,016	0,022	0,015	0,023	0,024	0,012	0,065	0,093	0,163	0,213	0,144	7,919	0,656	0,155
12	121	0.990	1,040	-0,032	-0,008	0,051	0,077	0,080	0,119	0,172	0,256	0,251	0,212	7,790	0,835	0,019
13	111	0.983	1,018	0,003	0,023	0,057	0,075	0,114	0,150	0,223	0,303	0,319	0,281	8,316	0,932	0,003
13	121	0.990	1,025	-0,008	-0,033	-0,017	-0,036	-0,013	0,037	0,089	0,162	0,172	0,138	8,118	0,997	-0,086
14	111	0.973	1,029	-0,011	0,018	0,034	0,029	0,047	0,136	0,230	0,323	0,322	0,306	8,347	0,720	0,081
14	121	0.991	1,045	-0,042	-0,022	0,010	0,011	0,046	0,107	0,222	0,307	0,315	0,306	9,046	0,624	0,049
15	111	0.945	1,017	-0,043	-0,008	-0,023	-0,009	-0,001	0,016	0,019	0,136	0,132	0,129	7,587	0,727	0,130
15	121	0.990	1,036	-0,035	-0,037	-0,035	-0,012	-0,005	0,060	0,068	0,171	0,126	0,055	8,401	0,616	0,053
16	111	0.968	1,026	-0,003	-0,021	-0,002	0,010	0,048	0,083	0,151	0,251	0,240	0,183	7,376	0,771	0,144
16	121	0.994	1,057	-0,094	-0,075	-0,078	-0,038	0,021	0,109	0,214	0,298	0,326	0,313	8,518	0,607	0,039
17	111	0.968	1,030	0,007	0,009	0,061	0,089	0,114	0,174	0,259	0,357	0,367	0,312	9,457	0,430	0,149
17	121	0.990	1,048	-0,000	0,030	0,060	0,055	0,059	0,126	0,206	0,288	0,284	0,271	9,201	0,656	0,011
18	111	0.963	1,015	0,016	0,011	-0,025	-0,014	-0,017	-0,007	0,018	0,094	0,068	0,058	7,101	0,832	0,140
18	121	0.989	1,020	-0,069	-0,138	-0,110	-0,175	-0,120	-0,115	-0,040	0,010	0,026	0,019	7,614	0,882	-0,006
19	111	0.965	1,023	0,017	-0,017	0,011	0,022	0,048	0,077	0,130	0,164	0,207	0,162	7,502	0,813	0,076
19	121	0.989	1,034	-0,013	-0,055	-0,029	-0,032	0,028	0,039	0,088	0,192	0,192	0,201	8,017	0,746	0,016
21	111	0.965	1,019	-0,043	-0,049	-0,004	0,013	0,019	0,066	0,094	0,183	0,136	0,151	8,083	0,640	0,215
21	121	0.991	1,023	0,014	-0,020	0,024	0,023	0,077	0,118	0,198	0,224	0,189	0,139	8,457	0,800	-0,037
22	111	0.963	1,030	-0,017	-0,025	0,010	0,021	0,070	0,125	0,219	0,335	0,385	0,356	7,692	0,771	0,113
22	121	0.987	1,071	-0,046	-0,022	-0,010	0,026	0,009	0,060	0,184	0,314	0,353	0,416	8,419	0,759	-0,049
23	111	0.941	1,015	-0,079	-0,077	-0,103	-0,111	-0,149	-0,121	-0,070	0,029	0,043	0,024	7,282	0,737	0,123
23	121	0.988	1,027	-0,021	-0,005	-0,087	-0,085	-0,066	-0,082	-0,121	-0,024	0,182	0,096	7,141	0,879	-0,013
24	111	0.971	1,022	0,006	0,015	0,013	0,042	0,042	0,082	0,159	0,230	0,252	0,222	7,471	0,855	0,062
24	121	0.994	1,037	-0,056	-0,067	0,005	0,008	-0,003	0,024	0,088	0,165	0,198	0,173	7,970	0,782	0,028
25	111	0.973	1,013	-0,012	-0,005	0,026	0,018	0,038	0,093	0,162	0,230	0,192	0,134	8,551	0,659	0,081
25	121	0.988	1,020	0,027	-0,017	0,055	0,067	0,033	0,060	0,135	0,188	0,124	0,069	8,310	0,815	-0,059
26	111	0.983	1,018	-0,011	0,007	0,017	0,030	0,035	0,101	0,182	0,270	0,285	0,209	8,245	0,774	0,066
26	121	0.991	1,027	0,025	-0,007	0,027	0,028	0,043	0,077	0,125	0,203	0,200	0,196	7,934	0,906	-0,077
27	111	0.985	1,015	-0,014	-0,021	-0,006	0,026	0,014	0,063	0,141	0,216	0,205	0,126	8,383	0,745	0,027
27	121	0.994	1,026	0,007	-0,063	-0,064	-0,022	-0,038	0,028	0,073	0,129	0,111	0,075	7,918	0,883	-0,016
28	111	0.979	1,017	-0,020	-0,052	-0,012	-0,030	-0,004	0,024	0,077	0,141	0,139	0,085	8,224	0,718	0,094
28	121	0.992	1,019	-0,028	-0,012	0,045	-0,030	-0,090	-0,037	-0,012	0,066	0,046	-0,001	7,837	0,989	-0,094
29	111	0.972	1,035	-0,008	0,007	0,020	0,042	0,071	0,128	0,252	0,378	0,431	0,371	7,887	0,770	0,083
29	121	0.991	1,086	0,004	-0,027	0,003	0,002	0,042	0,150	0,286	0,424	0,449	0,452	8,156	0,775	-0,013
30	111	0.981	1,026	0,016	0,011	0,031	0,053	0,061	0,100	0,163	0,249	0,268	0,219	7,950	0,862	0,055
30	121	0.983	1,030	-0,009	-0,030	-0,006	0,003	0,013	0,039	0,118	0,159	0,165	0,155	9,336	0,560	-0,039
31	111	0.978	1,015	0,020	0,028	0,048	0,070	0,095	0,125	0,164	0,238	0,222	0,132	8,840	0,617	0,135
31	121	0.993	1,031	0,026	0,039	0,048	0,041	0,029	0,082	0,121	0,148	0,120	0,085	8,621	0,830	-0,037
32	111	0.967	1,021	-0,021	-0,031	-0,068	-0,033	0,028	0,050	0,123	0,208	0,195	0,149	7,966	0,735	0,084
32	121	0.993	1,045	0,133	0,115	0,144	0,126	0,074	0,151	0,175	0,263	0,302	0,330	8,108	0,656	0,112
33	111	0.972	1,029	0,022	0,048	0,114	0,165	0,214	0,267	0,352	0,430	0,397	0,325	9,266	0,542	0,120
33	121	0.993	1,055	0,014	0,051	0,119	0,160	0,162	0,196	0,253	0,284	0,264	0,207	8,634	0,880	-0,074
34	111	0.976	1,031	0,015	0,005	0,028	0,044	0,067	0,110	0,180	0,242	0,262	0,220	8,187	0,794	0,099
34	121	0.990	1,017	-0,001	-0,010	0,017	-0,004	0,003	0,038	0,087	0,163	0,178	0,152	9,149	0,668	-0,006
35	111	0.970	1,027	0,006	0,009	0,030	0,059	0,094	0,143	0,236	0,352	0,343	0,308	8,087	0,706	0,161
35	121	0.993	1,058	0,028	0,030	0,062	0,071	0,093	0,175	0,266	0,351	0,341	0,307	8,813	0,730	0,009
36	111	0.957	1,016	-0,022	-0,072	-0,067	-0,039	-0,063	-0,052	-0,019	0,058	0,046	-0,027	7,032	0,848	0,120
36	121	0.994	1,031	-0,059	-0,024	-0,085	-0,077	-0,100	-0,047	0,014	0,130	0,159	0,129	7,165	0,962	-0,057
37	111	0.974	1,021	-0,010	0,013	0,030	0,034	0,067	0,084	0,150	0,221	0,215	0			

Dep	Type	R ² reprod.	RMSE	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	Intercept	log(sbati)	nb_piece
40	111	0.975	1.037	0.015	0.049	0.112	0.117	0.172	0.219	0.324	0.409	0.429	0.370	9.203	0.592	0.022
40	121	0.990	1.042	0.046	0.069	0.104	0.118	0.089	0.137	0.273	0.331	0.391	0.364	9.823	0.359	0.142
41	111	0.972	1.015	-0.017	-0.043	-0.062	-0.065	-0.048	-0.024	0.035	0.107	0.120	0.066	7.419	0.879	0.068
41	121	0.989	1.029	-0.058	-0.043	-0.058	-0.116	-0.161	-0.071	0.034	0.165	0.170	0.191	7.702	0.879	0.001
42	111	0.974	1.022	0.001	0.024	0.031	0.045	0.062	0.110	0.179	0.268	0.256	0.234	8.112	0.698	0.132
42	121	0.988	1.023	-0.049	-0.104	-0.083	-0.068	-0.064	0.000	0.097	0.177	0.137	0.077	7.713	0.810	0.035
43	111	0.945	1.021	-0.025	-0.036	-0.032	-0.018	-0.012	0.051	0.068	0.170	0.186	0.185	7.309	0.804	0.133
43	121	0.989	1.041	0.013	-0.017	-0.019	-0.039	-0.049	-0.040	0.112	0.193	0.236	0.161	7.356	0.901	0.022
44	111	0.976	1.030	-0.023	-0.001	0.039	0.061	0.104	0.168	0.266	0.354	0.339	0.287	8.870	0.625	0.115
44	121	0.992	1.045	0.001	-0.030	0.005	0.010	0.066	0.169	0.259	0.291	0.265	0.204	8.602	0.836	-0.026
45	111	0.981	1.012	-0.032	-0.031	-0.010	0.004	0.005	0.025	0.086	0.142	0.131	0.088	8.236	0.741	0.070
45	121	0.992	1.037	-0.042	-0.011	0.007	-0.001	-0.025	0.010	0.088	0.151	0.174	0.106	8.471	0.787	-0.044
46	111	0.963	1.017	-0.016	-0.006	-0.007	-0.002	0.022	0.025	0.103	0.206	0.209	0.194	7.531	0.861	0.050
46	121	0.991	1.025	-0.131	0.002	-0.032	0.037	0.008	0.022	0.072	0.132	0.220	0.167	7.704	0.819	0.035
47	111	0.971	1.029	0.004	-0.001	0.040	0.006	0.029	0.066	0.132	0.223	0.210	0.180	7.759	0.764	0.090
47	121	0.993	1.056	-0.054	-0.028	0.063	0.092	0.085	0.107	0.220	0.285	0.296	0.235	7.749	0.799	0.009
48	111	0.941	1.013	-0.009	0.037	-0.037	-0.046	0.021	0.047	0.091	0.200	0.211	0.175	7.852	0.721	0.100
48	121	0.978	1.052	0.009	-0.113	0.085	-0.045	-0.030	0.066	0.058	0.115	0.199	0.185	8.138	0.651	0.094
49	111	0.973	1.030	-0.003	0.001	0.017	0.025	0.065	0.124	0.230	0.334	0.332	0.299	7.929	0.728	0.133
49	121	0.991	1.100	0.003	0.011	0.046	0.049	0.079	0.228	0.383	0.463	0.460	0.432	8.462	0.746	-0.005
50	111	0.967	1.027	0.016	0.034	0.063	0.068	0.090	0.157	0.230	0.323	0.369	0.319	8.293	0.638	0.113
50	121	0.987	1.062	-0.038	-0.061	-0.024	-0.015	-0.029	0.003	0.139	0.202	0.243	0.264	8.717	0.667	0.015
51	111	0.974	1.018	-0.029	-0.016	-0.007	0.018	0.024	0.062	0.111	0.159	0.152	0.101	8.312	0.656	0.139
51	121	0.991	1.028	-0.005	-0.042	0.010	0.008	0.049	0.117	0.173	0.230	0.154	0.081	8.355	0.812	-0.017
52	111	0.943	1.020	-0.017	-0.042	-0.041	0.000	-0.030	-0.017	0.009	0.118	0.083	0.054	8.054	0.404	0.325
52	121	0.994	1.026	-0.077	-0.123	-0.098	-0.095	-0.094	-0.044	-0.021	0.040	0.116	0.064	7.642	0.812	0.013
53	111	0.963	1.021	-0.016	-0.049	-0.007	-0.002	-0.009	0.018	0.092	0.182	0.198	0.179	7.425	0.744	0.177
53	121	0.991	1.084	-0.006	-0.004	0.092	0.061	0.119	0.196	0.311	0.423	0.403	0.430	7.532	0.896	-0.014
54	111	0.974	1.030	-0.005	0.018	0.043	0.059	0.082	0.137	0.182	0.296	0.263	0.200	8.228	0.628	0.153
54	121	0.991	1.027	-0.034	-0.045	-0.010	0.027	0.073	0.090	0.153	0.189	0.145	0.133	8.454	0.741	-0.023
55	111	0.955	1.018	-0.011	-0.017	-0.001	0.004	-0.033	0.023	0.035	0.124	0.127	0.085	7.840	0.554	0.224
55	121	0.988	1.022	-0.067	-0.110	-0.009	-0.022	0.051	0.050	0.122	0.140	0.135	0.170	7.866	0.704	0.061
56	111	0.962	1.03	-0.028	-0.008	0.021	0.044	0.072	0.134	0.235	0.362	0.407	0.396	7.963	0.729	0.154
56	121	0.991	1.072	-0.061	-0.054	-0.025	-0.017	0	0.084	0.21	0.332	0.367	0.339	9.06	0.651	0.009
58	111	0.955	1.013	-0.026	-0.018	-0.031	-0.033	-0.061	-0.01	0.017	0.091	0.107	0.02	7.016	0.836	0.124
58	121	0.98	1.027	-0.089	-0.128	-0.04	0.077	-0.158	-0.281	-0.188	0.014	0.106	0.072	8.439	0.498	0.123
59	111	0.977	1.029	-0.005	0.002	0.02	0.021	0.04	0.086	0.142	0.206	0.171	0.127	7.393	0.921	0.07
59	121	0.991	1.024	-0.006	-0.026	0.022	0.035	0.022	0.049	0.12	0.147	0.095	0.023	9.173	0.651	-0.002
60	111	0.983	1.014	-0.001	0.014	0.033	0.043	0.061	0.093	0.145	0.194	0.158	0.111	8.693	0.705	0.038
60	121	0.991	1.018	-0.024	-0.048	-0.05	-0.034	0.002	-0.012	0.045	0.077	0.047	-0.024	8.323	0.899	-0.093
61	111	0.963	1.015	-0.047	-0.063	-0.036	-0.042	-0.043	0.023	0.078	0.165	0.143	0.142	7.192	0.881	0.071
61	121	0.99	1.023	-0.067	-0.008	-0.081	-0.051	-0.097	0.007	0.05	0.14	0.166	0.151	8.015	0.761	-0.044
62	111	0.98	1.028	-0.018	-0.018	0.007	0.02	0.024	0.059	0.113	0.183	0.18	0.127	7.758	0.842	0.041
62	121	0.985	1.029	-0.01	-0.014	0.002	0.057	0.048	0.048	0.11	0.128	0.141	0.113	9.089	0.609	0.058
63	111	0.956	1.017	0.014	0.014	0.036	0.021	0.058	0.073	0.103	0.19	0.185	0.157	7.48	0.777	0.169
63	121	0.992	1.02	0.026	0.013	0.023	0.026	0.017	0.069	0.14	0.206	0.189	0.144	8.182	0.802	0.013
64	111	0.964	1.021	0.001	0.045	0.074	0.089	0.095	0.132	0.223	0.31	0.303	0.278	8.421	0.739	0.074
64	121	0.985	1.043	-0.036	-0.006	0.029	0.034	0.051	0.085	0.168	0.25	0.251	0.229	8.592	0.829	-0.027
65	111	0.976	1.022	0.001	-0.012	-0.012	0.008	0.039	0.046	0.118	0.198	0.229	0.214	8.982	0.518	0.096
65	121	0.989	1.03	0.015	-0.044	-0.008	-0.05	-0.012	-0.002	0.059	0.146	0.15	0.212	9.482	0.367	0.096
66	111	0.981	1.04	-0.002	-0.001	0.015	0.016	-0.003	0.071	0.129	0.217	0.227	0.237	8.408	0.755	0.051
66	121	0.989	1.022	-0.017	-0.062	-0.044	-0.057	-0.053	-0.028	0.05	0.112	0.149	0.134	9.249	0.529	0.041
69	111	0.976	1.021	-0.003	0.019	0.058	0.094	0.131	0.195	0.297	0.394	0.339	0.258	8.584	0.73	0.119
69	121	0.992	1.038	0.008	0.019	0.082	0.117	0.166	0.235	0.294	0.335	0.276	0.21	8.606	0.877	-0.058
70	111	0.955	1.016	0.012	-0.021	-0.019	-0.012	0	0.016	0.085	0.182	0.151	0.107	8.052	0.55	0.202
70	121	0.988	1.019	-0.04	-0.097	-0.055	0.028	0.051	-0.003	0.058	0.118	0.153	0.068	7.915	0.702	0.028
71	111	0.967	1.018	-0.013	-0.029	-0.028	-0.016	-0.001	0.025	0.097	0.151	0.18	0.131	7.371	0.837	0.107
71	121	0.988	1.019	0.035	-0.047	-0.016	-0.03	0.011	0.032	0.073	0.174	0.173	0.149	7.219	0.956	-0.023
72	111	0.972	1.018	-0.004	-0.003	-0.01	-0.027	0.007	0.022	0.082	0.164	0.165	0.13	7.605	0.759	0.151
72	121	0.994	1.054	-0.013	-0.048	-0.044	-0.048	-0.024	0.017	0.144	0.253	0.273	0.255	7.912	0.837	-0.023
73	111	0.959	1.016	-0.016	-0.022	0.022	0.07	0.086	0.124	0.249	0.351	0.374	0.336	8.002	0.854	0.081
73	121	0.986	1.019	-0.005	0.006	0.026	0.062	0.06	0.094	0.169	0.263	0.296	0.273	9.08	0.663	0.055
74	111	0.974	1.01	-0.01	0.002	0.024	0.055	0.092	0.128	0.223	0.296	0.332	0.289	8.769	0.81	0.048
74	121	0.992	1.018	-0.008	0.02	0.034	0.047	0.087	0.119	0.173	0.217	0.234	0.228	8.772	0.808	0.04
75	111	0.966	1.013	0.031	0.092	0.146	0.188	0.33	0.287	0.345	0.267	0.206	0.242	8.771	1.062	0.028
75	121	0.984	1.005	0.009	0.026	0.103	0.139	0.187	0.257	0.268	0.273	0.209	0.16	9.021	0.96	0.019
76	111	0.984	1.022	-0.007	-0.011	0.014	0.021	0.035	0.073	0.137	0.201	0.208	0.162	8.149	0.796	0.046
76	121	0.991	1.034	-0.017	-0.047	-0.035	0.022	0.013	0.053	0.134	0.181	0.135	0.084	8.106	0.894	-0.049
77	111	0.987	1.014	-0.01	-0.005	0.03	0.051	0.074	0.106	0.166	0.227	0.196	0.112	9.499	0.55	0.077
77	121	0.995	1.018	0.021	-0.009	-0.01	0.031	0.054	0.083	0.121	0.17	0.109	0.05	9.396	0.645	-0.007
78	111	0.983	1.01	-0.004	-0.002	0.003	0.053	0.065	0.093	0.163	0.211	0.169	0.127	9.018	0.815	-0.008
78	121	0.993	1.019	-0.014	-0.026	-0.004	0.024									

Dep	Type	R ² reprod.	RMSE	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	Intercept	log(sbati)	nb_piece
84	111	0,981	1,019	-0,022	-0,021	0,007	-0,031	-0,025	0,086	0,159	0,234	0,247	0,172	7,755	0,989	0,004
84	121	0,99	1,027	-0,054	-0,099	-0,086	-0,105	-0,081	-0,076	-0,021	0,032	0,053	0,003	8,38	0,884	-0,122
85	111	0,974	1,032	-0,009	0,006	0,039	0,066	0,109	0,175	0,271	0,386	0,39	0,345	9,451	0,441	0,107
85	121	0,99	1,056	-0,023	0,004	0,03	0,006	-0,004	0,055	0,186	0,291	0,311	0,298	8,908	0,692	0,011
86	111	0,967	1,018	-0,001	-0,004	0,002	0,002	0,018	0,038	0,07	0,146	0,153	0,088	7,655	0,696	0,189
86	121	0,993	1,058	0,058	0,045	0,059	0,029	0,051	0,124	0,238	0,32	0,305	0,286	8,38	0,729	-0,022
87	111	0,963	1,02	-0,012	-0,005	-0,002	0,002	0,016	0,052	0,09	0,182	0,152	0,104	7,652	0,701	0,174
87	121	0,993	1,037	-0,004	0,016	0,035	0,033	0,036	0,118	0,227	0,3	0,273	0,22	7,695	0,855	-0,017
88	111	0,955	1,017	-0,019	-0,016	-0,026	-0,006	0,022	0,067	0,107	0,208	0,191	0,192	8,363	0,47	0,221
88	121	0,982	1,026	-0,005	-0,059	0,014	-0,014	0,044	0,066	0,12	0,201	0,208	0,198	8,811	0,502	0,055
89	111	0,967	1,012	-0,037	-0,04	-0,032	-0,031	-0,066	-0,033	0,023	0,075	0,069	0,005	7,825	0,732	0,111
89	121	0,991	1,017	-0,022	0,01	-0,007	-0,04	-0,077	-0,043	0,007	0,061	0,098	0,074	8,346	0,691	0,018
90	111	0,984	1,017	0,014	0,014	0,044	0,039	0,059	0,118	0,114	0,193	0,162	0,078	8,038	0,798	0,033
90	121	0,988	1,017	0,034	-0,002	0,011	0,023	0,032	0,085	0,103	0,173	0,045	0,073	7,157	0,979	-0,019
91	111	0,991	1,01	-0,002	0,013	0,034	0,061	0,087	0,125	0,184	0,229	0,193	0,112	9,812	0,581	0,013
91	121	0,995	1,016	-0,026	-0,055	-0,036	-0,025	0,004	0,057	0,089	0,137	0,068	-0,002	9,236	0,694	-0,026
92	111	0,983	1,009	-0,026	0,01	0,034	0,108	0,169	0,215	0,257	0,274	0,232	0,143	9,477	0,829	0,011
92	121	0,993	1,013	-0,006	-0,003	0,057	0,111	0,154	0,208	0,238	0,264	0,209	0,144	8,641	1,003	-0,03
93	111	0,987	1,014	-0,005	-0,003	0,031	0,071	0,124	0,179	0,239	0,266	0,228	0,171	9,874	0,603	-0,008
93	121	0,992	1,023	-0,034	-0,042	0,028	0,063	0,107	0,17	0,214	0,234	0,178	0,097	8,821	0,859	-0,064
94	111	0,985	1,011	0,001	0,033	0,054	0,111	0,131	0,206	0,26	0,319	0,282	0,197	9,894	0,663	-0,016
94	121	0,993	1,021	-0,008	-0,006	0,038	0,075	0,118	0,171	0,209	0,238	0,181	0,136	8,915	0,878	-0,051
95	111	0,991	1,014	-0,013	0,01	0,05	0,066	0,106	0,15	0,211	0,255	0,199	0,131	9,714	0,62	0
95	121	0,996	1,019	-0,005	-0,033	-0,022	0,021	0,054	0,086	0,127	0,157	0,104	0,043	9,274	0,703	-0,038

7 Projection passée

Principes théoriques

Régression Lasso

La régression Lasso (*Least Absolute Shrinkage and Selection Operator*) est une extension de la régression linéaire introduisant une régularisation ℓ_1 qui permet à la fois de contraindre les coefficients et d'effectuer une sélection automatique des variables. Elle est particulièrement utile dans les contextes de forte colinéarité ou lorsque le nombre de variables explicatives est élevé par rapport au nombre d'observations.

Formellement, pour un vecteur de réponses $y \in \mathbb{R}^n$ et une matrice de covariables $X \in \mathbb{R}^{n \times p}$, la régression Lasso consiste à résoudre le problème suivant :

$$\mathcal{L}(\beta) = \|y - X\beta\|^2 + \lambda\|\beta\|_1,$$

où $\lambda \geq 0$ est un paramètre de régularisation, et $\|\beta\|_1 = \sum_{j=1}^p |\beta_j|$ désigne la norme ℓ_1 du vecteur de coefficients. Plus λ est élevé, plus les coefficients sont poussés vers zéro.

Cette propriété de parcimonie rend le Lasso particulièrement adapté aux situations où il est souhaitable d'identifier un sous-ensemble pertinent de variables explicatives, ce qui peut renforcer l'interprétabilité du modèle et améliorer ses performances prédictives en contexte de bruit.

Dans cette étude, la régression Lasso est utilisée comme alternative robuste pour modéliser la relation entre variables explicatives (notamment macroéconomiques) et indices de prix immobiliers. Sa capacité à ignorer automatiquement les variables peu informatives permet d'éviter le sur-ajustement, tout en identifiant les déterminants les plus stables de la dynamique des prix. Cela est particulièrement pertinent dans un cadre de projection long terme, où la parcimonie et la stabilité structurelle sont recherchées.

Régression Ridge

La régression Ridge est une extension de la régression linéaire ordinaire (moindres carrés) conçue pour améliorer la robustesse des estimations en présence de multicolinéarité ou de données faiblement informatives. Elle introduit une pénalisation de type ℓ_2 sur la norme des coefficients afin de limiter les effets d'instabilité et de sur-apprentissage.

Formellement, pour un vecteur cible $y \in \mathbb{R}^n$ et une matrice de covariables $X \in \mathbb{R}^{n \times p}$, la régression Ridge cherche à minimiser la fonction suivante :

$$\mathcal{L}(\beta) = \|y - X\beta\|_2^2 + \alpha\|\beta\|_2^2,$$

où $\alpha \geq 0$ est un hyperparamètre contrôlant l'intensité de la régularisation ℓ_2 . Lorsque $\alpha = 0$, on retrouve le modèle linéaire classique ; lorsque $\alpha > 0$, la pénalisation introduit un biais contrôlé qui réduit la variance des estimations, au prix d'une perte potentielle d'ajustement. Cette approche est particulièrement adaptée dans le contexte de cette étude, et sera utilisée dans le but de créer un historique plus conséquent de manière à pouvoir simuler des scénarios de long terme. Les séries macroéconomiques explicatives sont potentiellement corrélées (par exemple, taux d'intérêt long et court), et le nombre d'années disponibles peut être limité, ce qui légitimise l'utilisation d'une telle méthode. La pénalisation ℓ_2 permet d'obtenir une estimation plus stable des relations entre les variables macroéconomiques et les indices de prix hédoniques.

Régression Elastic Net

La régression Elastic Net combine les avantages des régularisations Lasso (ℓ_1) en introduisant une double pénalisation dans la fonction de coût. Cette méthode a été conçue pour surmonter les limitations de chacune des approches prises isolément : la régression Ridge tend à conserver l'ensemble des variables explicatives (même celles peu informatives), tandis que le Lasso peut être instable en présence de variables fortement corrélées, n'en sélectionnant qu'une seule parmi plusieurs équivalentes.

La régression Elastic Net résout le problème d'optimisation suivant :

$$\mathcal{L}(\beta) = \|y - X\beta\|_2^2 + \lambda_1\|\beta\|_1 + \lambda_2\|\beta\|_2^2 \quad (1)$$

où λ_1 et λ_2 sont deux hyperparamètres contrôlant respectivement l'intensité des pénalisations ℓ_1 (effet de sélection) et ℓ_2 (effet de stabilisation). Ce compromis permet à la fois de sélectionner les variables pertinentes et de maintenir des estimations stables, même en cas de forte colinéarité ou de faible taille d'échantillon.

Dans le cadre de cette étude, l'Elastic Net constitue une solution particulièrement adaptée pour reconstruire des indices hédoniques immobiliers à un niveau fin de segmentation, croisant département et type de bien. Les avantages principaux de l'Elastic Net dans ce contexte sont les suivants :

- Robustesse face à la multicolinéarité : contrairement au Lasso, l'Elastic Net peut maintenir plusieurs variables corrélées dans le modèle, ce qui est crucial lorsque les effets croisés sont structurellement redondants.
- Stabilité des coefficients : l'introduction d'une composante Ridge assure une meilleure stabilité des coefficients d'année utilisés pour construire les indices hédoniques temporels.
- Sélection automatique des effets significatifs : les effets non informatifs (ex. années atypiques ou combinaisons spécifiques) peuvent être automatiquement ignorés, réduisant le sur-apprentissage, malgré un faible nombre d'observations.

Paramètres et validation du modèle

TABLE 10 – Paramètres et de l'Elastic Net

Dep	Type	R ² reprod.	RMSE reprod.	α	l_1	Inflation	Tx Court	Rdt immo	Tx Long	Pop. Nat.	Pop. loc.	R ² proj.	RMSE proj.
1	111	0,962	2,586	0,006	0,9	0,041	0	0	0	0,074	0	-234,001	5,613
1	121	0,901	2,054	0,005	0,7	0,028	0	0	0	0,034	0	-5,275	6,939
2	111	0,812	2,284	0,019	0,5	0,027	0	0	0	0	-0,012	-5,02	2,462
2	121	0,171	5,695	0,025	1	0,011	0	0	0	0	0	-6,202	3,818
3	111	0,95	1,025	0,004	0,1	0,034	0,018	0,013	0	0,045	0,033	-7,674	4,821
3	121	0,9	2,639	0,002	1	0,033	0	-0,009	0	0,073	0,029	-1,016	5,279
4	111	0,929	2,553	0,001	0,5	0,055	0,046	0	0	0,032	0,007	-209,415	3,362
4	121	0,831	3,297	0,017	0,1	0,045	0,043	0	0	0	0,027	-0,579	1,508
5	111	0,986	1,393	0,001	0,7	0,043	0,073	0	-0,036	0,077	-0,027	-7,054	3,648
5	121	0,952	2,126	0	0,9	0,059	0	-0,017	0,046	0,065	-0,014	-1030,314	11,222
6	111	0,985	1,189	0,003	1	0,034	0	-0,002	0	0	0,06	-2435,922	4,347
6	121	0,993	0,801	0	1	0,011	0,004	0	-0,006	0,002	0,073	-161,044	2,262
7	111	0,994	0,814	0,003	0,1	0,055	0,008	0,004	0,024	0	0,06	-104,61	2,019
7	121	0,866	2,543	0,008	0,1	0,036	0,021	0,001	0,007	0	0,035	-5,234	4,939
8	111	0,908	1,77	0,026	0,1	0,025	0	-0,004	0	0,029	0	-97,4	3,838
8	121	0,687	4,175	0,015	1	0,021	0	0	0	0,029	0	-9,073	6,035
9	111	0,96	1,993	0,004	0,7	0,025	0	-0,016	0	0	0,068	-30,766	7,12
9	121	0,825	3,819	0,014	0,5	0	0	-0,023	0	0	0,066	-1836,532	10,919
10	111	0,671	3,639	0,022	1	0,014	0	0	0	0,02	0	-177,49	3,436
10	121	0,833	3,94	0,018	1	0,039	0	0	0	0,032	0	-3,229	5,149
11	111	0,958	1,614	0,001	0,95	0,033	0,005	-0,005	0,001	0,047	0	-11,377	4,413
11	121	0,97	1,29	0	1	0,047	0	-0,012	0,025	0,177	-0,154	-27,556	1,702
12	111	0,955	1,627	0,005	0,1	0,042	0,039	-0,001	-0,016	0,029	0	-0,582	1,38
12	121	0,979	1,619	0,002	0,1	0,041	0,011	0,008	-0,001	0,066	0	-3,826	3,501
13	111	0,98	1,919	0,004	1	0,027	0	0	0	0,072	0,021	-107,045	1,388
13	121	0,906	2,527	0,011	0,1	0,046	0,023	0	0,002	0,013	0,024	-296,12	6,614
14	111	0,976	2,382	0,01	0,5	0,019	0	0	-0,013	0	0,108	-17,252	2,389
14	121	0,973	2,626	0,004	0,7	0,026	0	0,004	-0,006	0	0,113	-11,579	7,374
15	111	0,981	0,934	0,002	0,1	0,064	0,045	0,021	0,009	0,03	0,016	0,123	1,953
15	121	0,841	2,814	0,057	0,1	0,035	0	0	0	0,024	0	-7,24	4,868
16	111	0,94	2,695	0,045	0,1	0,037	0	-0,005	0	0,06	0	-993,339	5,303
16	121	0,884	6,185	0,018	1	0,021	0	0	0	0,117	0	-6,638	12,344
17	111	0,986	1,954	0,008	0,5	0,027	0	0	-0,011	0	0,108	-94,742	3,412
17	121	0,981	1,773	0,002	1	0,028	0	0	0	0	0,088	-2271,509	0,805
18	111	0,233	3,344	0,207	0,1	0,005	0	0	0	0,002	0	0,393	0,642
18	121	-0,001	6,377	0,356	0,1	0	0	0	0	0	0	-15,952	13,738
19	111	0,95	1,825	0,005	0,1	0,021	0	-0,017	-0,001	0,046	-0,015	-17,072	3,731
19	121	0,955	2,136	0,007	0,1	0,059	0	-0,01	0,037	0,045	-0,019	-234,721	9,55
21	111	0,95	1,813	0,007	0,1	0,045	0	0,011	0,024	0,047	0	-10,027	7,035
21	121	0,882	3,204	0,052	0,1	0,014	0	0,002	-0,018	0,037	0,015	-318,315	12,504
22	111	0,983	2,425	0,005	1	0,029	0	0	0	0	0,127	-30,475	4,673
22	121	0,996	1,284	0	1	0,03	-0,044	-0,018	0,045	0	0,154	-1,804	3,796
23	111	0,892	2,026	0,001	1	0,09	0,018	-0,001	0,052	0	0,01	-2,475	7,046
23	121	0,98	1,231	0	1	0,048	0,087	-0,029	-0,028	0	0,001	-9,736	3,351
24	111	0,987	1,247	0,001	0,1	-0,016	0,011	0,006	-0,051	0,065	0,048	-17,048	1,361
24	121	0,95	2,123	0,001	0,5	0,058	0,008	0	0,026	0,051	0,001	-3,652	5,834
25	111	0,924	2,527	0,009	0,5	0,038	0	0	0	0,013	0,032	-4,038	1,292
25	121	0,928	1,702	0,001	0,5	0,03	-0,015	0	0	0,024	0	-6,149	3,591
26	111	0,988	1,351	0,001	0,1	0,027	0,045	0,016	-0,055	0,177	-0,118	-29,218	3,131
26	121	0,987	1,007	0	1	0,017	0	0	-0,004	0,124	-0,064	-17,612	5,481
27	111	0,938	2,324	0,032	0,1	0,041	0	-0,004	0	0,045	-0,004	-52,635	5,076
27	121	0,542	4,5	0,025	1	0,013	0	0	0	0,017	0	-641,079	8,871
28	111	0,885	2,316	0,027	0,1	0,03	0	-0,008	0	0,033	0	-21,016	4,66
28	121	0,659	2,456	0,01	0,5	0,024	0	0	0	0	0,018	-1,21	2,055
29	111	0,979	2,888	0,003	0,9	0,039	0	-0,006	0	0	0,131	-82,832	3,734
29	121	0,965	4,47	0,011	1	0,026	0	0	0	0	0,159	-1510,84	7,953
30	111	0,976	1,721	0,003	0,7	0,029	0	-0,004	0	0,015	0,056	-5,074	2,049
30	121	0,931	2,085	0,005	0,99	0,017	0	0	0	0	0,054	-174,615	5,711
31	111	0,953	1,869	0,052	0,1	0,02	0	0	-0,015	0,021	0,024	-1,655	1,648
31	121	0,873	1,731	0,02	0,1	0,012	0	0,002	-0,009	0,014	0,012	-3,187	2,698
32	111	0,874	3,585	0,013	1	0,023	0	0	0	0	0,057	-109,647	11,056
32	121	0,886	3,797	0,001	1	0,024	0,024	0	0	0,036	0,033	-1,004	10,069
33	111	0,985	2,258	0,028	0,1	0,015	0	0,002	-0,034	0,056	0,058	-13,306	4,226
33	121	0,987	1,265	0,001	1	0	0	0,004	-0,034	0	0,074	-6,846	2,036
34	111	0,961	2,184	0,032	0,1	0,026	0	-0,006	0	0,038	0,034	-15,917	3,196
34	121	0,957	1,608	0,001	1	0,048	0,021	0	0,007	0,035	0	-6379,29	3,885
35	111	0,976	2,489	0,009	0,1	0,054	0,014	0	0	0,057	0,036	-459,778	6,592
35	121	0,945	3,64	0,013	0,99	0,029	0	0	0	0,092	0	-5,589	3,607
36	111	0,84	1,645	0,002	1	0,034	0	-0,017	0	0	0	-32,193	6,388
36	121	0,938	2,31	0,002	0,1	0,053	0,022	0	0,005	0,191	0,16	-1,686	4,725
37	111	0,971	1,549	0,007	0,5	0,029	0	0	-0,001	0,017	0,038	-17,584	2,16
37	121	0,931	2,616	0,011	0,1	0,046	0	-0,01	0	0	0,051	-15,93	6,933
38	111	0,966	2,521	0,004	0,5	0,038	0	-0,002	0	0	0,086	-13,53	2,807
38	121	0,796	3,31	0,014	1	0,024	0	0	0	0	0,028	-6,559	1,611
39	111	0,994	1,061	0,001	0,1	0,05	0,004	-0,002	0,028	0,052	-0,037	-175,139	6,94
39	121	0,772	3,957	0,025	1	0,005	0	0	0	0,042	0	-2,296	2,14

Dep	Type	R ² reproduction	RMSE	α	l_1	Inflation	Tx Court	Rdt immo	Tx Long	Pop. Nat.	Pop. loc.	R ² proj.	RMSE proj.
40	111	0,989	2,002	0,003	1	0,014	0	0	-0,025	0	0,129	-13,935	2,851
40	121	0,969	2,867	0,002	1	0,028	0	-0,009	0	0	0,109	-0,123	2,486
41	111	0,403	5,026	0,272	0,1	0,009	0	0	0	0,012	0	-167,014	11,182
41	121	0,92	3,375	0,003	0,1	0,087	0	0,008	0,058	0,137	0,1	-10,739	9,608
42	111	0,977	1,759	0,001	1	0,04	0	-0,003	0,007	0	0,072	-4454,216	4,736
42	121	0,601	6,011	0,038	0,9	0,028	0	0	0	0,019	0	-37,535	14,729
43	111	0,973	1,542	0,007	0,1	0,058	0,028	0	0,016	0,069	-0,026	-37,9	7,706
43	121	0,946	2,548	0	1	0,079	0,011	-0,033	0,02	0,093	-0,061	-95,825	6,396
44	111	0,959	3,287	0,006	1	0,041	0	0	0	0,073	0,024	-20,544	5,173
44	121	0,825	5,721	0,037	0,5	0,028	0	0	0	0,025	0,045	-15048,569	7,504
45	111	0,935	1,621	0,003	0,9	0,03	0	-0,003	0	0,032	0	-8,68	4,873
45	121	0,877	2,671	0,008	1	0,03	0	-0,007	0	0,034	0	-1,137	3,029
46	111	0,98	1,402	0	1	0,033	0	-0,012	0,003	0,009	0,056	-9,577	2,572
46	121	0,687	5,569	0,034	1	0	0	0	0	0	0,05	-0,926	8,535
47	111	0,954	2,029	0,002	1	0,031	0,001	0	0	0,062	0,012	-295,459	3,318
47	121	0,948	3,068	0,012	1	0,037	0	0	0	0,075	0	-1,291	3,954
48	111	0,903	3,043	0,036	0,1	0,032	0,014	0	0	0,057	-0,016	-26,396	2,234
48	121	0,682	5,461	0,011	1	0	0	-0,012	0	0,062	0	-292,521	7,63
49	111	0,949	3,614	0,014	0,5	0,036	0	-0,005	0	0,067	0,031	-1752,517	6,688
49	121	0,911	7,227	0,022	1	0,033	0	0	0	0,018	0,119	-1715,449	6,487
50	111	0,979	2,267	0,002	1	0,019	0	-0,004	-0,011	0,118	0,019	-6,575	2,275
50	121	0,966	2,4	0	0,7	0,001	-0,037	-0,011	0,011	0,138	0,055	-10,346	6,24
51	111	0,925	1,908	0,034	0,1	0,025	0	0	0	0,024	-0,016	-20,688	6,726
51	121	0,833	3,744	0,074	0,1	0,023	0	0,002	-0,018	0,024	-0,012	-493,947	5,952
52	111	0,954	1,113	0,001	1	0,054	0,001	-0,005	0,021	0,01	0	-20,264	3,846
52	121	0,832	3,01	0,002	0,7	0,056	0,031	-0,009	0,01	0,02	0	-6,418	10,081
53	111	0,983	1,253	0,001	1	0,027	0	-0,015	0	0	-0,065	-17,795	3,362
53	121	0,981	2,925	0,011	0,5	0,027	0	0,004	0	0,052	-0,092	-12,774	1,027
54	111	0,974	1,871	0,035	0,1	0,035	0	0	-0,005	0,061	-0,01	-123,923	2,946
54	121	0,949	1,897	0,035	0,1	0,012	0	0,002	-0,015	0,046	-0,018	-41,025	10,794
55	111	0,945	1,319	0	1	0,058	0,038	0,009	0,001	0,005	0	-19,289	2,466
55	121	0,834	3,72	0,008	1	0,008	0	0	0	0,047	-0,019	-11,292	11,3
56	111	0,985	2,455	0,002	0,5	0,054	0,006	-0,008	0,016	0	0,128	-21,172	6,576
56	121	0,958	3,95	0,003	0,5	0,068	0	-0,012	0,026	0	0,117	-11,105	10,353
58	111	0,842	2,033	0,002	1	0,044	0,013	-0,008	0	0,003	0	-2,623	2,428
58	121	0,511	7,702	0,027	1	0,029	0	-0,051	0	0	0	-0,641	5,457
59	111	0,935	2,039	0,035	0,1	0,022	0	0,001	-0,008	0,028	0,017	-25,961	1,408
59	121	0,907	1,686	0,02	0,1	0,019	0	0,003	-0,02	0,007	0,009	-185,531	4,059
60	111	0,96	1,418	0,016	0,1	0,019	0	0,002	-0,013	0,04	0	-4078,014	2,885
60	121	0,693	2,231	0,059	0,1	0,025	0	0	-0,002	0	0	-36,947	7,435
61	111	0,909	2,646	0,01	0,1	0,06	0	0	0,035	0,027	-0,017	-11,247	8,035
61	121	0,857	3,598	0,001	0,95	0,075	0,032	0,004	0,027	0,032	0	-3,872	7,139
62	111	0,961	1,564	0,025	0,1	0,032	0	-0,007	0	0,017	-0,028	-15,855	3,683
62	121	0,987	0,667	0,001	0,1	-0,02	-0,064	-0,048	0,008	0,009	-0,065	-16,731	2,102
63	111	0,973	1,219	0,01	0,1	0,032	0,011	0	0	0,042	0	-7,028	1,93
63	121	0,992	0,708	0	1	0	0,003	-0,001	-0,039	0,145	-0,097	-3,414	2,813
64	111	0,993	1,098	0,001	1	0,03	0	0	-0,003	0	0,088	-124,169	0,691
64	121	0,982	1,583	0,001	1	0,032	0	-0,003	0,001	0	0,079	-3,306	3,649
65	111	0,987	1,172	0,007	0,1	0,026	0,008	-0,01	0	0,028	0,044	-46356,124	5,705
65	121	0,954	1,939	0,001	0,5	0,049	0,001	0,001	0,048	0,014	0,042	-59,075	5,866
66	111	0,966	1,995	0,007	0,1	0,046	0,002	0	0,026	0	0,068	-3344,984	5,858
66	121	0,916	2,325	0,006	0,1	0,049	0	-0,014	0,028	0	0,042	-105,966	8,896
69	111	0,961	3,243	0,038	0,1	0,039	0	0,001	-0,018	0,057	0,032	-979,049	5,172
69	121	0,95	3,04	0,048	0,1	0,009	0	0,012	-0,031	0,042	0,037	-119,099	4,649
70	111	0,695	4,141	0,025	1	0,022	0	0	0	0,019	0	-114,584	6,539
70	121	0,896	2,362	0,001	1	0	-0,051	-0,056	0	0	-0,064	-14,329	7,629
71	111	0,928	2,171	0,001	1	0,04	0,015	-0,011	0	0,011	-0,031	-71,092	5,692
71	121	0,891	2,749	0,017	0,1	0,035	0,005	-0,016	0,003	0,044	0	-28,678	9,713
72	111	0,974	1,222	0,001	1	0,032	0	-0,007	0	0,061	0,027	-11,221	0,778
72	121	0,608	8,952	0,052	1	0,003	0	0	0	0,058	0	-164,639	8,171
73	111	0,99	1,802	0,001	0,7	0,045	0	-0,014	0	0	0,116	-29,021	4,308
73	121	0,996	0,845	0	1	0,037	0	-0,014	0,007	0	0,087	-15,336	1,002
74	111	0,973	2,436	0,001	0,5	0,034	0	-0,012	0	0,072	0,028	-113,02	5,502
74	121	0,972	1,692	0,005	1	0,012	0	0	0	0,075	0	-36,017	2,372
75	111	0,956	2,772	0,002	0,1	-0,09	-0,038	0,015	-0,054	0,225	0,106	-0,458	1,924
75	121	0,928	3,025	0,005	0,5	-0,001	0	0,032	-0,029	0,063	0	-9,251	1,372
76	111	0,952	1,962	0,005	0,7	0,026	0	0	0	0,05	0,009	-74,855	2,939
76	121	0,875	2,741	0,054	0,1	0,027	0	0	-0,009	0,029	0,007	-101,695	8,344
77	111	0,947	2,009	0,042	0,1	0,023	0	0	-0,018	0,025	0,019	-57,056	3,72
77	121	0,814	2,603	0,083	0,1	0,015	0	0	-0,015	0,011	0,008	-68,775	8,91
78	111	0,941	1,996	0,033	0,1	0,021	0	0	-0,013	0,039	0,006	-763,414	5,802
78	121	0,852	2,325	0,058	0,1	0,011	0	0,001	-0,016	0,024	0,003	-134,748	8,061
79	111	0,986	1,178	0,001	1	0,064	0,037	0	0	0,06	-0,021	-61,493	4,704
79	121	0,968	2,765	0,004	0,1	0,15	0,06	-0,002	0,064	0,139	-0,114	-11,874	20,455
80	111	0,957	1,304	0,02	0,1	0,021	0	0	0	0,012	-0,028	-213,105	4,916
80	121	0,619	3,213	0,088	0,1	0,009	0	0	0	0	-0,025	-96,206	8,724
81	111	0,927	2,586	0,008	1	0,024	0	0	0	0	0,058	-412,749	2,937
81	121	0,975	1,758	0,002	1	0,022	0	-0,004	-0,002	0	0,075	0,4	1,94
82	111	0,923	3,043	0,009	1	0,035	0	0	0	0,055	0	-8,996	4,215
82	121	0,968	2,815	0,006	0,1	0,058	0,007	0,001	0,026	0,099	0	-18,307	10,829
83	111	0,977	1,91	0,001	0,7	0,043	0	-0,02	0,009	0	0,08	-96,255	6,364

Dep	Type	R ² reproduction	RMSE	α	l_1	Inflation	Tx Court	Rdt immo	Tx Long	Pop. Nat.	Pop. loc.	R ² proj.	RMSE proj.
83	121	0,955	1,726	0,001	0,5	0,045	0	-0,012	0,024	0	0,046	-101,499	5,122
84	111	0,852	4,486	0,06	0,1	0,033	0	-0,001	0	0,029	0,034	-11,897	3,83
84	121	-0,001	5,134	0,336	0,1	0	0	0	0	0	0	-11,16	9,109
85	111	0,982	2,516	0,006	0,7	0,04	0	-0,003	0	0	0,118	-52,327	3,227
85	121	0,975	2,391	0	1	0,084	0,019	0	0,032	-0,182	0,258	-12,334	4,233
86	111	0,951	1,345	0,002	0,7	0,035	0,01	-0,005	-0,002	0,026	0	-4329,371	1,901
86	121	0,868	5,08	0,019	1	0,022	0	0	0	0,076	0	-1,47	4,691
87	111	0,943	1,707	0,002	0,9	0,04	0,008	0	0	0,031	0	-105,694	5,917
87	121	0,926	3,53	0,003	1	0,043	0	0	0	0,072	0	-110,547	2,047
88	111	0,898	3,124	0,01	1	0,018	0	-0,001	0	0,061	0	-46,298	6,604
88	121	0,936	2,556	0,001	1	0,04	0,009	-0,001	0,01	0,062	0	-1247,036	8,354
89	111	0,844	1,742	0,001	1	0,045	0,009	-0,005	0,004	0	0	-3,34	3,784
89	121	-0,001	5,183	0,323	0,1	0	0	0	0	0	0	0,195	0,983
90	111	0,881	2,298	0,048	0,1	0,02	0	0	-0,01	0,014	-0,015	-2,106	1,231
90	121	0,648	3,22	0,012	1	0,019	0	0,006	0	0	-0,01	-18,652	7,726
91	111	0,946	2,015	0,035	0,1	0,013	0	0	-0,027	0,044	0,001	-2550,003	3,853
91	121	0,656	3,493	0,124	0,1	0,02	0	0	-0,01	0,007	0	-26,876	6,686
92	111	0,907	3,633	0,069	0,1	0	0	0,013	-0,039	0,032	0,033	-52,471	9,415
92	121	0,923	3,001	0,064	0,1	0	0	0,015	-0,033	0,028	0,031	-211,641	4,69
93	111	0,916	3,257	0,054	0,1	0,009	0	0,005	-0,025	0,044	0,023	-741,84	6,974
93	121	0,885	3,486	0,032	0,1	0,006	0	0,012	-0,035	0,044	0,01	-79,174	14,861
94	111	0,95	2,875	0,047	0,1	0,015	0	0	-0,029	0,042	0,033	-6103,365	2,803
94	121	0,918	2,735	0,062	0,1	0,008	0	0,008	-0,02	0,022	0,032	-177,794	5,352
95	111	0,948	2,238	0,026	0,1	0,014	0	0,008	-0,026	0,039	0,015	-26,446	3,353
95	121	0,789	2,946	0,085	0,1	0,011	0	0	-0,019	0,013	0,009	-1691,228	10,667

8 Réseaux de neurones

Notations des réseaux de neurones

Dimensions fondamentales :

- n : dimension de l'entrée (nombre de variables explicatives),
- m : nombre d'observations dans le jeu d'entraînement,
- L : nombre total de couches du réseau (hors couche d'entrée),
- t_ℓ : nombre de neurones dans la couche ℓ ($\ell = 1, \dots, L$).

Données d'entrée et sortie :

- $X \in \mathbb{R}^{m \times n}$: matrice des observations en entrée (chaque ligne est une observation),
- $Y \in \mathbb{R}^{m \times 1}$ (ou $\mathbb{R}^{m \times q}$ en sortie multivariée) : cibles à prédire,
- \hat{Y} : prédictions du modèle.

Paramètres du réseau :

- $W^{[\ell]} \in \mathbb{R}^{t_{\ell-1} \times t_\ell}$: matrice des poids reliant la couche $\ell - 1$ à la couche ℓ ,
- $b^{[\ell]} \in \mathbb{R}^{t_\ell}$: vecteur des biais associés à la couche ℓ ,
- θ : ensemble des paramètres du réseau, soit $\theta = \{W^{[\ell]}, b^{[\ell]}\}_{\ell=1}^L$,
- $\varphi(\cdot)$: fonction d'activation.

Variables intermédiaires (réseaux *feed-forward*) :

- $Z^{[\ell]} = A^{[\ell-1]} \cdot W^{[\ell]} + b^{[\ell]}$: pré-activation de la couche ℓ ,
- $A^{[\ell]} = \varphi^{[\ell]}(Z^{[\ell]})$: activation de la couche ℓ , avec $\varphi^{[\ell]}$ la fonction d'activation,
- $A^{[0]} = X$: convention pour l'entrée.

Fonction de perte et optimisation :

- $\mathcal{L}(y, \hat{y})$: fonction de perte élémentaire,
- $J(\theta) = \frac{1}{m} \sum_{i=1}^m \mathcal{L}(y^{(i)}, \hat{y}^{(i)})$: fonction objectif.

Hyperparamètres :

- α : taux d'apprentissage,
- λ : coefficient de régularisation (L1 ou L2),
- d_{emb} : dimension d'un *embedding* (si variable catégorielle),
- t_ℓ : nombre de neurones dans la couche ℓ ,
- p : taux de *dropout* (probabilité de désactivation d'un neurone à l'entraînement),
- B : taille des *mini-batches*,
- E : nombre d'*epochs*.

Réseaux récurrents :

- x_t : entrée à l'instant t ,
- h_t : état caché à l'instant t ,
- W_h, W_x, W_y : matrices de poids respectives pour h_{t-1} , x_t et la sortie,
- b, b' : biais associés,
- $\phi(\cdot)$: fonction d'activation de l'état caché,
- \hat{y}_t : sortie prédite à l'instant t .

Unités LSTM :

- i_t, f_t, o_t : portes d'entrée, d'oubli et de sortie,
- \tilde{c}_t : candidat mémoire,
- c_t : état mémoire à l'instant t .

Problèmes de gradient dans les RNN standards

L'apprentissage des réseaux de neurones récurrents (RNN) au moyen de la rétropropagation du gradient à travers le temps est systématiquement confronté à deux écueils : *vanishing* et *exploding*¹.

Lorsqu'on propage l'erreur sur de longues séquences, la suite de produits de dérivées peut décroître exponentiellement (si la norme des poids est < 1) ou, à l'inverse, croître sans limite (si elle est > 1), rendant l'optimisation instable ou inefficace. Plus rigoureusement, la dérivée du coût par rapport à un poids donné s'écrit comme un produit itératif des jacobiniennes des fonctions d'activation successives ; la norme du gradient décroît ou explose de façon exponentielle avec la longueur T de la séquence. Ce phénomène s'avère particulièrement handicapant pour les tâches nécessitant la prise en compte de dépendances temporelles étendues. Mathématiquement, considérons un RNN simple dont l'état caché se met à jour par

$$h_t = \phi(W_h h_{t-1} + W_x x_t + b),$$

et dont la fonction de perte instantanée est \mathcal{L}_t . En rétropropagation à travers le temps, le gradient d'un poids W_h à l'instant $t-\tau$ se réécrit² :

$$\frac{\partial \mathcal{L}_t}{\partial W_h} = \underbrace{\left(\prod_{k=1}^{\tau} \frac{\partial h_{t-k+1}}{\partial h_{t-k}} \right)}_{\mathbf{J}^{(\tau)}} \frac{\partial h_{t-\tau}}{\partial W_h},$$

1. [R. Pascanu et al., 2013].

2. [P. Ottou, 2017].

où $\frac{\partial h_{t-k+1}}{\partial h_{t-k}} = \text{diag}(\phi'(z_{t-k+1})) W_h^\top$. La matrice produit $\mathbf{J}^{(\tau)}$ transporte l'erreur sur τ pas de temps. Si λ_{\max} (resp. λ_{\min}) est la plus grande (resp. la plus petite) valeur singulière de W_h , on obtient l'inégalité classique

$$\|\mathbf{J}^{(\tau)}\|_2 \leq (\lambda_{\max} \|\phi'\|_\infty)^\tau, \quad \|\mathbf{J}^{(\tau)}\|_2 \geq (\lambda_{\min} \|\phi'\|_\infty)^\tau.$$

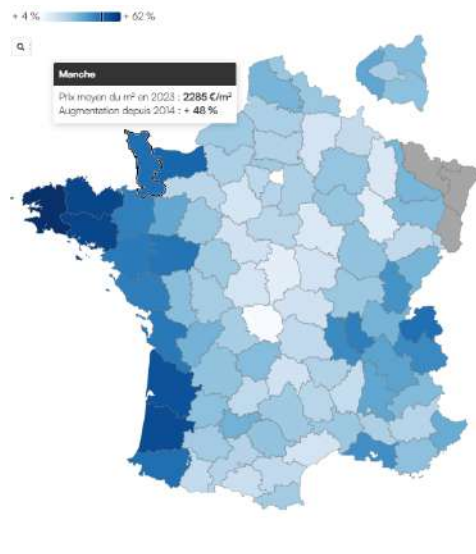
- *Vanishing gradients* : si $\lambda_{\max} \|\phi'\|_\infty < 1$, la norme décroît exponentiellement et l'apprentissage « oublie » les contributions lointaines.
- *Exploding gradients* : si $\lambda_{\min} \|\phi'\|_\infty > 1$, la norme explose, entraînant des oscillations numériquement ingérables.

Empiriquement, même de très légers écarts autour de 1 suffisent à faire tomber la norme à un niveau très faible ou à la faire exploser. Les tâches où l'information pertinente se situe loin dans le passé deviennent quasiment impossibles à entraîner avec un RNN naïf.

.9 Analyses - Modélisation segmentée

Evolution du prix en Bretagne

FIGURE 5 – Évolution du prix moyen d'achat au m² entre 2014 et 2023



Source : Données DVF

Impact du type de bien sur la LTV

Les cartes segmentées selon le type de bien présentent des tendances globalement similaires, bien que des nuances apparaissent.

Pour les appartements, trois grands *clusters* se distinguent nettement : la façade Atlantique Ouest, les Alpes et la Bretagne, avec également un niveau particulièrement élevé à Paris. Pour la Méthode 2, les LTV propres à ces *clusters* apparaissent plus hétérogènes. En revanche, la carte des maisons révèle une tendance géographique plus vague. La Bretagne ressort à nouveau comme une zone à forte LTV, mais les dynamiques sont moins lissées spatialement, traduisant une variabilité plus marquée selon les départements.

FIGURE 6 – Carte des LTV optimales par type de bien - Méthode 1 GSEs

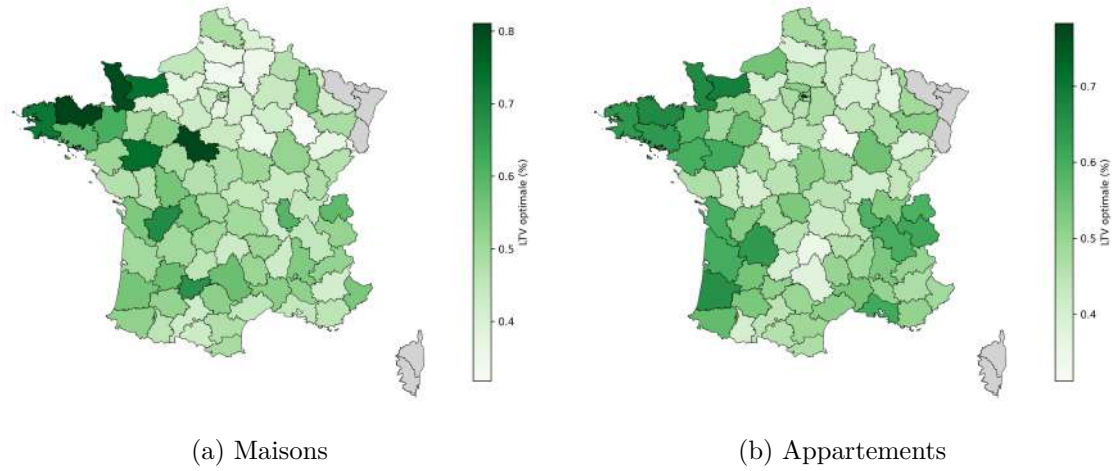


FIGURE 7 – Carte des LTV optimales par type de bien - Méthode 2 LSTM

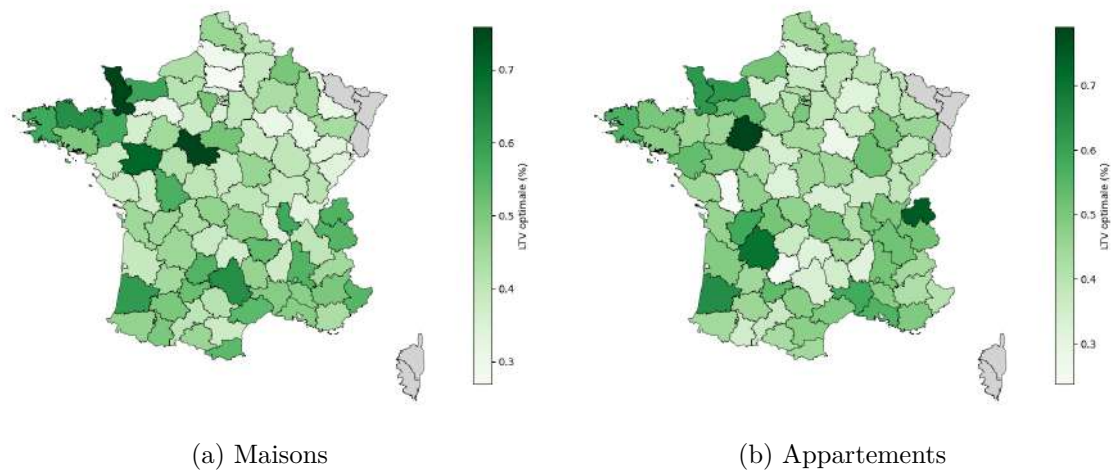
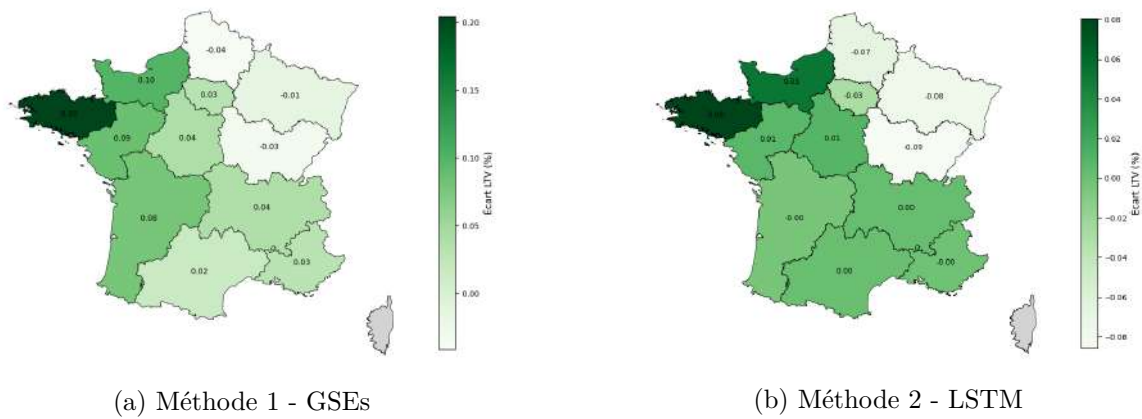


FIGURE 8 – Cartes des écarts de LTV par région

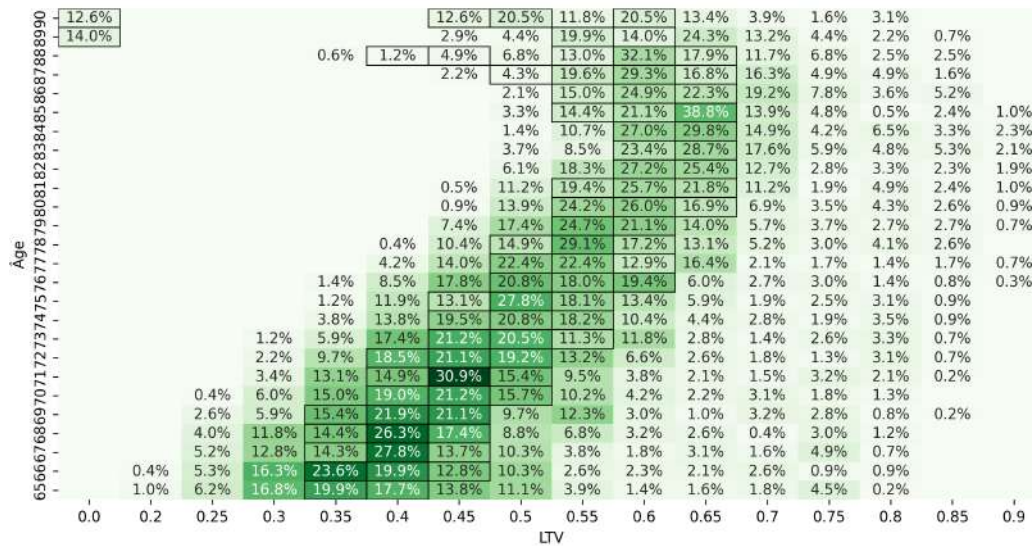


Tranches d'âges

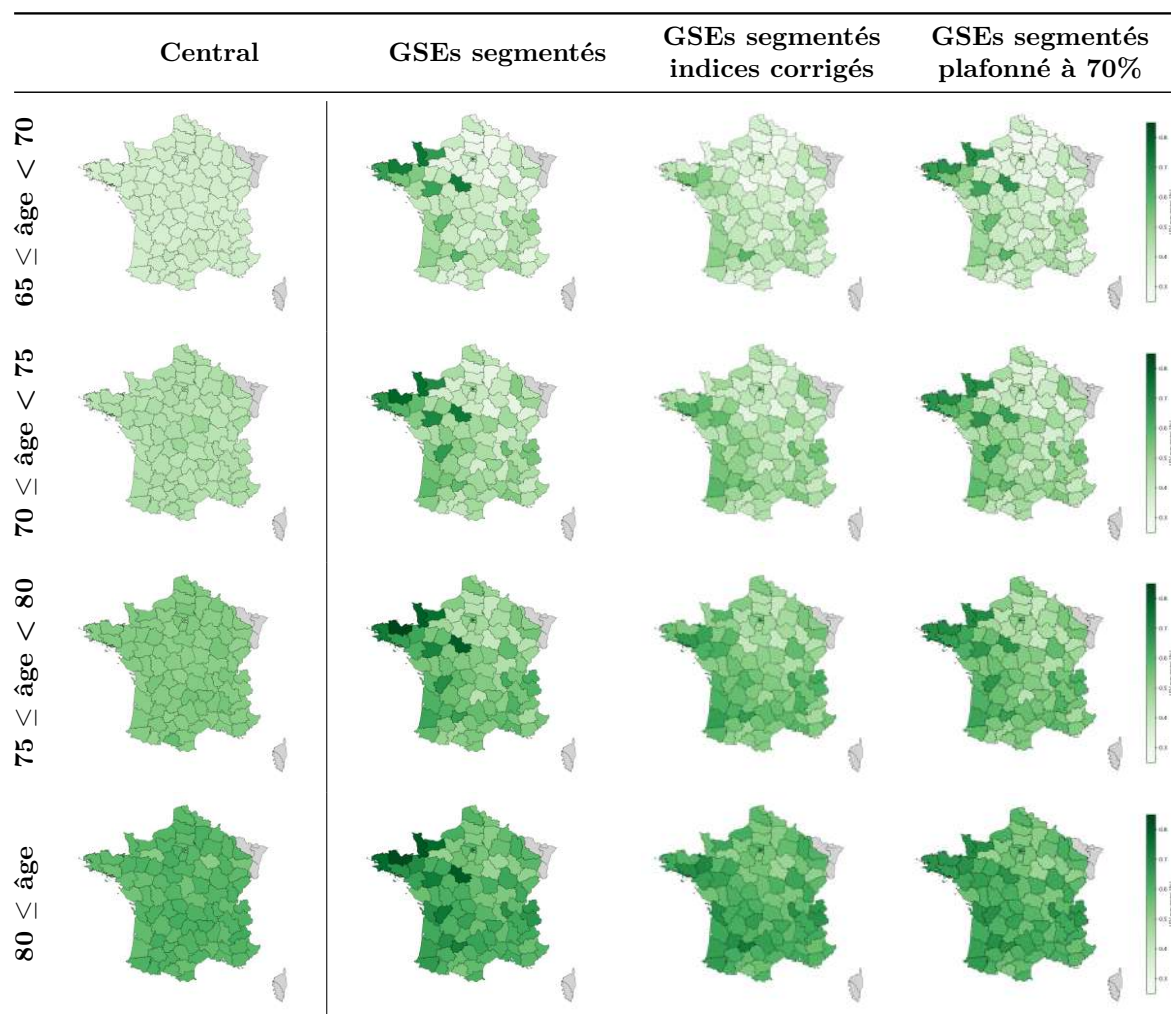
TABLE 11 – Indicateurs par tranche d'âge et par méthode

	Central				Méthode 1 – GSEs				Méthode 2 – LSTM			
	Global	70–	70–80	80+	Global	70–	70–80	80+	Global	70–	70–80	80+
Rendement <i>flat</i> (%)	9,8	13,2	9,8	7,0	9,9	13,2	9,9	7,0	9,8	13,9	10,2	7,0
Rendement capitalisé (%)	5,3	5,4	5,3	5,1	5,3	5,4	5,3	5,2	5,4	5,6	5,4	5,2
Prêt moyen (€)	88 329	70 341	89 341	107 962	97 223	81 729	97 922	114 442	87 171	58 125	86 468	123 022
Rendement <i>flat</i> p5 (%)	4,4	6,0	4,3	3,2	4,8	6,4	4,8	3,5	4,9	6,7	5,0	3,8
Probabilité NNEG (%)	35,0	36,5	37,5	30,0	38,0	40,3	40,3	32,6	28,4	22,4	28,3	31,7
LTV moyenne (%)	47,2	37,3	47,5	58,9	52,0	43,3	52,1	62,5	46,6	30,8	46,0	67,1
C. au rdt <i>flat</i> (%)	13,2	–	–	52,4	15,2	–	0,2	59,8	13,9	–	0,0	55,2
C. au rdt <i>flat</i> p5 (%)	6,6	–	1,3	24,0	1,3	–	0,3	4,7	–	–	–	–
C. à la proba. NNEG (%)	46,2	50,8	57,3	21,1	80,1	95,5	96,0	33,6	5,9	–	1,2	21,2
C. à la valeur NNEG (%)	33,5	49,2	41,4	0,8	2,9	4,5	3,4	0,2	79,8	100,0	98,8	21,9
Part refusée (%)	0,4	–	–	1,7	0,4	–	–	1,7	0,4	–	–	1,7

FIGURE 9 – Matrice de fréquence - LTV optimale GSEs × Age



Correction des LTV jugées trop forte - Méthode 1 GSEs

FIGURE 10 – LTV moyenne par département et tranche d'âge – Central *vs.* Segmenté

Tarification prudente

TABLE 12 – Statistiques agrégées par tranche d'âge — Tarification prudente

	Prudent - global	70-	70-80	80+
Rendement <i>flat</i>	9,9 %	13,8 %	10,1 %	7,1 %
Rendement capitalisé	5,4 %	5,6 %	5,4 %	5,2 %
Prêt moyen (€)	81 831	56 662	82 822	110 058
Rendement <i>flat</i> au 5 ^{ème} percentile	4,9 %	6,7 %	4,9 %	3,6 %
Probabilité NNEG	28,5 %	23,4 %	29,8 %	29,8 %
LTV moyenne	43,7 %	30,0 %	44,1 %	60,1 %
Prop. du ptf. contrainte par le rdt. <i>flat</i>	13,7 %	–	0,1 %	54,1 %
Prop. du ptf. contrainte par le rdt. <i>flat</i> p5	1,3 %	–	0,3 %	4,7 %
Prop. du ptf. contrainte par la proba. NNEG	12,9 %	3,6 %	14,8 %	20,8 %
Prop. du ptf. contrainte par la valeur NNEG	70,4 %	96,4 %	84,6 %	14,1 %
Prop. du ptf. plafonnée à 70%	1,2 %	–	0,2 %	4,6 %
Prop. du ptf. refusée	0,4 %	–	–	1,7 %

TABLE 13 – Statistiques agrégées par zone géographique — Tarification prudente

	PM	AM	AP	Ih	P
Rendement <i>flat</i>	9,8 %	10,0 %	9,7 %	9,9 %	10,2 %
Rendement capitalisé	5,4 %	5,4 %	5,4 %	5,4 %	5,5 %
Prêt moyen (€)	80 146	70 712	83 380	99 746	152 191
Rendement <i>flat</i> au 5 ^{ème} percentile	4,8 %	5,0 %	4,8 %	4,9 %	4,9 %
Probabilité NNEG	28,6 %	28,8 %	28,5 %	28,0 %	24,3 %
LTV moyenne	42,9 %	44,9 %	47,5 %	40,9 %	59,7 %
Prop. du ptf. contrainte par le rdt. <i>flat</i>	14,3 %	13,0 %	13,9 %	12,5 %	1,3 %
Prop. du ptf. contrainte par le rdt. <i>flat</i> p5	2,0 %	0,7 %	0,3 %	–	–
Prop. du ptf. contrainte par la proba. NNEG	13,6 %	11,8 %	11,8 %	13,5 %	–
Prop. du ptf. contrainte par la valeur NNEG	68,7 %	73,1 %	71,9 %	73,8 %	68,0 %
Prop. du ptf. plafonnée à 70%	0,9 %	1,1 %	1,8 %	–	30,7 %
Prop. du ptf. refusée	0,5 %	0,4 %	0,3 %	0,2 %	–

FIGURE 11 – Matrice de fréquence LTV central - LTV prudente

