

**Proposition d'une nouvelle modélisation de l'actif
immobilier pour les générateurs de scénarios
économiques**

Mémoire d'actuariat présenté pour l'obtention du

**Master droit économie et gestion
Mention actuariat**

Et l'admission à l'Institut des Actuares

Mémoire soutenu le 16 juin 2017

par M. Sébastien GALLET

Caractère confidentiel : non

Jury :

Président : Michel FROMENTEAU

Membres : Edith BOCQUAIRE

Gwenaël BILLIOTTE

Jean-Claude RAZAFINDRAKOTO

Pierre PETAUTON

David FAURE

François WEISS

Remerciements

Je souhaite adresser mes remerciements à Jean-Claude, Eric et Jacques pour leurs conseils avisés.

Je remercie également les professeurs du CNAM qui m'ont permis de m'enrichir intellectuellement et d'évoluer dans ma vie professionnelle.

Je souhaite enfin exprimer toute ma gratitude à Karen qui m'a toujours soutenu, voire supporté, et sans qui cela n'aurait sans doute pas été possible.

Table des matières

Remerciements	2
Table des matières	3
Introduction	4
I ETAT DES LIEUX	6
I.1 GENERATEURS DE SCENARIOS ECONOMIQUES : POURQUOI ? COMMENT ?	6
I.1.1 Pourquoi les GSE ?	6
I.1.2 Comment fonctionne un GSE ?	6
I.2 MARCHE IMMOBILIER D'HABITATION FRANÇAIS	8
I.2.1 Données utilisées	8
I.2.2 Analyse des données	11
II DEVELOPPEMENT DU MODELE DU MARCHE IMMOBILIER.....	13
II.1 PRESENTATION DES HYPOTHESES ET DEVELOPPEMENTS.....	13
II.1.1 Normalité vs. log normalité	13
II.1.2 Invariance de la distribution des groupes par passage du marché local au global.....	15
II.1.3 Acheteur/vendeur et prix moyen: modélisation stochastique	16
II.2 ANALYSE SIMPLE DU MODELE	28
II.2.1 Analyse de l'équation différentielle.....	28
II.2.2 Choc, retour à l'équilibre et application aux bulles spéculatives	29
III APPLICATION AU MARCHE D'HABITATION FRANÇAIS DE 1970-2015.....	32
III.1 DRIFT DES ACHETEURS = POUVOIR D'ACHAT IMMOBILIER	32
III.1.1 Evolution du pouvoir d'achat immobilier.....	32
III.1.2 Application au modèle et résultat intermédiaire	34
III.2 ESTIMATION DES VOLATILITES ACHETEUR/VENDEUR	35
III.2.1 Importance des crises	36
III.2.2 Principe d'évolution des volatilités Acheteur/Vendeur	37
III.2.3 Calibrage des crises de volatilités Acheteurs/Vendeurs	38
III.3 « EFFET TUNNEL » ET PREVISION A COURT TERME	43
III.3.1 « Effet tunnel »	43
III.3.2 Essai de prévision à court terme	45
IV GSE (GENERATEUR DE SCENARIOS ECONOMIQUES)	46
IV.1 APPLICATION DU MODELE DYNAMIQUE STOCHASTIQUE DE L'IMMOBILIER A LA FORMATION D'UN GSE SIMPLIFIE	46
IV.1.1 Principe général et simplification	46
IV.1.2 Tirages et résultats intermédiaires	46
IV.2 COMPARAISON DU MODELE IMMOBILIER A CELUI UTILISE DANS UN GSE TYPE AHLGRIM	48
IV.2.1 Intégration du modèle dynamique stochastique pour l'immobilier	48
IV.2.2 Backtesting	51
Conclusion	55
Bibliographie	59
Table des annexes	61
Liste des figures	114

Introduction

Dans le cadre de leur pilotage technique, les assureurs se doivent de mettre en œuvre une gestion de leur actif basée sur la projection des flux réaliste mais aussi conservative. Les différentes contraintes réglementaires successives qu'ont été les modèles MCEV puis les techniques de calcul des provisions dans le cadre IFRS et enfin Solvabilité II les poussent à recourir à des modèles stochastiques prenant en compte la volatilité des performances des actifs. Ces modèles de projections d'actifs doivent ainsi respecter des contraintes :

- De non sous estimation des rendements défavorables
- Intégrer les contraintes économiques long-terme (plus de 30 ans)
- Intégrer l'inflation et les dépendances entre les actifs

Les modèles répondant à cette problématique et basés sur la génération de scénario cohérent sont appelés des Générateurs de Scénarios Economiques (GSE). Ils comportent classiquement quatre grandes classes d'actif que sont les obligations, les actions, le monétaire et l'immobilier. Ces modèles sont en constantes évolutions et bénéficient des avancées théoriques sur les différentes classes d'actifs. On peut citer ici les modèles de GSE de Wilkie [1986], Brennan et Xia [2000] et plus récemment le modèle d'Ahlgrim [2005] sur lequel je reviendrai par la suite. Néanmoins tous ces modèles, lorsqu'ils modélisent l'immobilier, le modélise comme les autres actifs financiers à savoir comme une action ou éventuellement comme l'inflation. Hors, ces modèles n'intègrent pas les particularités du marché immobilier.

L'immobilier est un actif très différent des autres actifs du marché financier. Il est peu liquide et les biens sont peu divisibles et peu substituables. Substituable voulant dire que leurs détenteurs peuvent les remplacer par un autre actif s'ils jugent le couple rendement/risque plus avantageux. L'arbitrage, notion centrale de l'évaluation d'actif est ici difficile à définir. On ne peut pas réduire une décision d'achat pour habitation propre à un calcul de rendement suivi d'un arbitrage. Enfin, l'évolution des prix varie lentement par rapport aux marchés financiers et il y a de toute évidence une mémoire des prix. C'est donc un marché fortement inefficent au sens de Fama¹.

L'évaluation de la plupart des actifs financiers est basée sur l'utilisation de modèles stochastiques adaptés aux particularités de chaque marché. On peut distinguer par exemple les options simples communément modélisées par Black-Scholes (type action) ou encore les obligations (à travers l'inflation) modélisées par un modèle de Vasicek. Ils sont basés sur une modélisation par variables aléatoires complétée par une notion d'arbitrage venant de la substituabilité des biens échangés. En fait, tous les actifs substituables sont ainsi évalués par rapport au taux d'emprunt sans risque. Aucun modèle de référence ne prend en compte spécifiquement les particularités d'un marché de biens non substituables et en conséquence il n'existe pas, à ma connaissance, de modèle de marché spécifique à l'immobilier.

Un modèle de propagation des prix de l'immobilier devrait inclure ces particularités ainsi que d'autre issue de l'analyse multi variée qui va être présentée dans la première partie. Ces particularités donnent donc un sens voir une certaine urgence à l'établissement d'un modèle spécifique d'évolution des prix sur le marché de l'immobilier qui puisse être intégré

¹ Eugène Fama, « Efficient Capital Works », 1970 Review of Theoretical and Empirical Works.

dans un générateur de scénarios économiques (GSE). Cela sera toute l'ambition de mon mémoire.

Mais tout d'abord et pour comprendre le cheminement qui m'a mené au modèle d'évolution des prix de l'immobilier, je commencerai par un état des lieux sur l'utilisation des GSE et la présentation des données du marché immobilier français.

I Etat des lieux

Il faut dans un premier temps revenir sur le deux sujets qui constituent le fondement de cette étude à savoir les générateurs de scénarios économiques (GSE) et le marché immobilier.

I.1 Générateurs de Scénarios Economiques : Pourquoi ? Comment ?

L'objet de cette partie est de revenir sur la notion de scénario économique pour en définir l'utilisation ainsi que son fonctionnement.

I.1.1 Pourquoi les GSE ?

Un scénario économique correspond à une projection de grandeurs économiques et financières sur un horizon de temps. Si la génération de scénarios économiques est aujourd'hui une problématique incontournable pour les organismes assureurs c'est principalement du fait de la consécration des méthodes de valorisation financière dans l'évaluation des engagements d'assurance. En effet, les méthodes de valorisation dites market-consistent sont devenues des standards tant dans les problématiques de communication financière (Market Consistent Embedded Value) que prudentielles avec Solvabilité II. Par exemple, dans le cadre de Solvabilité II, la génération de scénarios économiques intervient au moins à deux moments précis dans l'évaluation de la solvabilité globale d'un organisme assureur :

- Le calcul des provisions techniques relatives aux engagements d'assurance contenant des options et garanties financières
- L'estimation du capital de solvabilité par le biais d'un modèle interne

Enfin les GSE sont également devenus incontournables en matière de gestion d'organisme d'assurance et plus particulièrement dans la recherche d'allocations stratégiques d'actifs

I.1.2 Comment fonctionne un GSE ?

Un GSE consiste à définir un ensemble de modèles stochastiques (une structure) représentant les différents actifs ainsi que leurs corrélations. Soit parce que l'on injecte directement la corrélation dans la génération des nombres aléatoires (résidus), soit parce que les formules stochastiques des modèles font appel à une même variable sous-jacente.

Je ne présenterai ici que le modèle d'Ahlgrim qui est l'un des modèles de référence. En effet, afin d'accentuer la sensibilisation des actuaires à la gestion des risques financiers, la Casualty Actuarial Society (CAS) et la Society Of Actuaries (SOA), deux associations professionnelles reconnues, se sont associées pour solliciter des recherches sur le développement de modèles intégrés. Le modèle d'Ahlgrim s'inscrit dans ce cadre, et couvre, entre autres, les taux d'intérêt, les rendements des actions, les taux de dividende, les taux d'inflation et les revenus de l'immobilier. Il peut être résumé par le schéma suivant :

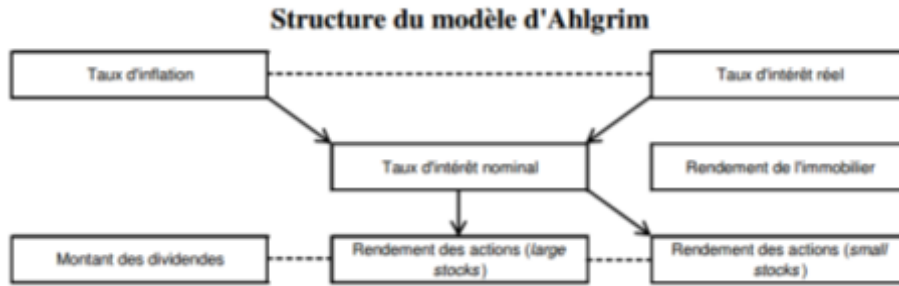


Figure 1 :Structure du modèle d' Ahlgrim

Chaque actif est modélisé par une formule mathématique intégrant une partie aléatoire (Gaussienne) que l'on appelle résidu et qui est, dans le cas du modèle d'Ahlgrim, corrélé aux autres actifs. Le détail des formules est présenté en annexe 1. Ces formules, comportant systématiquement plusieurs composantes aléatoires Gaussiennes, nécessitent de « tirer » des scénarios pour en avoir une représentation utilisable. Cela revient à générer des nombres aléatoires respectant les corrélations et ainsi calculer des scénarios respectant la cohérence entre les actifs modélisés (immobilier, action, obligation...). On génère dans ce cas des scénarios qui prennent la forme d'une matrice de trois dimensions :

- Les actifs et variables économiques
- Le temps
- Les scénarios (aussi appelés trajectoires)

Ce qui peut être visualisé de la façon suivante :

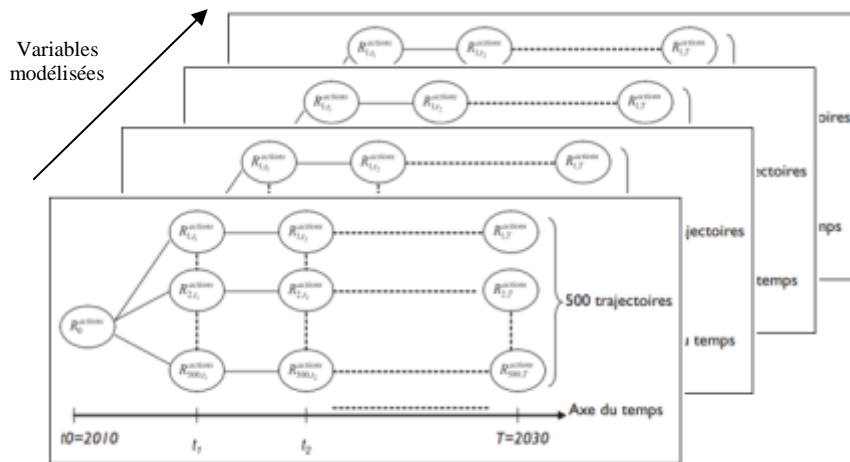


Figure 2 :Schéma du cube des tirages du GSE

Si la génération de scénarios économiques est aujourd'hui industrialisée dans beaucoup d'organismes d'assurance vie, notamment chez ceux qui publient des Embedded Value, il n'en demeure pas moins que la crise financière de 2008 doit amener les générateurs à évoluer. La notion de corrélation se retrouve souvent au centre des critiques. En effet, en cas de crise, les corrélations évoluent rapidement et rendent caduc les paramètres des modèles financiers. Il faut donc renforcer la compréhension des mécanismes sous-jacents à chaque actif et l'immobilier est l'actif dont le modèle couramment utilisé est notoirement le moins réaliste. En effet, dans le modèle d'Ahlgrim, l'immobilier est modélisé par un processus d'Ornstein Uhlenbeck qui s'écrit mathématiquement de la façon suivante :

$$dq_t = k_q(\mu_q - q_t)dt + \sigma_q dB_{q,t}$$

- k_q est la vitesse de retour à la moyenne.

- μ_q est le taux moyen à long terme.

- σ_q est l'écart type de l'erreur.

- $B_{q,t}$ est un mouvement brownien.

C'est une marche aléatoire de volatilité σ avec une force de retour à une moyenne imposée (μ le taux moyen long terme). Cette force de rappel est modulée par la variable k qui peut être vu comme la vitesse de retour à la moyenne imposée μ .

Cela se traduit par le fait que l'évolution de la variable modélisée dépend de trois paramètres et une variable aléatoire. La vitesse de retour à la moyenne et l'écart type sont estimés sur les données historiques et le taux moyen long terme doit être cohérent des projections économiques globales à long terme. La variable aléatoire est corrélée aux autres actifs et cette corrélation est calculée sur un historique difficile à définir en terme de durée car en période de crise, les corrélations changent violemment et mettent à mal les modèles calibrés sur des périodes de « calme économique ». En conclusion, ce modèle est simpliste et facilement utilisable mais il ne rend pas compte de la réalité et des spécificités du marché immobilier que je vais vous présenter dans la suite.

I.2 Marché immobilier d'habitation Français

Ce mémoire ne portera que sur l'étude et la modélisation du marché d'habitation (et non du marché des locaux d'entreprise). C'est en effet un choix de simplification. L'immobilier d'entreprise pourrait ultérieurement être intégré à cette étude mais ne devrait pas sensiblement en changer les résultats théoriques.

Le marché de l'immobilier d'habitation est fondamentalement différent des autres marchés. C'est un marché de bien peu substituable (uniquement par le marché de la location qui est lui-même lié au marché de la vente) avec un niveau d'information très bas. Seuls quelques paramètres sur le bien suffisent à en évaluer un prix d'ailleurs très fluctuant selon les estimations des différents acteurs. L'objectif va être dans cette étude de définir un modèle d'évolution des prix immobilier basé sur d'autre variable, plus prédictibles et/ou peu volatiles permettant une explication de l'évolution passée et une prédiction des évolutions futures. Pour cela il faut avant tout définir les variables d'intérêts et ce travail débute par la constitution d'une base de données sur un ensemble de variables économiques le plus large possible et sur une période de temps suffisamment longue pour pouvoir calibrer et estimer la qualité d'un modèle.

I.2.1 Données utilisées

Les données qui servent de base à cette étude sont celles que Jacques Friggit² a rassemblées pour ses activités au sein du Conseil Général de l'environnement et du développement durable CGEDD³. Il a produit de nombreuses études sur le l'évolution des prix de

² ingénieur général au [Conseil général de l'environnement et du développement durable](http://www.cgedd.developpement-durable.gouv.fr/)

³ <http://www.cgedd.developpement-durable.gouv.fr/>

l'immobilier⁴. Les données sont disponibles librement par un lien web⁵. Elles comportent entre autres :

- les prix immobiliers (indices de prix de vente et de loyer des logements)
- nombre et montant des ventes immobilières
- taux d'intérêt
- durée des crédits immobiliers
- dette immobilière des ménages
- valeur d'investissements en logement, en actions, en obligations ou en or
- prix à la consommation
- indice du coût de la construction (ICC)
- population
- nombre et revenu des ménages

Ces séries portent sur la France mais également, pour certaines variables, sur les USA et le Royaume-Uni. Dans cette étude, seul le marché français sera étudié. Cette base de données est disponible sur une période de 1800 à aujourd'hui, cependant, pour des raisons d'homogénéité temporelle des données (stabilité politique, évolution des techniques financières, lois sociales...) j'ai choisi de n'utiliser ces données que de 1970 à aujourd'hui. Le marché de l'immobilier est représenté dans cette base par l'indice INSEE-Notaires. En effet, une méthode a été mise au point par l'INSEE pour produire des indices de prix réguliers et fiables des logements anciens. L'indice est ainsi défini sur le site de l'INSEE : « *La difficulté à estimer des indices de prix pour les logements tient au fait que le prix d'un logement donné est rarement observé et que la qualité des logements évolue au cours du temps. Pour traiter cette difficulté, la méthode retenue repose sur des modèles économétriques (modèles hédoniques), fondés sur l'estimation des prix relatifs des caractéristiques des logements.* »⁶. C'est en fait une évolution relative à une année (ici 2000) et nettoyée des effets des variations qualitatives du marché. Par exemple, si dans une année ne sont vendus que des appartements 5 pièces alors que l'année précédente seul les 2 pièces se sont vendus, cela ne veut pas dire que les prix auraient soudainement augmentés sur le marché immobilier Français. Il faut donc « nettoyer » la partie du prix liée au nombre de pièces et à l'emplacement pour comparer uniquement la partie marché du bien immobilier. C'est ce que fait l'indice INSEE-Notaires.

Si l'on compare l'évolution de l'indice des prix immobilier en France (y compris Paris) à Paris, on obtient le graphique suivant :

⁴ <http://www.cgedd.developpement-durable.gouv.fr/prix-immobilier-evolution-1200-a1048.html>

⁵ http://www.cgedd.developpement-durable.gouv.fr/IMG/xls/valeur-immobilier-1800-2015_cle2a7a51.xls

⁶ L'intégralité de la description de l'indice ainsi que des résultats est disponible sur le site <http://www.INSEE.fr/fr/publications-et-services>

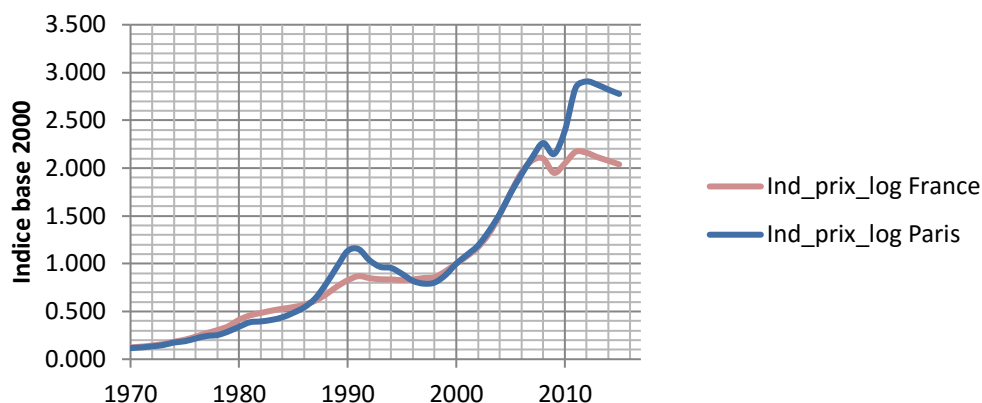


Figure 3 : Evolution de l'indice de prix immobilier France / Paris (Base 2000)

Il faut préciser que l'indice INSEE-Notaires ne commence qu'en 1992. Pour les années antérieures, Jacques Friggit a reconstitué l'indice à partir de l'indice de Gaston Duon⁷.

On remarque immédiatement que les prix à Paris et en France ont eu une évolution parallèle entre 1970 et 2009 (avec plus de volatilité pour Paris) mais qu'entre 2010 et 2011 une divergence extrêmement forte a été enregistrée.

Dans le GSE, les actifs sont modélisés hors inflation et l'inflation est modélisée à part et peut-être ajoutée si besoin. Dans cette logique, voyons comment la correction de l'inflation impacte l'évolution de l'indice de prix immobilier.

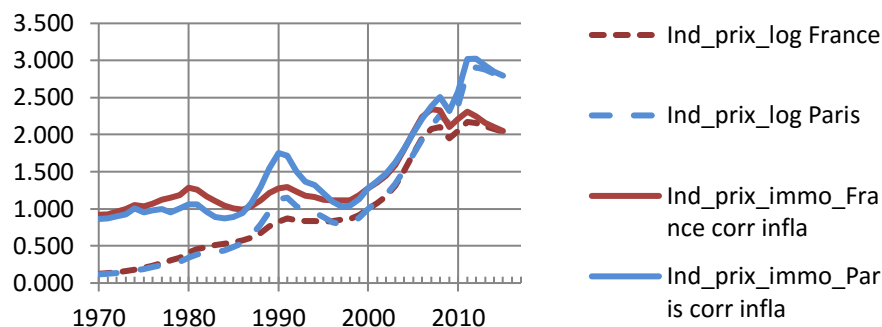


Figure 4 : Comparaison des indices immobiliers corrigés de l'inflation

La correction d'inflation ne change pas sensiblement les aspérités des courbes. Cependant cette correction apporte une meilleure représentation de l'évolution du marché car nettoyé de l'effet « monnaie ». C'est l'indice corrigé de l'inflation que le modèle tentera de répliquer dans la suite.

Si l'on regarde l'évolution des Taux d'intérêts long terme (reflet des prêts immobiliers aux particuliers) et de l'inflation :

⁷Le prix des logements sur le long terme, CGEDD, Jacques Friggit, p.19 (http://www.cgedd.developpement-durable.gouv.fr/IMG/pdf/evolution-prix-immobilier-friggit_cle0c611b.pdf)

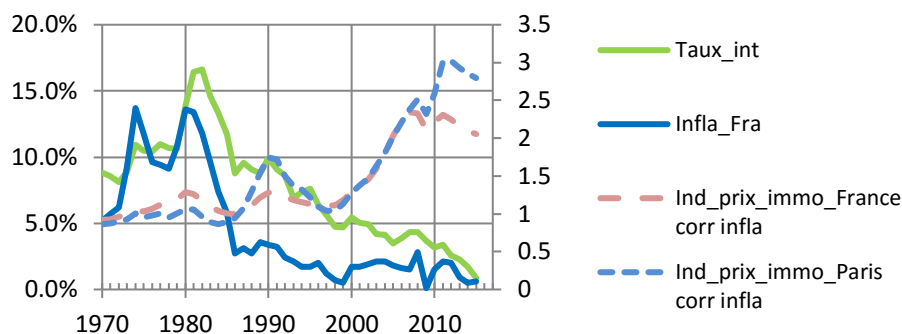


Figure 5 : Evolution comparée des taux d'intérêts, inflation et indices immobiliers

La figure 5 laisse penser qu'un lien peut-être établi entre le Taux d'intérêt long terme, l'inflation et l'évolution des prix immobilier. Une analyse technique des données doit pouvoir confirmer les liens et corrélations entre ces variables.

I.2.2 Analyse des données

J'ai cherché dans un premier temps les corrélations entre le marché de l'immobilier et les autres variables économiques. J'ai pour cela utilisé différentes méthodes d'analyse multi variée et temporelle.

I.2.2.1 Analyse en composantes principales et méthode dérivées

J'ai réalisé une étude complète sur le sujet que je joins en annexe n°7 et dont je ne reporte ici que les conclusions.

L'étude (qui regroupe plusieurs méthodes d'analyse multivariée: ACP, ACM, PCR, DisQual, PLSR, FDA), permet de conclure que l'immobilier est corrélé au taux d'emprunt mais non linéairement. Il en va de même pour la durée de l'emprunt. Il existe également des années de baisse qui ne corresponde pas aux autres et qui semble être liée à des crises (économiques ou financières). Les chocs ont donc un rôle important dans la formation des prix et ce point doit se retrouver dans le modèle qui sera défini par la suite.

Enfin, les années se suivent et se ressemblent. On parle alors d'autoregressivité. Les années passées influence le présent et ce point mérite une étude plus spécifique.

I.2.2.2 Analyse de l'autorégressivité : modèle ARMA/ARIMA

L'analyse temporelle des prix de l'immobilier d'habitation corrigés de l'inflation sera réalisée à l'aide de l'outil classique qu'est le modèle ARMA/ARIMA. C'est un modèle qui définit la variable l'année n comme étant une combinaison linéaire des variables en $n-i$ (i allant de 1 à p , on parle de AR(p)) et la somme d'une autre combinaison linéaire des résidus en $n-j$ (j allant de 1 à q , on parle de MA(q)).

Dans notre cas, on l'applique sur les différences entre chaque période (n et $n+1$) de l'indice immobilier. C'est donc une Intégration (d'ordre 1) et on parlera alors de modèle ARIMA ($p,1,q$).

Ce qui visuellement est l'analyse des courbes des différences annuelles de l'indice l'immobilier suivantes :

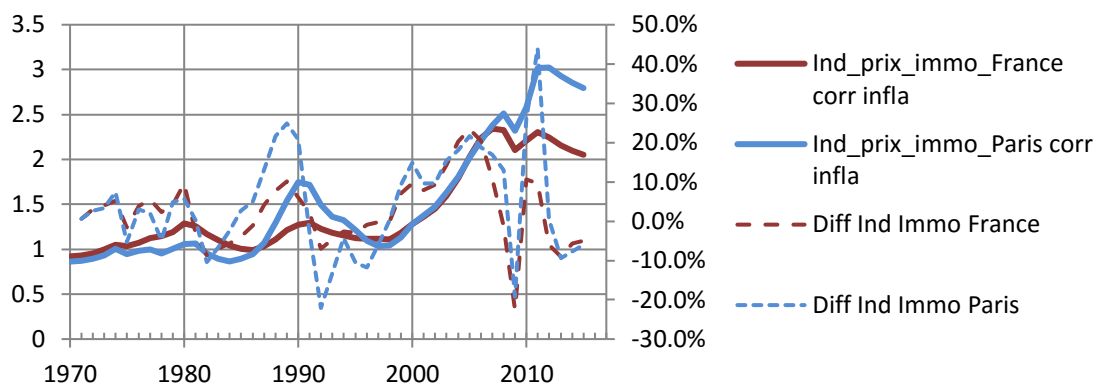


Figure 6 : Courbe des différences d'indice sur chaque période

Le critère de sélection des meilleurs modèles sera le critère d'information d'Akaike dit critère AIC⁸. Le meilleur modèle au sens de ce critère étant le modèle avec le plus faible AIC. Le tableau ci-dessous reprend les résultats du critère AIC pour différent p et q et pour l'indice immobilier à Paris. Les calculs ont été réalisés avec le logiciel R.

p	q	AIC
0	0	-66,82
0	1	-86,60
0	2	-84,77
1	0	-81,43
1	1	-84,81
1	2	-82,89
2	0	-80,85
2	1	-83,49
2	2	-82,40
3	0	-81,13
3	1	-83,30
3	2	-81,61
4	0	-79,49
4	1	-81,50
4	2	-79,61

On remarque que ARIMA (0,1,1) donne le meilleur résultat mais que l'on a des résultats proches pour les modèles ARIMA(1 ou 2 ou 3, 1,1). Cela confirme l'autorégression de la série temporelle sur des périodes relativement courtes (1 à 3 ans) et peu d'influence des résidus.

Le résultat est peu différent lorsque l'on réalise le même test sur la différence de l'indice des prix immobilier France. Enfin, ce résultat est important car il doit être retrouvé ou exprimé dans le modèle qui sera développé.

Nous avons maintenant les éléments qui doivent se retrouver dans le modèle (Taux et durée d'emprunt, influence des chocs, autorégressivité sur 1 à 3 ans). Contrairement au marché

⁸ Le **critère d'information d'Akaike** est une mesure de la qualité d'un modèle statistique proposée par Hirotugu Akaike en 1973. Le critère d'information d'Akaike s'écrit comme suit:

$$AIC = 2.k - 2. \ln(L)$$

où k est le nombre de paramètres à estimer du modèle et L est le maximum de la fonction de vraisemblance du modèle.

financier, chaque vente est unique et pourtant au global, le marché immobilier semble suivre des règles sous-jacentes.

II Développement du modèle du marché immobilier

Cette partie va définir les bases mathématiques et le développement théorique qui va conduire à l'élaboration du modèle stochastique d'évolution du prix moyen sur le marché immobilier.

Par souci de simplification, il est fait le choix de ne considérer que 2 types d'acteurs : les acheteurs et les vendeurs. En effet, on pourrait considérer par exemple, 4 groupes que sont les acheteurs, les vendeurs, les acheteurs/vendeurs (qui vendent pour acheter) et les institutionnels. Chacun des ces 4 groupes ayant des contraintes et objectifs différents mais le modèle mathématique serait alors considérablement complexifié. C'est pourquoi j'ai fait le choix de ne considérer le marché immobilier que comme une négociation entre un groupe de vendeurs et un groupe d'acheteurs homogènes.

II.1 Présentation des hypothèses et développements

Pour définir un modèle mathématique, il faut d'abord faire des hypothèses simplificatrices. Ces simplifications ont pour objectif de ne conserver que les principaux effets sous-jacents aux phénomènes observés. Les simplifications devant par la suite être testées pour s'assurer de leurs acceptabilités. Concernant la modélisation du marché immobilier d'habitation parisien qui va être étudié par la suite, je pose deux hypothèses simplificatrices principales et préalables au développement mathématique :

La première est que le temps va être discrétisé en années. Cela revient à considérer que le marché de l'immobilier est ouvert un jour par an, à date fixe et qu'il n'est pas possible de réaliser une vente un autre jour de l'année. Ce pas de temps correspond à celui des données.

La deuxième est de définir l'ensemble des acheteurs et des vendeurs de bien comme deux variables aléatoires. A chaque période, chaque membre du groupe des acheteurs fait une proposition pour tenter d'acquérir le bien convoité. Vu de l'extérieur, les propositions de prix du groupe se comportent comme une variable aléatoire centrée autour de sa moyenne, les vendeurs étant également modélisés sur le même principe. Suivant cette dernière hypothèse, il faut donc maintenant donner une forme à la répartition des offres des acheteurs et des vendeurs au sein de leur groupe respectif. En terme mathématique, on parle de fonction de densité de probabilité.

II.1.1 Normalité vs. log normalité

Les vendeurs de bien immobilier doivent se positionner sur un marché dont la seule valeur de référence est un prix au m² pour une localisation (approximative). Ils vont donc naturellement, selon les caractéristiques propres de leur bien, leur propre évaluation et leur situation (pressé de vendre ou non par exemple) se placer autour de la valeur de référence qui n'est en fait qu'une estimation de la valeur moyenne de la période présente basée sur la valeur moyenne de la période passée. Les biens sont par nature très variés (taille, étages, balcon, parking, voisinage, proximité école,...) ainsi que les vendeurs (optimiste/pessimiste, pressé/non pressé, avec une bonne/mauvaise référence, ...).

La troisième hypothèse consiste à interpréter les prix proposés par les vendeurs comme étant le produit des coefficients représentant le bien et le vendeur (coef étage X coef parking

X coef pressé X ...) par le prix moyen de vente de la période précédente. C'est d'ailleurs la méthode utilisée par les agents immobiliers lorsqu'ils font l'estimation d'un bien.

Considérons dans un premier temps qu'un seul critère différencie les biens d'un micro-marché (marché d'un type de bien sur une zone géographique limitée).

S'il on devait tirer au hasard un vendeur parmi le groupe des vendeurs, le coefficient appelé α pourrait être considéré comme un paramètre aléatoire.

Le prix de vente proposé par un vendeur i s'écrit alors

$$P_i = \alpha_i \cdot P_0$$

Avec P_0 le prix référence qui est la moyenne des ventes sur la période passée. On remarque que $\alpha_i > 0$ pour que les propositions des vendeurs soit > 0

Si l'on postule que toutes choses égales par ailleurs, les propositions des vendeurs (P_i) doivent rester « relativement » stables alors en prenant la moyenne des propositions des vendeurs, et en notant \tilde{P} le prix moyen demandé par les vendeurs alors

$$\tilde{P} = \frac{1}{Nb \text{ vendeurs}} \sum_i^{Nb \text{ vendeurs}} \alpha_i \cdot P_0$$

$$\tilde{P} = P_0 \cdot \frac{\sum_i^{Nb \text{ vendeurs}} \alpha_i}{Nb \text{ vendeurs}}$$

Pour que les prix soient stables en l'absence de modification des conditions de marché, il faut que la moyenne (arithmétique) des propositions des vendeurs soit proche de la moyenne des ventes sur la période passée. Pour rappel, nous sommes toujours dans le cadre d'un micro-marché (local géographiquement et même type de bien). Soit la conséquence suivante :

$$\forall Nb \text{ vendeurs}, \frac{\sum_i^{Nb \text{ vendeurs}} \alpha_i}{Nb \text{ vendeurs}} \approx 1 \text{ avec } \alpha_i \in]0; +\infty[$$

On en déduit que les moyennes des différentes valeurs que peut prendre le coefficient α_i doit être très proche de 1 et donc que la variable aléatoire α est centrée sur 1 et > 0

En pratique, pour la description du marché immobilier du prix au m², le coefficient serait compris entre 0,1 et 10 au maximum. Soit $\log(\alpha_i) \in [-2,3; +2,3]$

Le produit des coefficients doit lui aussi rester compris entre 0,1 et 10. Ces deux contraintes impliquent que chaque coefficient participe de façon « homogène » à la volatilité totale de l'ensemble (pas de coefficient prépondérant sur la volatilité)

On peut donc considérer les coefficients comme étant centrés, de faibles amplitudes et sans coefficient prépondérant en terme de volatilité.

Ainsi, si l'on considère la formation du prix d'un bien comme étant le produit de n coefficients (balcon, parking, standing,...), ils devront tous avoir individuellement les propriétés décrites. Dans ce cas, une offre s'écrira :

$$P_i = \alpha_i \cdot \beta_i \cdot \gamma_i \dots P_0$$

Avec $\alpha_i, \beta_i, \gamma_i \dots$ le produit de n variables aléatoires quasi-centrées et de moment 2 homogènes (sans prépondérance). Cela permet d'utiliser le théorème central limite dans les conditions de Lyapounov⁹.

⁹ <http://mathworld.wolfram.com/LyapunovCondition.html>

En effet, le théorème central limite de Laplace démontre que toute somme de variables aléatoires indépendantes et identiquement distribuées et centrées tend vers une variable aléatoire gaussienne. L'extension du TCL¹⁰ étant que lorsque qu'il s'agit d'un produit de variable aléatoire, la convergence en loi est log-normale.

Lyapounov ajoute qu'il n'est pas nécessaire d'être identiquement distribué à condition que pour $n \rightarrow +\infty$ (nombre de variable aléatoire), le rapport des moments d'ordre 2 et 3 $\rightarrow 0$ (pas de variable prépondérante en terme d'écart type).

Sans connaître précisément le nombre de coefficients utilisés pour l'estimation du prix des biens, on peut raisonnablement postuler que la répartition des prix de vente (proposés par les vendeurs) tende vers une distribution log-normale autour de la moyenne des estimations du prix moyen au m² (issu du prix au m² de la période passée).

En ce qui concerne les acheteurs, on peut reprendre entièrement la démonstration faite pour les vendeurs mais cette fois avec des variables aléatoires différentes (apport, taux d'emprunt, taux d'endettement accepté, revenus, ...). On peut ainsi postuler qu'ils sont eux aussi log-normalement distribués autour de la moyenne des propositions des acheteurs. Les estimations du prix moyen acheteur et vendeur doivent la plupart du temps diverger et évoluer différemment pour rendre compte de la variation des prix et du volume des ventes que l'on observe sur le marché.

Cela est vrai pour un bien type (ex : appartement 2 pièces dans le 5^{ème} arrondissement de paris). On verra au paragraphe II.1.2 comment ceci peut-être généralisé à l'ensemble du marché.

II.1.2 Invariance de la distribution des groupes par passage du marché local au global

Le modèle concerne à ce stade un type de bien sur un marché local. Pour étendre le raisonnement au marché immobilier dans son ensemble, il faut pouvoir trouver une référence commune faisant abstraction des disparités qualitatives et quantitatives entre les différentes zones géographiques.

C'est ce que fait l'indice INSEE-Notaires. L'approche hédonique consiste à postuler l'existence d'un prix implicite commun à tous les biens qui sera corrigé selon les propriétés propres du bien pour retrouver le prix facial. Les coefficients de correction de chaque paramètre de différenciation (balcon, étage, parking...) sont appelés des coefficients hédoniques. Ils sont ensuite évalués par la méthode des moindres carrés¹¹.

En reprenant le résultat précédent, on peut écrire localement (voir paragraphe II.1.1) pour la proposition des vendeurs et avec

$$P_{i,local,type} = \alpha_{i,local} \cdot \alpha_{i,parking} \cdot \alpha_{i,standing} \dots P_{0,local,type}$$

avec

- $P_{i,local,type}$ le prix de vente des vendeurs i avec :
 - o local - la localisation,
 - o type - le type bien (T1, T2 ...)
- les coefficients α correspondants aux critères de différenciation. On parle de coefficients hédoniques.
- $P_{0,local,type}$ un prix pour un bien référence en localisation et en type

¹⁰ Théorème central limite

¹¹ L'intégralité de la description de l'indice ainsi que des résultats est disponible sur le site <http://www.INSEE.fr/fr/publications-et-services>

Soit

$$\log(P_{0,local,type}) = \log(P_{i,local,type}) - \sum_{k=1}^{nb\ coef} \alpha_{i,k}$$

$\alpha_{i,k}$ étant les coefficients permettant de corriger la référence $P_{0,local,type}$ par rapport aux particularités du bien (k regroupant les critères de surface, parking, standing, erreur propre à l'individu i...)

L'indice INSEE des notaires permet d'écrire au niveau national :

$$\log(P_{0\ national}) = \log(P_{0\ local,type}) - \sum_{j=1}^{nb\ coef} \delta_j$$

δ_j étant les coefficients permettant de corriger la référence par rapport au type de bien et sa localisation.

Il vient donc :

$$\log(P_{i,local,type}) = \log(P_{0\ national}) + \sum_{k=1}^{nb\ coef} \alpha_{i,k} + \sum_{j=1}^{nb\ coef} \delta_j$$

De même que pour les coefficients α_k (voir paragraphe II.1.1) les coefficients δ_j sont quasiment centrés avec un écart type de faible amplitude à condition qu'il n'y ait pas de mode trop important au sein d'un des coefficients. C'est pour cela que l'on traitera la région parisienne et le reste du territoire séparément. En effet, on peut penser que la répartition du coefficient « région » (ou un équivalent) va avoir une variance importante par rapport aux variances des autres critères et ralentir fortement la convergence en loi log-normale de la distribution des vendeurs (condition de Lyapounov).

Ainsi, le nombre de coefficients composants le prix étant augmenté, la convergence en loi Log-normale est donc plus rapide que lorsque l'on observe l'ensemble du groupe de Vendeur au niveau local. Il est important de noter ici que la référence de prix est l'indice INSEE qui est une référence « nettoyée » des effets qualitatifs et des disparités géographiques. C'est pour cette raison qu'il sera l'indicateur modélisé dans le reste de l'étude. Plus précisément, nous utiliserons l'indice INSEE pour Paris. En effet, c'est le marché qui a été le mieux suivi et dont le volume permet de supposer une distribution log-normale des propositions des acteurs. Cela évite ainsi de mixer des marchés trop différents qui ralentiraient la convergence en loi log-normale. Désormais muni d'une distribution des acheteurs et des vendeurs autour de leurs moyennes respectives nous allons pouvoir développer un modèle dynamique de l'indice des prix de vente moyen pouvant s'appliquer à l'immobilier Parisien.

II.1.3 Acheteur/vendeur et prix moyen: modélisation stochastique

Une autre hypothèse simplificatrice du modèle ayant probablement le plus d'impact sur la formulation mathématique est que les vendeurs, pour estimer leur bien, se basent sur les prix moyens des ventes réalisées la période précédente. Cette partie est donc consacrée à la modélisation du prix de vente moyen. C'est le prix moyen issu de la rencontre du groupe des acheteurs et des vendeurs : soit du point de vu mathématique, de deux variables aléatoires

Log-normalement distribuée (voir II.1.2). On peut illustrer ce point par le schéma suivant avec P les prix de vente et t le temps:

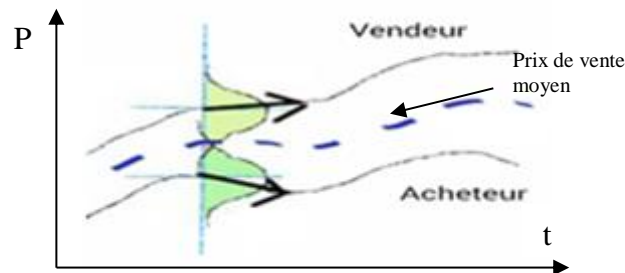


Figure 7 : schéma de principe de l'approche stochastique

Il faut ainsi établir une règle pour comptabiliser la conclusion d'une vente. Ainsi, la vente va se réaliser lorsque l'offre de l'acheteur sera égale ou supérieure à l'offre du vendeur, et l'hypothèse est faite que la vente se fera au prix du vendeur.

Alors, pour un bien type, le prix moyen des ventes sera l'espérance mathématique de cette vente.

Concernant la volatilité qui sera utilisée dans les fonctions de densités, elle est constituée de différentes informations :

- Les différents types de biens qui représentent la dispersion « basique » du marché en terme de prix (ou d'indice des prix INSEE). C'est la dispersion lorsque l'offre et la demande de bien n'évoluent pas.
- Une dispersion des acheteurs/vendeurs due au stress et à la concurrence qui poussent les acteurs à faire évoluer leurs budgets ou à partir sur un autre marché (exemple pour un acheteur : en passant d'un appartement 3 pièces à 2 pièces). Cela peut également être interprété comme une variation de l'offre et/ou de la demande.

Moyennant de nombreuse hypothèse, il parait possible de faire un lien entre la volatilité implicite qui va être utilisée ici et la théorie classique de l'offre et de la demande. Cela demande notamment des hypothèses sur le comportement des nouveaux entrant/sortant (devancer ou suivre les prix du marché). Cependant, j'ai choisi d'utiliser la volatilité des fonctions de densité des groupes acheteur/vendeur comme variable représentant la variation de l'offre et de la demande ainsi que les comportements associés.

II.1.3.1 Calcul du prix moyen des ventes réalisées

Afin de simplifier le modèle, considérons un marché ouvert une fois par période T sans cout de transaction et constitué de deux groupes (Acheteurs et vendeurs). Chaque acheteur et vendeur se rencontrent aléatoirement et se font des propositions. La vente est faite lorsque la proposition de l'acheteur est supérieure à celle du vendeur. La vente est alors conclue au prix du vendeur. Nous pouvons modéliser la proposition globale du groupe des acheteurs en appelant X_{A_0} la moyenne des offres des acheteurs lors du précédent marché. A l'ouverture du nouveau marché, chaque acheteur se positionne en fonction de l'évolution commune au groupe (le drift, soit par exemple une baisse des taux valable pour tous les acheteurs) et une partie propre.

On peut donc écrire l'évolution de la proposition d'un acheteur pour le marché T_1 (par rapport au marché T_0) comme étant :

$$\frac{dX_{A_T}}{X_{A_T}} = \mu_A + \sigma_A \cdot U_A$$

Avec μ_A , appelé drift ou terme de dérive, l'évolution en pourcentage de la capacité d'achat immobilier moyen, σ_A volatilité implicite des acheteurs et U variable aléatoire normale centrée réduite $\sim \mathcal{N}(0,1)$. Ce qui peut s'écrire également :

$$X_{A_T} = X_{A_0} e^{\mu_A + \sigma_A \cdot U_A}$$

On peut remarquer la proximité avec un mouvement Brownien géométrique mais il n'y a pas ici de notion de diffusion ni de temps, le marché est ici instantané toutes les périodes T. De même pour le groupe des vendeurs :

$$X^*_{V_T} = X_{V_0} e^{\mu_V + \sigma_V \cdot U_V}$$

μ_V étant l'évolution moyenne en % demandée par les acheteurs sur la période. $X^*_{V_T}$ est le prix demandé par le vendeur mais ce n'est pas le prix proposé du point de vu de l'acheteur. Il y a, entre les deux acteurs, les frais de notaire ainsi que la commission de l'agent immobilier. Nous ferons donc la différence entre $X^*_{V_T}$ prix proposé par le vendeur et X_{V_T} prix de vente proposé vu par l'acheteur. Il vient donc

$$X_{V_T} = X_{V_0} e^{\mu_V + \sigma_V \cdot U_V + \mu_{fees}}$$

Avec μ_{fees} taux continu des frais de la vente.

Le prix moyen des ventes peut alors s'écrire selon le principe des probabilités conditionnelles:

$$E[X_V \mid X_A > X_V] = \frac{E[X_V \cdot I_{X_A > X_V}]}{Proba(X_A > X_V)}$$

Avec $E[X_V \mid X_A > X_V]$ = *Espérance de X_V sachant que $X_A > X_V$*
 $I_{X_A > X_V}$ la fonction indicatrice qui vaut 1 quand $X_A > X_V$ et 0 sinon

Commençons par déterminer $Proba(X_A > X_V)$:

$$X_A > X_V \Leftrightarrow X_{A_0} e^{\mu_A + \sigma_A U_A} > X_{V_0} e^{\mu_V + \sigma_V U_V + \mu_{fees}}$$

Soit

$$\ln(X_{A_0}) + \mu_A + \sigma_A U_A > \ln(X_{V_0}) + \mu_V + \sigma_V U_V + \mu_{fees}$$

Ou encore

$$\ln\left(\frac{X_{A_0}}{X_{V_0}}\right) + (\mu_A - \mu_V - \mu_{fees}) > (\sigma_V \cdot U_V - \sigma_A \cdot U_A)$$

$(\sigma_V \cdot U_V - \sigma_A \cdot U_A)$ est une somme de variables normales centrées. C'est donc une variable aléatoire normale centrée de volatilité $(\sigma_V^2 + \sigma_A^2 - 2 \cdot \sigma_V \cdot \sigma_A \cdot \rho)$. ρ étant le coefficient de corrélation entre les deux groupes. On écrira dans la suite:

$$\bar{\sigma} = \sqrt{\sigma_V^2 + \sigma_A^2 - 2 \cdot \sigma_V \cdot \sigma_A \cdot \rho}$$

On peut alors simplifier l'écriture et la compréhension en notant :

$$\sigma_V \cdot U_V - \sigma_A \cdot U_A = \bar{\sigma} \cdot U$$

Avec U variable aléatoire normale centrée réduite $\sim \mathcal{N}(0,1)$

D'où l'inégalité

$$U < \frac{\ln\left(\frac{X_{A_0}}{X_{V_0}}\right) + (\mu_A - \mu_V - \mu_{fees})}{\bar{\sigma}}$$

On pose alors :

$$d_1 = \frac{\ln\left(\frac{X_{A_0}}{X_{V_0}}\right) + (\mu_A - \mu_V - \mu_{fees})}{\bar{\sigma}}$$

On a ainsi : $Proba(X_A > X_V) = Proba(U < d_1) = N(d_1)$

N() étant la fonction de répartition de la loi normale centrée réduite.

Concernant le terme $E[X_V \cdot I_{X_A > X_V}]$:

On peut considérer que la proposition de l'acheteur s'échange contre la proposition du vendeur au prix du vendeur augmenté des frais. C'est pourquoi nous pouvons nous remettre dans le cadre de l'évaluation de Margrabe¹² d'option d'échange. En effet

$$X_{A_T} = X_{A_0} e^{\mu_A + \sigma_A \cdot U_A} = X_{A_0} \cdot e^{\frac{\sigma_A^2}{2}} \cdot e^{(\mu_A - \frac{\sigma_A^2}{2})T + \sigma_A \cdot \sqrt{T} \cdot U_A} \text{ pour } T=1$$

$$X_{V_T} = X_{V_0} e^{\mu_V + \sigma_V \cdot U_V + \mu_{fees}} = X_{V_0} e^{\frac{\sigma_V^2}{2}} \cdot e^{(\mu_V + \mu_{fees} - \frac{\sigma_V^2}{2})T + \sigma_V \cdot \sqrt{T} \cdot U_V} \text{ pour } T=1$$

L'espérance peut donc directement s'écrire en prenant X_V comme numéraire :

$$E[X_V \cdot I_{X_A > X_V}] = X_{V_0}' \cdot e^{\mu_V + \mu_{fees}} \cdot N(d_2)$$

$$\text{Avec } d_2 = \frac{\ln\left(\frac{X_{A_0}'}{X_{V_0}'}\right) + \mu_A - \mu_V - \frac{1}{2} \bar{\sigma}^2}{\bar{\sigma} \cdot \sqrt{T}} \text{ et } \begin{cases} X_{V_0}' = X_{V_0} \cdot e^{\frac{\sigma_V^2}{2}} \\ X_{A_0}' = X_{A_0} \cdot e^{\frac{\sigma_A^2}{2}} \end{cases}$$

Soit

¹²William Margrabe, "The Value of an Option to Exchange One Asset for Another", Journal of Finance, Vol. 33, No. 1, (March 1978), pp. 177-186. Voir également « FINANCE DE MARCHE », Roland Portait & Patrice Poncet, p.388

$$E[X_V \cdot I_{X_A > X_V}] = X_{V_0} \cdot e^{\mu_V + \frac{\sigma_V^2}{2} + \mu_{fees}} \cdot N(d_2)$$

avec

$$d_2 = \frac{\ln\left(\frac{X_{A_0}}{X_{V_0}}\right) + (\mu_A - \mu_V - \mu_{fees})}{\bar{\sigma}} - \frac{\sigma_V^2 - \sigma_V \cdot \sigma_A \cdot \rho}{\bar{\sigma}}$$

et

$$d_2 = d_1 - \frac{\sigma_V^2 - \sigma_V \cdot \sigma_A \cdot \rho}{\bar{\sigma}}$$

Et peut ainsi noter :

$$E[X_{V_1} \setminus X_{A_1} > X_{V_1}] = X_{V_0} \cdot e^{\mu_{V_1} + \frac{\sigma_{V_1}^2}{2} + \mu_{fees}} \cdot \frac{N(d_2)}{N(d_1)}$$

Il faut noter que d_2 et d_1 sont évalués au temps 1

Par la suite nous noterons $E[X_{V_1} \setminus X_{A_1} > X_{V_1}] = \tilde{X}_1$ la valeur moyenne des ventes (frais inclus) à la période $T=1$.

Ces résultats étant non triviaux, il a été souhaitable de trouver une autre manière de calculer les prix de vente moyen afin de pouvoir confirmer les résultats analytiques obtenus ci-dessus. J'ai pour cela opté pour une autre approche souvent utilisée en calcul probabiliste à savoir le calcul intégral des fonctions de densité.

II.1.3.2 Méthode des Densité de probabilité Acheteur/vendeur et densités conjointes

Cette partie est développée dans l'annexe n°1. Ne sont présentés ici que les résultats.

Cette méthode consiste à utiliser les densités de probabilité des acheteurs et des vendeurs (pour mémoire de fonction Log-normale). Le prix de vente moyen étant alors de forme intégrale.

La densité de la variable aléatoire X_A , le prix proposé par un acheteur s'écrit alors :

$$f_{X_A}(x) = \frac{1}{\sigma_A \sqrt{2\pi}} \cdot \frac{1}{x} \cdot e^{-\frac{\left[\ln\frac{x}{X_{A_0}} - \mu_A T\right]^2}{2\sigma_A^2}} \quad \text{pour } x \geq 0$$

De même pour la densité des propositions de vendeurs X_V

$$f_{X_V}(y) = \frac{1}{\sigma_V \sqrt{2\pi}} \cdot \frac{1}{y} \cdot e^{-\frac{\left[\ln\frac{y}{X_{V_0}} - \mu_V T\right]^2}{2\sigma_V^2}} \quad \text{pour } y \geq 0$$

Il y a vente lorsque $X_A \geq X_V$ au prix de X_V . Sinon il n'y pas vente. Nous retrouvons donc la formule de probabilités conditionnelles :

$$\tilde{X}_{X_A \geq X_V} = \frac{E[X_V \cdot I_{X_A > X_V}]}{\text{Proba}(X_A > X_V)}$$

Avec $\tilde{X}_{X_A \geq X_V}$ le prix de vente moyen sur ce marché que l'on peut également écrire par les densités de probabilités :

$$\tilde{X}_{X_A \geq X_V} = \frac{\int_{-\infty}^{+\infty} \int_y^{+\infty} y \cdot f_{X_A}(x) \cdot f_{X_V}(y) \cdot dx \cdot dy}{\int_{-\infty}^{+\infty} \int_y^{+\infty} f_{X_A}(x) \cdot f_{X_V}(y) \cdot dx \cdot dy}$$

X_A et X_V sont considérés indépendants. Il y a vente si $X_A \geq X_V$ soit si $x \geq y$ d'où la borne inférieure d'intégration. Afin de faciliter l'intégration, on réalise le changement de variable en posant $z = \frac{x}{y}$ et $v = y$, qui donne $x = vz$ et $J = v = y$ (Jacobien). Les contraintes deviennent :

$$\begin{cases} v \geq 0 \text{ (prix de vente)} \\ z \geq 1 \text{ (condition de la vente)} \end{cases}$$

On peut ainsi écrire la fonction de densité: $f_{\frac{X_A}{X_V}, X_V} = f_{X_A, X_V}(vz, y) \cdot v$ ce qui permet d'écrire

$$\tilde{X}_{X_A \geq X_V} = \frac{\int_0^{+\infty} \int_1^{+\infty} v^2 \cdot f_{X_A}(vz) \cdot f_{X_V}(v) \cdot dz \cdot dv}{\int_0^{+\infty} \int_1^{+\infty} v \cdot f_{X_A}(vz) \cdot f_{X_V}(v) \cdot dz \cdot dv}$$

En appelant N la fonction de répartition de la loi $\mathcal{N}(0,1)$, on obtient alors (cf annexe n°1)

$$\tilde{X}_{X_A \geq X_V} = \frac{\int_0^{+\infty} \frac{1}{\sigma_{V2}\sqrt{2\pi}} \cdot e^{-\frac{[\ln v - \mu_{V2}]^2}{2\sigma_{V2}^2}} \left[1 - N\left(\frac{\ln v - \mu_{A2}}{\sigma_{A2}}\right) \right] \cdot dv}{\int_0^{+\infty} \frac{1}{\sigma_{V2}\sqrt{2\pi}} \cdot \frac{1}{v} \cdot e^{-\frac{[\ln v - \mu_{V2}]^2}{2\sigma_{V2}^2}} \left[1 - N\left(\frac{\ln v - \mu_{A2}}{\sigma_{A2}}\right) \right] \cdot dv}$$

J'ai comparé les deux méthodes sur plusieurs ensembles de variables et les prix moyens de vente calculés sont identiques (cf. annexe n°2). Le modèle analytique est ainsi validé. Il va servir de base au développement du modèle dynamique-stochastique d'évolution des prix sur un marché de bien non substituable.

II.1.3.3 Boucle de rétroaction des vendeurs

C'est le point essentiel et l'originalité du modèle. Le vendeur, lorsqu'il arrive sur le marché ou lorsqu'il révisé sa proposition, utilise implicitement les coefficients hédoniques vus précédemment pour adapter sa proposition en fonction des particularités de son bien. Mais il a besoin également d'une évaluation de la dynamique du marché dans son ensemble. Il n'a pour cela que l'évolution du prix moyen des ventes comme référence. Il fera donc évoluer sa proposition d'une période à l'autre en fonction des prix moyens constatés sur la période précédente (moins les frais de ventes) et de leurs variations entre les deux périodes précédentes. Je ne retiens ici que les deux dernières périodes car cela correspond au résultat de l'analyse temporelle (cf §I.2.2.2). Soit chronologiquement l'évolution des indices :

$$X_{V_0} \xrightarrow{\times e^{\left(\frac{\sigma_{V_0}^2}{2} + \mu_{fees}\right)} \times \frac{N(d_2)_0}{N(d_1)_0}} \tilde{X}_0 \xrightarrow{\times e^{\mu_{V_1} - \mu_{fees}}} X_{V_1} \xrightarrow{\times e^{\left(\frac{\sigma_{V_1}^2}{2} + \mu_{fees}\right)} \times \frac{N(d_2)_1}{N(d_1)_1}} \tilde{X}_1 \xrightarrow{\times e^{\mu_{V_2} - \mu_{fees}}} \dots$$

Proposition des vendeurs à T_0
Prix moyen constaté à T_0 à la fin du
Proposition des vendeurs à T_1 (net vendeur)
Prix moyen constaté à T_1 à la fin du

X_{V_1} la proposition moyenne des vendeurs à T=1 sera donc

$$X_{V_1} = \tilde{X}_0 \cdot e^{\mu_{V_1} - \mu_{fees}}$$

Ou encore

$$X_{V_1} = X_{V_0} \cdot e^{\mu_{V_1} + \frac{\sigma_{V_0}^2}{2}} \cdot \frac{N(d_2)_0}{N(d_1)_0}$$

Avec :

- μ_{V_1} fonction de l'évolution de marché pour la période de T=-1 à T=0
- d_2 et d_1 fonction de X_{V_0} , X_{A_0} , σ_{V_0} , σ_{A_0} et μ_{fees}

Pour le prix de vente moyen à T=1, il faut écrire :

$$\tilde{X}_1 = \tilde{X}_0 \cdot e^{\mu_{V_1} + \frac{\sigma_{V_1}^2}{2}} \cdot \frac{N(d_2)_1}{N(d_1)_1}$$

Avec les différents paramètres pris au temps T=1.

Nous n'avons pas encore abordé la nature de μ_{V_T} et le rôle qu'il joue dans l'établissement de l'équilibre des prix.

II.1.3.3.1 Coefficient d'anticipation ξ

Au moment du marché à T=1, le vendeur va utiliser le prix moyen de la dernière période mais aussi la variation (la tendance) des deux périodes précédentes T=-1 à T=0 pour se positionner. En effet, l'étude du paragraphe I.2.2 nous montre que les acteurs vont considérer que le passé récent va se reproduire. Mais il est probable également que les différents vendeurs ne vont pas tous en tenir compte de la même manière. Nous introduisons ici un coefficient d'anticipation ξ_i propre au vendeur i tel que :

$$\mu_{V_{1,i}} = \xi_i \cdot \ln\left(\frac{\tilde{X}_0}{\tilde{X}_{-1}}\right) + \mu_{V_{négo,i}}$$

Avec $\mu_{V_{négo,i}}$ le taux d'« ouverture des négociation » des vendeurs. En effet, tous les vendeurs, s'étant fait une opinion sur le prix de leur bien, vont tenter de surperformer le marché ou dans tout les cas vont créer une marge de négociation supplémentaire.

Ces deux variables ad hoc font entrer cette étude dans le cadre de la finance comportementale et seront, nous le verrons par la suite, centrale dans la dynamique du modèle.

On cherche ξ et $\mu_{V_{négo}}$ les variables du groupe des vendeurs dans son ensemble. Ils sont l'agrégation des comportements individuels ξ_i et $\mu_{V_{négo,i}}$.

Soit en reprenant la modélisation log-normal du groupe des vendeurs qui s'écrit

$$\frac{dX_{V_T}}{X_{V_T}} = \mu_V + \sigma_V \cdot U_V$$

Avec μ_V l'anticipation moyenne du groupe des vendeurs et en considérant que les différents vendeurs ont des coefficients d'anticipation et de négociation répartis autour de leur moyenne en suivant une loi normale, nous pouvons écrire :

$$\frac{dX_{V_{1,i}}}{X_{V_{1,i}}} = (\xi + \sigma_\xi \cdot U_\xi) \cdot \ln\left(\frac{\tilde{X}_0}{\tilde{X}_{-1}}\right) + \mu_{V_{négo}} + \sigma_V \cdot U_V$$

Avec :

- $\xi = E[\xi_i]$ dont un estimateur non biaisé est $\frac{\sum \xi_i}{\text{Nb Vendeurs}}$
- $\mu_{V_{négo}} = E[\mu_{V_{négo,i}}]$ dont un estimateur non biaisé est $\frac{\sum \mu_{V_{négo,i}}}{\text{Nb Vendeurs}}$
- chaque U variable aléatoire normale centrée réduite $\sim \mathcal{N}(0,1)$

Le terme $\sigma_V \cdot U_V$ reprenant à la fois la répartition objective (type de bien) et subjective (marge de négociation). Nous pouvons ici agréger ces deux volatilités car elles sont considérées non corrélées.

Ce qui permet de réécrire

$$X_{V_{1,i}} = X_{V_0} e^{\xi \cdot \ln\left(\frac{\tilde{X}_0}{\tilde{X}_{-1}}\right) + \mu_{V_{négo}} + \ln\left(\frac{\tilde{X}_0}{\tilde{X}_{-1}}\right) \cdot \sigma_\xi \cdot U_\xi + \sigma_V \cdot U_V}$$

Soit en définissant une nouvelle variable aléatoire normale centrée réduite $U_{\xi V}$

$$X_{V_{1,i}} = X_{V_0} e^{\xi \cdot \ln\left(\frac{\tilde{X}_0}{\tilde{X}_{-1}}\right) + \mu_{V_{négo}} + \sigma_{\xi V} \cdot U_{\xi V}}$$

Avec

$$\sigma_{\xi V}^2 = \sigma_V^2 + \left(\ln\left(\frac{\tilde{X}_0}{\tilde{X}_{-1}}\right)\right)^2 \sigma_\xi^2 + 2 \cdot \ln\left(\frac{\tilde{X}_0}{\tilde{X}_{-1}}\right) \cdot \sigma_\xi \cdot \sigma_V \cdot \rho_{\xi V}$$

Il faudrait donc en théorie remplacer σ_V par $\sigma_{\xi V}$ dans l'ensemble des expressions pour le calcul des prix moyens. Cependant, l'étude de l'annexe 4 permet de montrer que l'on peut négliger cette correction si l'on considère qu'il n'y a pas de corrélation entre la dispersion « de fond » des vendeurs (quelque soit le prix moyen et son évolution) et la dispersion « de tendance » (qui dépend de la variation des prix) soit $\rho_{\xi V} = 0$. Dans ce cas, nous pouvons donc bien utiliser la formule simplifiée :

$$\mu_{V_1} = \xi \cdot \ln\left(\frac{\tilde{X}_0}{\tilde{X}_{-1}}\right) + \mu_{V_{négo}}$$

On peut alors réécrire l'évolution de l'indice des prix de vente comme suit :

$$\frac{\tilde{X}_1}{\tilde{X}_0} = e^{\xi \cdot \ln\left(\frac{\tilde{X}_0}{\tilde{X}_{-1}}\right) + \frac{\sigma_{V_1}^2}{2} + \mu_{V_{négo}}} \cdot \frac{N(d_2)_1}{N(d_1)_1}$$

Soit en généralisant pour une période i quelconque

$$\boxed{\frac{\tilde{X}_{i+1}}{\tilde{X}_i} = e^{\xi \cdot \ln\left(\frac{\tilde{X}_i}{\tilde{X}_{i-1}}\right) + \frac{\sigma_{V_{i+1}}^2}{2} + \mu_{V_{négo}}} \cdot \frac{N(d_2)_{i+1}}{N(d_1)_{i+1}}} \quad (1)$$

Cette formule d'évolution de l'indice des prix est ce que j'appelle le « modèle dynamique-stochastique de l'indice des prix sur un marché de biens non substituables »

Soit sous une forme différente :

$$\ln(\tilde{X}_{i+1}) - \ln(\tilde{X}_i) = \xi \cdot [\ln(\tilde{X}_i) - \ln(\tilde{X}_{i-1})] + \frac{\sigma_{V_{i+1}}^2}{2} + \mu_{V_{négo}} + \ln \frac{N(d_2)_{i+1}}{N(d_1)_{i+1}} \quad (2)$$

Cette formule n'est pas simple à appréhender. Il est souhaitable de pouvoir l'approximer pour pouvoir en étudier le comportement et valider son utilisation dans le cadre d'un GSE. Pour cela, comme il est d'usage pour les équations différentielles, on va utiliser une discrétisation par des schémas de différences finis mais les termes non linéaires doivent être approximés afin de faciliter la résolution analytique.

II.1.3.3.2 Linéarisation

Le terme $\ln \frac{N(d_2)}{N(d_1)}$ est non calculable analytiquement. Cependant, afin de simplifier la résolution, nous ferons une approximation de cette fonction par une droite affine fonction de $\ln \left(\frac{X_{A_i}}{\tilde{X}_i} \right)$. Soit

$$\ln \frac{N(d_2)_i}{N(d_1)_i} = \alpha \cdot \ln \left(\frac{X_{A_i}}{\tilde{X}_i} \right) + \beta$$

En effet, si l'on trace $\ln \frac{N(d_2)}{N(d_1)}$ en fonction de $\ln \left(\frac{X_{A_i}}{\tilde{X}_i} \right)$ pour différente valeur de volatilité, on remarque qu'une linéarisation donnera une bonne approximation.

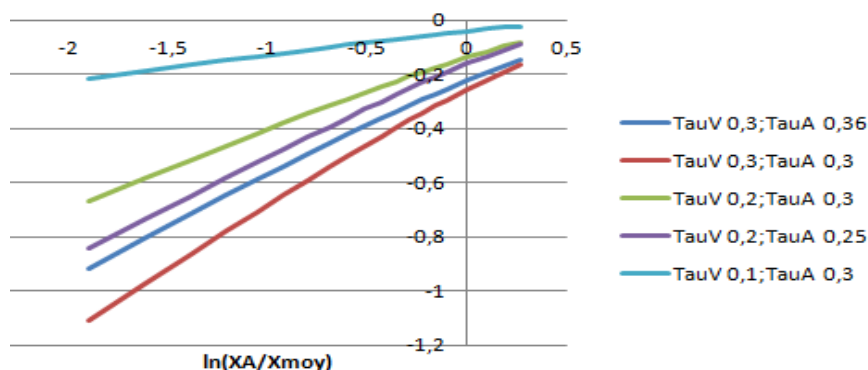


Figure 8 : $\ln \frac{N(d_2)}{N(d_1)}$ vs. $\ln \left(\frac{X_{A_i}}{\tilde{X}_i} \right)$

De plus, α et β dépendent de σ_V et σ_A et il est également possible par la méthode des moindres carré d'obtenir une linéarisation de ces deux variable en fonction des volatilités (cf. Annexe 4).

Par la méthode des moindres carrés, on peut donc approximer α par la formule

$$\alpha = \frac{\sigma_V - 0,688 \cdot \sigma_A + 0,124}{0,409}$$

Valable pour $20\% < \sigma_A, \sigma_V < 35\%$ (Erreur inférieur à 10% sur α)

De même pour β

$$\beta = \frac{0,467 \cdot \sigma_A - \sigma_V - 0,042}{0,630}$$

Valable pour $20\% < \sigma_A, \sigma_V < 35\%$ (Erreur inférieur à 10%).

Cela permet donc d'écrire l'équation dynamique de l'indice des prix :

$$\ln(\tilde{X}_1) - \ln(\tilde{X}_0) = \xi \cdot [\ln(\tilde{X}_0) - \ln(\tilde{X}_{-1})] + \frac{\sigma_{V_1}^2}{2} + \mu_{V_{négo}} + \alpha \cdot \ln(X_{A_1}/\tilde{X}_1) + \beta$$

Pour simplifier l'écriture nous poserons $\ln(\tilde{X}_i) = Y_i$ soit :

$$Y_1 - Y_0 = \xi \cdot [Y_0 - Y_{-1}] + \frac{\sigma_{V_0}^2}{2} + \mu_{V_{négo}} - \alpha \cdot Y_1 + \alpha \cdot \ln(X_{A_1}) + \beta$$

Avec $\alpha = f_1(\sigma_V, \sigma_A)$ et $\beta = f_2(\sigma_V, \sigma_A)$ et f_1, f_2 fonctions affines.

Cette linéarisation va permettre d'écrire une équation différentielle simplifiée (avec coefficients constants) de l'évolution de l'indice des prix de vente moyen.

II.1.3.3 Approximation par une équation différentielle d'ordre 2

Afin de retrouver une formulation classique d'équation différentielle nous allons réécrire l'équation à l'aide des schémas de différence fini. Cela consiste à linéariser les dérivées de la façon suivante :

$$\frac{dY}{dt} \approx \frac{Y_{i+1} - Y_i}{T}$$

$$\frac{d^2Y}{dt^2} \approx \frac{Y_{i+1} - 2 \cdot Y_i + Y_{i-1}}{T^2}$$

Avec $T=1$ (période de durée 1 an)

L'hypothèse sous jacente étant que les variations sont faibles d'une période à l'autre et varient lentement. Soit si l'on reformule le premier terme de l'équation:

$$Y_1 - Y_0 - \xi \cdot [Y_0 - Y_{-1}] + \alpha \cdot Y_1 = \xi \cdot (Y_1 - 2Y_0 + Y_{-1}) + \xi \cdot Y_0 - Y_0 + Y_1 - \xi \cdot Y_1 + \alpha \cdot Y_1$$

Ce qui permet de réécrire l'équation sous la forme différentielle suivante :

$$\boxed{\xi \cdot \frac{d^2Y}{dt^2} + (1 - \xi + \alpha) \cdot \frac{dY}{dt} + \alpha \cdot Y_0 = \frac{\sigma_V^2}{2} + \mu_{V_{négo}} + \alpha \cdot \ln(X_{A_1}) + \beta} \quad (3)$$

Cette équation différentielle linéaire du second ordre à coefficients constants est une approximation de la dynamique de l'indice des prix. Elle va permettre d'étudier l'évolution sur plusieurs périodes T.

Par analogie, on peut dire que le terme d'inertie ξ est équivalent à la masse dans un système masse-ressort. De même, $(1 - \xi + \alpha)$ peut être associé à un coefficient d'amortissement qui agit en fonction de la vitesse d'évolution des prix. Enfin, la force de rappel vers l'équilibre

dépend uniquement de α et donc de σ_A et σ_V (mesure du niveau de tension en offre et demande).

La solution de ce type d'équation différentielle est alors de la forme :

$$Y = Y_{Homogène} + Y_{Particulière}$$

La solution homogène est la solution de l'équation

$$\xi \cdot d^2Y + (1 - \xi + \alpha) \cdot dY + \alpha \cdot Y = 0$$

Pour cela on résout l'équation du second ordre

$$\Delta = (1 - \xi + \alpha)^2 - 4\alpha\xi$$

$$\Delta = 0 \Rightarrow \xi^2 - 2\xi(1 + 3\alpha) + 1 + 2\alpha + \alpha^2 = 0$$

C'est de nouveau une équation du second degré qui dépend de α .

$$\Delta_2 = 4(3\alpha + 1)^2 - 4(\alpha + 1)^2 = 0 \Rightarrow \alpha = 0$$

Il vient que $\Delta_2 > 0$ si $\alpha > 0$ (ce qui est toujours le cas). Et donc, $\Delta_2 > 0$ implique qu' ξ s'annule en deux points.

$$\xi_{12} = (1 + 3\alpha) \pm \sqrt{(3\alpha + 1)^2 - (\alpha + 1)^2}$$

L'application numérique donne, pour $\sigma_V = 20\%$ et $\sigma_A = 30\%$ (soit $\alpha = 0,27$), $\xi_{12} = 0,52$ et $3,1$. Faisons l'analyse du signe de Δ :

ξ	0,52		3,1		
Signe Δ	+	0	-	0	+

La solution de l'équation différentielle sera donc de trois types présentés dans le graphique suivant :

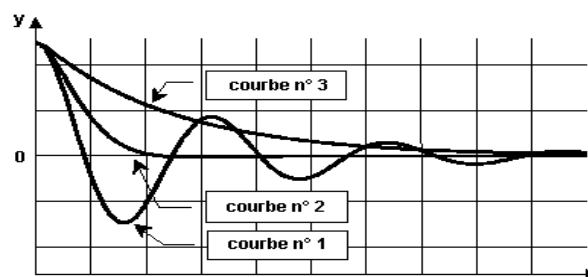


Figure 9 : Solutions d'une équation différentielle du second ordre à coefficients constants

- « oscillation amortie sous critique » pour $0,52 < \xi < 3,1$ (Figure 9, courbe n°1)

$$Y_H = e^{\frac{-(1-\xi+\alpha)}{2\xi} \cdot T} \left[A \cdot \cos\left(\frac{\sqrt{-(1-\xi+\alpha)^2 + 4\alpha\xi}}{2\xi} \cdot T\right) + B \cdot \sin\left(\frac{\sqrt{-(1-\xi+\alpha)^2 + 4\alpha\xi}}{2\xi} \cdot T\right) \right]$$

- « oscillation amortie sur critique » pour $\xi < 0,52$ et $\xi > 3,1$ (Figure 9, courbe n°3)

$$Y_H = A e^{\frac{-(1-\xi+\alpha)-\sqrt{(1-\xi+\alpha)^2-4\alpha\xi}}{2\xi} T} + B e^{\frac{-(1-\xi+\alpha)+\sqrt{(1-\xi+\alpha)^2-4\alpha\xi}}{2\xi} T}$$

- « oscillation amortie critique » pour $\xi = 0,52$ ou $\xi = 3,1$ (Figure 9, courbe n°2)

$$Y_H = (A.T + B). e^{\frac{-(1-\xi+\alpha)}{2\xi} T}$$

Avec T le numéro de la période considérée. Les constantes A et B dépendent des conditions aux limites. Si l'on se place dans le cas des oscillations sous-critique (qui sera le cas naturel dans le cadre de cette étude) et

$$\text{pour } T = 0, Y_H + Y_P = \ln(\tilde{X}_0)$$

Ainsi

$$A = \ln(\tilde{X}_0) - Y_P$$

si l'on considère une stabilité du marché avant = 0, soit $(Y_H + Y_P)'_0 = 0$ alors

$$B = \frac{-(1-\xi+\alpha)}{\sqrt{-(1-\xi+\alpha)^2+4\alpha\xi}} \cdot (\ln(\tilde{X}_0) - Y_P)$$

Notez bien que pour $\xi = 1 + \alpha$ (soit 1,27), le terme d'amortissement s'annule. C'est une oscillation libre. Au-delà, les oscillations s'amplifient d'une période à l'autre et les prix divergent. On peut donc raisonnablement éloigner ces possibilités et postuler que $\xi < 1,27$ (la plupart du temps) et donc que les prix tendent vers la stabilité en absence de perturbation. Maintenant que nous avons la solution homogène de l'équation différentielle, nous devons trouver une solution particulière.

La solution Particulière est très simple à trouver car c'est la solution de l'équation lorsque les dérivées (quelque soit l'ordre) s'annulent et que les prix se stabilisent. Ce qui donne

$$Y_P = \frac{\frac{\sigma_V^2}{2} + \mu_{V_{négo}} + \beta}{\alpha} + \ln(X_A)$$

On peut également dire que c'est la limite de la variable Y lorsque T est très grand et que les oscillations sont entièrement dissipées. Soit :

$$\lim_{T \rightarrow \infty} \ln(\tilde{X}_T) = Y_P$$

Nous avons ainsi la limite de convergence du prix moyen de vente (toutes choses égales par ailleurs). Ce point est très important et sera développé avec la notion de prix d'équilibre (paragraphe III.3). Cette approximation du modèle permet également de définir la période propre par

$$T_{propre} = 2 \cdot \pi \cdot \sqrt{\frac{\xi}{\alpha}}$$

Ce modèle approximé possède des propriétés intéressantes car en rapport avec des situations économiques réelles que nous allons voir dans le chapitre suivant.

II.2 Analyse simple du modèle

Nous avons à ce stade modélisé le groupe des acheteurs et le groupe des vendeurs par des variables aléatoires et l'on a démontré que l'on pouvait approcher leurs fonctions de répartition par une loi log-normale. Il a été défini une fonction d'échange permettant de calculer, en fonction des paramètres des variables aléatoires, le prix moyen d'échange (de vente) entre les deux groupes. Il a enfin été introduit une boucle de rétroaction conforme aux observations et qui consiste prendre en compte le passé récent dans l'établissement de l'offre des vendeurs. Nous allons voir, dans ce paragraphe, le comportement du modèle dans quelques cas simples. Pour mémoire, la formule est

$$\tilde{X}_i = \tilde{X}_{i-1} \cdot e^{\mu v_i + \frac{\sigma v_i^2}{2} + \mu v_{négo}} \cdot \frac{N(d_2)_i}{N(d_1)_i}$$

Il est donc facile de réaliser un tableur effectuant la récurrence d'une année à l'autre sans devoir calculer les équations d'ondes définies dans le paragraphe précédent. Cela facilite la réalisation des calculs et utilise le résultat exact.

II.2.1 Analyse de l'équation différentielle

Comme nous venons de la voir, le modèle dynamique-stochastique défini peut-être approximé par une équation différentielle du second ordre. Je n'ai pu trouver dans la littérature beaucoup de publication faisant référence à une équation d'évolution d'une variable économique régie par une telle équation différentielle. On peut voir dans la publication de Bohumil Stadnik¹³ qu'il est fait usage de solution type masse-ressort mais pour la volatilité en elle-même et sans justification de son utilisation. Ce point est tout de même intéressant car dans un modèle pleinement comportementaliste, il faudrait lier les volatilités et les prix (ce qui n'est pas le cas dans cette étude).

L'autre publication que j'ai pu trouver et utilisant une équation différentielle du même type est le fruit des travaux de Bouchaud et Cont¹⁴. Ils y définissent une équation différentielle d'ordre 2 d'équilibre des prix. Cependant, dans leur équation, le terme d'inertie provient d'un coefficient liant la différence entre l'offre et la demande à la variation des prix (épaisseur du marché ou liquidité) alors que dans mon cas, c'est ξ , la croyance en la reproduction du marché (coefficient comportemental) qui joue ce rôle. De plus, le terme d'amortissement est introduit par J.-P. Bouchaud et R. Cont par une décroissance logarithmique de l'offre et de la demande alors quelle est consécutive à l'utilisation du prix moyen des ventes (rapport $\frac{N(d_2)_i}{N(d_1)_i}$ pour la période i) dans le cadre de ma proposition.

L'établissement de ces deux équations sont donc difficilement comparable car ne venant pas de la modélisation des mêmes phénomènes.

Pour l'équation trouvée, la grande majorité des bouquets de variables « réalistes » conduisent à un système type masse ressort avec amortissement sous-critique.

Par exemple, si l'on fait varier le coefficient d'anticipation pour des volatilités fixées on obtient :

¹³ Bohumil Stadnik¹³, « Spring oscillations within financial markets », *Procedia, Social and behavioural Sciences*

¹⁴ J.-P. Bouchaud et R. Cont, « A Langevin approach to stock market fluctuations and crashes », *The European Physical Journal B*, 1998

Variation du coef d'anticipation

(avec $X_{a0}=1$; $X_{v0}=1,1$)

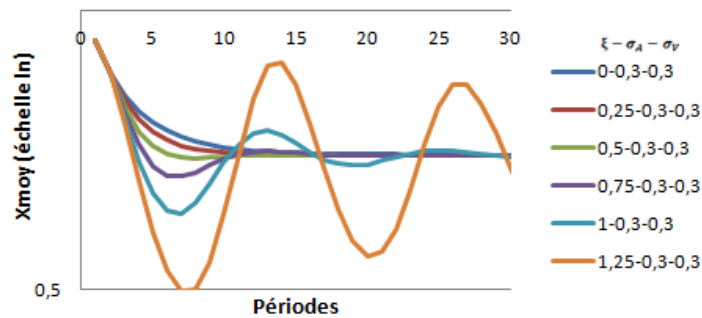


Figure 10 : effets du coefficient d'anticipation

On remarque que c'est une variable agissant sur la sensibilité dynamique comme l'on pouvait s'attendre. Cependant quelque soit ξ , la limite de convergence ne change pas. Il est difficile à ce stade d'avoir une évaluation précise de ξ . On peut légitimement penser que la valeur devrait être proche de 1 bien que n'étant pas constante a priori. En effet, cela correspond à l'anticipation (moyenne du groupe des vendeurs) que l'évolution des prix sur l'année dernière en % va se reproduire cette année à l'identique. Ne pouvant pas à ce stade justifier d'une valeur ou formule plus précise, j'utiliserai la valeur $\xi=1$. De la même manière, si l'on fixe le coefficient d'anticipation et que l'on fait varier les volatilités des 2 groupes :

Variation des Tau A et V

(avec $X_{a0}=1$; $X_{v0}=1,1$)

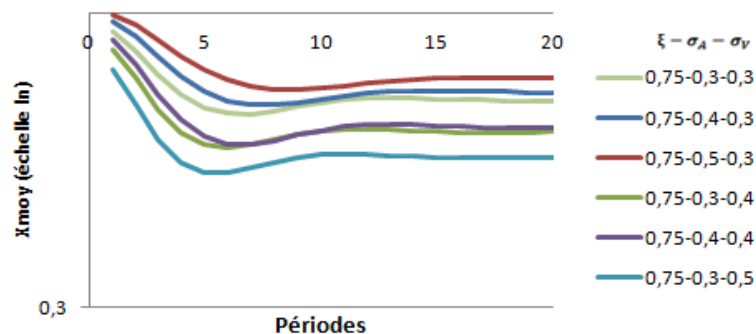


Figure 11 : effets des volatilités avec $\xi = 0,75$

On remarque immédiatement que c'est la valeur finale (ou d'équilibre) qui varie ce qui est bien conforme au résultat obtenu pour la solution particulière de l'équation différentielle. Cette dynamique répond particulièrement bien à deux contraintes de l'évolution du prix de l'immobilier vu dans le paragraphe I.2 à savoir :

- L'autoregressivité due à la continuité des équations différentielles.
- La réponse aux chocs qui peuvent même dans certains cas être amplifiée. A noter qu'ici le choc n'est pas un choc de prix mais un choc de volatilité (= variation offre-demande).
- Le prix d'équilibre du marché dépend de la volatilité de repos, soit le niveau d'offre et de demande hors situation de crise des groupes acheteurs/vendeurs.

II.2.2 Choc, retour à l'équilibre et application aux bulles spéculatives

Un autre résultat remarquable est, comme pour un système masse-ressort, un retour à l'équilibre sous forme d'oscillations. Afin de mieux comprendre la dynamique des prix

moyens on réalise une évolution rapide sur quelques périodes de la volatilité des vendeurs à la baisse. Cela pourrait être une information poussant les vendeurs à conserver leur bien. Puis un retour, après un temps d'attente, à la volatilité initiale qui peut être comprise comme un retour au niveau d'offre et de demande d'équilibre ou de base. Cela produit la réponse suivante :

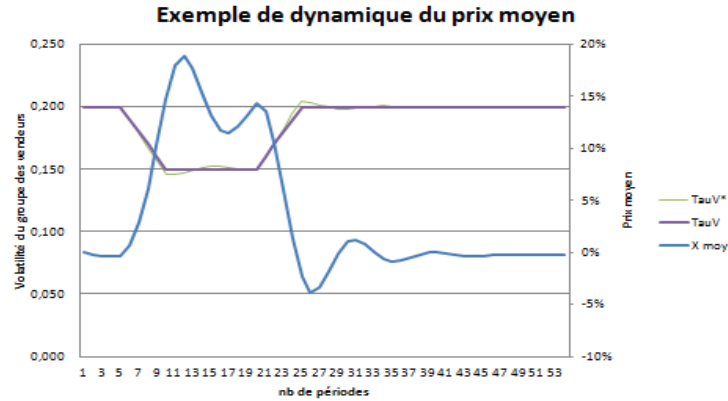


Figure 12 : exemple de choc de volatilité des vendeurs

Il est intéressant de voir qu'il est très facile de générer des bulles qui ont la particularité d'avoir deux points hauts. La condition étant qu'il y ait un temps d'attente suffisant long entre l'évolution rapide de la volatilité post-choc et le retour à la normale. Ce phénomène a été relevé fréquemment et donne donc du crédit supplémentaire à cette modélisation. On trouve par exemple sur le web le graphique humoristique suivant:

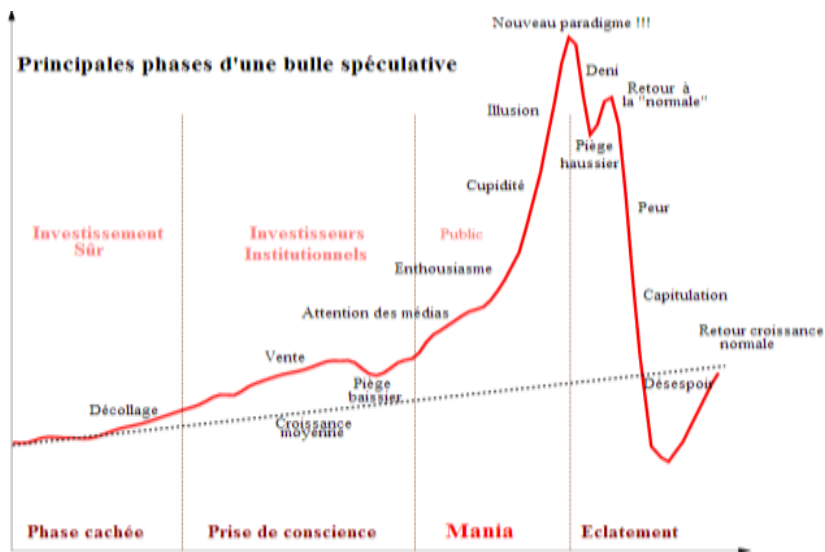


Figure 13 : graphique humoristique sur les bulles spéculatives

On retrouve également une analogie intéressante avec les travaux de Gallegati, Palestrini et Barkley-Rosser sur la PFD (Period of financial Distress) qui se définit par le fait que les cracks sont toujours précédés d'une zone d'incertitude consécutive à une forte augmentation. Ils précisent également que le début du crack (et donc la fin de la PFD) est dû aux contraintes

économiques¹⁵. Ce dernier point n'est pas en accord avec le modèle présenté ici car c'est un retour à la volatilité d'équilibre qui fait dégonfler la bulle.

Maintenant que nous avons un modèle crédible d'évolution de l'indice des prix (Formule (1), p.22), nous allons l'adapter à l'évolution de l'indice immobilier français sur les 45 dernières années.

¹⁵ Gallegati, Palestrini, Barkley-Rosser, "The Period of Financial Distress in Speculative Markets: Interacting Heterogeneous Agents and Financial Constraints", Mimeo, James Madison University

III Application au marché d'habitation français de 1970-2015

L'objectif de cette partie est de confirmer l'intérêt et la valeur opérationnelle du modèle en l'appliquant au marché immobilier français de 1970 à 2015.

III.1 Drift des acheteurs = Pouvoir d'achat immobilier

Le modèle dynamique stochastique défini précédemment fait appel à différentes variables dont les principales étant le drift (dérive) et la volatilité des acheteurs. Le drift noté μ va être représentatif de l'évolution de la moyenne des comportements des acheteurs.

Pour mémoire

$$X_{A_T} = X_{A_0} e^{\mu_A + \sigma_A U_A}$$

Avec σ_A volatilité implicite des acheteurs sur la période T et enfin U variable aléatoire normale centrée réduite $\sim \mathcal{N}(0,1)$.

III.1.1 Evolution du pouvoir d'achat immobilier

Concernant les acheteurs, le drift représente la part commune au groupe de l'évolution de la capacité d'achat immobilier. Comme nous considérerons l'évolution des salaires hors inflation, le drift représente en fait l'évolution moyenne du pouvoir d'achat immobilier.

Nous ferons ici une hypothèse majeure : La part du salaire moyen des ménages consacrée à l'immobilier est constante au cours du temps. μ_A ne dépend alors que de la variation des conditions d'emprunt (annuité et nombre d'années de remboursement) et des salaires.

L'annuité moyenne de remboursement s'écrivant

$$A = \text{Salaire moyen} * \tau_{\text{endettement}}$$

Avec A annuité et $\tau_{\text{endettement}}$ le taux d'endettement étant pris égale à 33% du salaire moyen par ménage¹⁶

Nous cherchons à modéliser l'évolution de l'indice des prix immobilier corrigés de l'inflation. Cela implique que nous devons utiliser les salaires corrigés de l'inflation. Nous utiliserons dans la suite :

$$A = \text{Salaire moyen}_{\text{corr. inflation}} * \tau_{\text{endettement}}$$

De plus la formule du crédit définit les annuités de remboursement constant comme :

$$A = \frac{K \tau}{1 - (1 + \tau)^{-n}}$$

K montant moyen de l'emprunt, τ le taux d'emprunt moyen, n le nombre d'années d'emprunt moyen.

Les remboursements constants sont largement majoritaires en France depuis les années 80 mais ce n'était pas le cas entre 1973 et 1985 car l'inflation élevée permettait d'envisager des remboursements progressifs. Le modèle ne tient pas compte de ce point toujours par souci de simplification.

¹⁶ La règle des 33% de taux d'endettement fait partie de la loi Neiertz, loi visant à limiter le surendettement qui avait été établie en 1989 mais déjà tacitement en vigueur depuis plus longtemps.

On peut donc écrire la variation du pouvoir d'achat immobilier en fonction des variations des variables sous-jacentes :

$$1 + \frac{dK}{K} = 1 + \frac{1}{K} \left(\frac{dK}{d\tau} \cdot d\tau + \frac{dK}{dA} \cdot dA + \frac{dK}{dn} \cdot dn \right) = e^{\mu_A}$$

avec

$$\begin{cases} \frac{dK}{d\tau} = A \cdot \frac{n \cdot (1 + \tau)^{-n-1} \cdot \tau - 1 + (1 + \tau)^{-n}}{\tau^2} \\ \frac{dK}{dA} = \frac{1 - (1 + \tau)^{-n}}{\tau} \\ \frac{dK}{dn} = \frac{A}{\tau} \cdot \ln(1 + \tau) \cdot (1 + \tau)^{-n} \end{cases}$$

Et l'évolution de la capacité moyenne d'achat immobilier d'une période i à $i+1$ s'écrit

$$X_{A_{i+1}} = X_{A_i} \left(1 + \frac{dK}{K_i} \right)$$

En reprenant les données historiques issues des travaux de Jacques Friggit (voir paragraphe I.2) on obtient l'évolution de la capacité moyenne d'achat immobilier comparée à l'évolution de l'indice des prix immobilier à Paris :

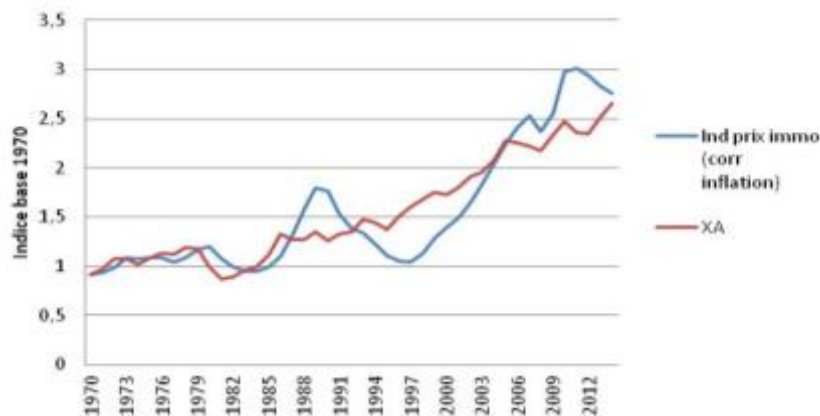


Figure 14 : Evolution historique des propositions moyennes des acheteurs

On remarque immédiatement que la capacité d'achat pour chaque période ($X_A(T)$) a évolué parallèlement à l'indice des prix immobiliers. Dit autrement, l'effort fait aujourd'hui pour acheter un type d'appartement par un acheteur est resté en tendance le même pendant ces 45 dernières années.

En effet, la capacité d'achat immobilier a fortement augmenté suite à la baisse des taux d'emprunt immobilier (de 10% dans les années 70 à 2% aujourd'hui). Cela c'est conjugué à l'allongement des prêts de 12 ans en moyenne en 1970 à 18 en 2015. Il y a d'ailleurs une relation entre ces deux variables qui ne sera pas considérée toujours pas souci de simplification.

On remarque sur ce graphique que l'indice immobilier c'est écarté à plusieurs reprises et de façons significatives du pouvoir d'achat. Cela s'est produit aussi bien à la hausse qu'à la baisse.

III.1.2 Application au modèle et résultat intermédiaire

En appliquant simplement le modèle défini dans la première partie, en utilisant $X_a(T)$ défini ci-dessus et en prenant des valeurs arbitraires mais réaliste des paramètres libres à savoir :

σ_A constant	30%
σ_V constant	18%
$\mu_{V_{négo}}$	10%
μ_{fees}	7%
ξ	1

On obtient la comparaison entre l'indice immobilier INSEE-Notaires et le modèle :

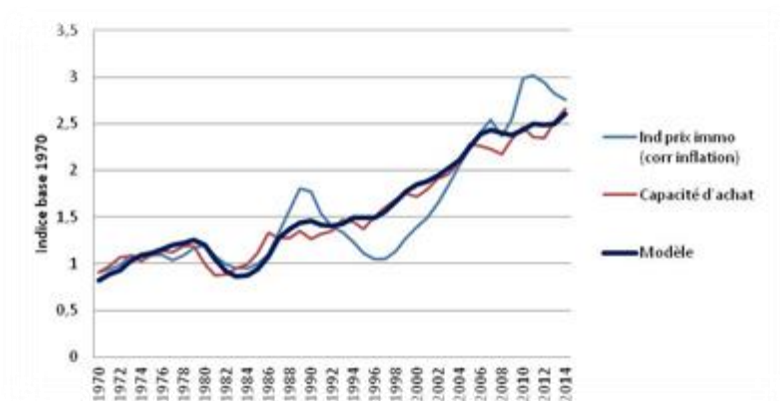


Figure 15 : 1ere application simple du modèle

Ce résultat intermédiaire est très intéressant car les variations de tendances sont captées par ce modèle. L'équation différentielle « lisse » les aspérités de la variation de la capacité moyenne d'achat. On retrouve donc bien une courbe de l'indice modélisé avec une propriété d'autoregressivité proche de celle de la courbe de l'indice INSEE.

On peut également comparer la probabilité de vente modélisée aux variations de vente réelles enregistrées sur la période.

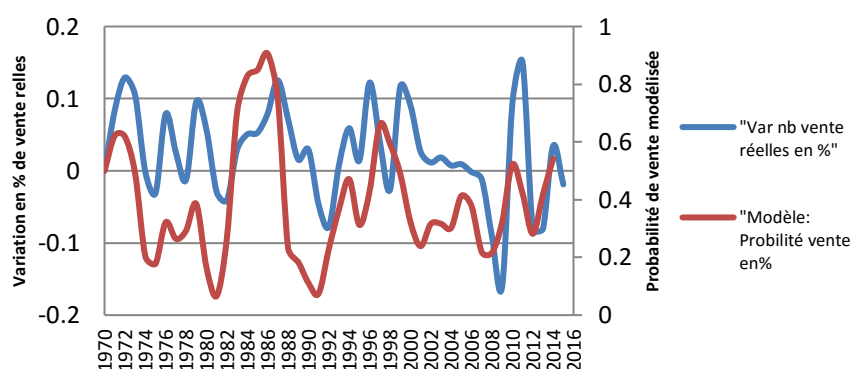


Figure 16 : Probabilité de vente du modèle vs. Variation des volumes de vente historique.

Bien que ces deux grandeurs intrinsèquement différentes et représentées sur deux axes différents ne soient pas quantitativement comparables, on peut tout de même constater une bonne similitude dans le nombre et la longueur des périodes ainsi que dans l'orientation des évolutions.

Maintenant que nous avons une estimation de l'évolution de la capacité d'achat immobilier des acheteurs, il reste à déterminer l'évolution des autres variables d'influences que sont les volatilités implicites des groupes des acheteurs et des vendeurs.

III.2 Estimation des volatilités acheteur/vendeur

Nous avons vu que les volatilités des groupes acheteurs et vendeurs devaient être vu comme une mesure de la compétition au sein du groupe. C'est-à-dire son niveau de désaccord interne. Mais l'origine de cette hétérogénéité peut-être due à plusieurs causes. Certaines viennent de l'intérieur même du groupe (endogène) et d'autres imposent au groupe un changement qui modifie sa volatilité (exogène).

Les causes endogènes peuvent être :

- D'avantage de compétition du à une plus grande densité (augmentation du nombre d'individu d'un groupe. Attention ici car c'est bien l'opposé du résultat statistique qui est postulé. En effet, en statistique, quand un groupe augmente en nombre d'individu, sa variance diminue proportionnellement à $1/\sqrt{n+dn}$ mais on considère ici que les nouveaux arrivant choisissent un sous-marché où ils sont compétitifs. Ainsi, chaque nouvel arrivant tendra à faire augmenter la dispersion du groupe. Nous respectons ainsi le fait que l'augmentation de la volatilité=compétition des acheteurs tend toujours à faire monter les prix alors que l'augmentation de la volatilité=compétition des vendeurs tend toujours à les faire baisser. C'est la loi de l'offre et la demande. On peut d'ailleurs visualiser ce phénomène en représentant les densités de probabilité des lois Log-Normale pour différentes volatilités :
- De la même façon, lorsqu'un groupe diminue en taille, sa volatilité diminue car la compétition diminue et il n'est donc plus nécessaire de s'éloigner de la moyenne pour conclure une vente (que l'on soit acheteur ou vendeur)

Les causes exogènes peuvent-être :

- Les crises incitent, selon les acheteurs, à avancer l'achat ou au contraire à repousser le projet. Cela peut se traduire de la même façon par une augmentation ou par une baisse de la volatilité.
- Les crises bancaires peuvent inciter les banques à ne plus délivrer de crédit ou à durcir les conditions. Cela empêche les acheteurs qui veulent se démarquer de le faire et diminue simplement le nombre d'acheteurs potentiels. La tendance sera donc une diminution de la volatilité du groupe d'acheteurs.

On cherche désormais une évolution des volatilités acheteur/vendeur qui permette de retrouver l'évolution de l'indice immobilier tout en étant cohérente avec l'historique économique de la période concernée et les comportements d'un groupe d'humain en général (pas de saut sans raison apparente).

On peut par exemple avec un algorithme simple d'optimisation trouver une solution notée n^0 s'ajustant bien aux données :



Figure 17 : Indice immobilier calibré solution n°0

Avec comme évolution des volatilités

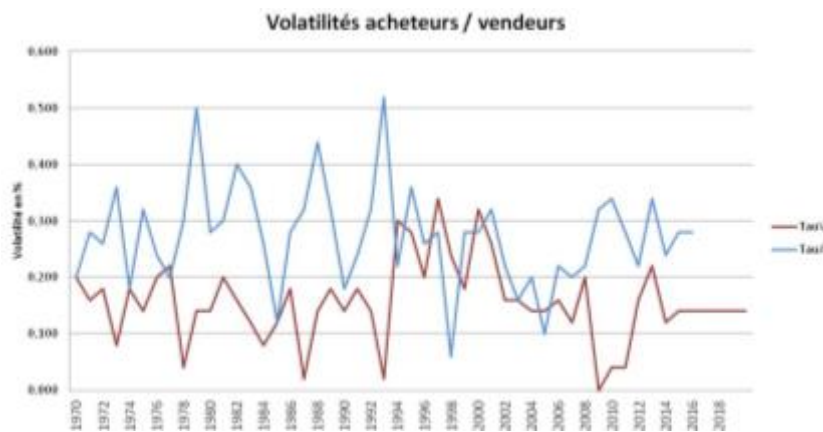


Figure 18 : Volatilités calibrées solution n°0

C'est une solution parmi une infinité car il y a pour chaque valeur de l'indice, à une année donnée, une infinité de couples $\{\sigma_A; \sigma_V\}$ qui correspondent. Cependant pour parvenir à reproduire l'indice historique, on remarque qu'il y a systématiquement des sauts de volatilités aussi bien chez les acheteurs que chez les vendeurs. On remarque également que l'algorithme n'est pas parvenu à « recoller » l'indice immobilier historique à partir de 2008. Cela signifie que les volatilités ne sont pas indépendantes d'une année à l'autre et qu'il faut calibrer des périodes et non seulement les années indépendamment les unes des autres. L'idée a donc été de relier les volatilités à des faits historiques et ainsi recréer l'indice à partir d'évènements passés qui donneraient de la cohérence ainsi qu'une valeur explicative au modèle.

III.2.1 Importance des crises

Au vue de l'étude de l'analyse multi variée (paragraphe I.2.2.1) l'indice immobilier semble d'avantage sensible aux chocs économiques qu'aux évolutions lentes des paramètres économiques classiques. C'est pourquoi il est apparu naturel de modéliser l'évolution des volatilités Acheteurs et Vendeurs comme une suite de chocs conséquences directes des chocs financier et bancaire qui se sont déroulés ces 40 dernières années.

Pour cela, j'ai retenu 4 grandes crises Française ayant pu impacter la décision des protagonistes du marché immobilier français :

- 1979 : deuxième choc pétrolier avec la peur d'une nouvelle spirale inflationniste
- 1993 : Crise monétaire et fin de la SME (System Monétaire Européen). Les différents pays européen essuient des attaques spéculatives sur leurs monnaies. Les Taux long sont devenus inférieurs aux Taux court (voir figure ci-dessous) ce qui a eu comme conséquence la réticence des banques à produire des prêts longs termes et donc un tarissement du crédit immobilier.

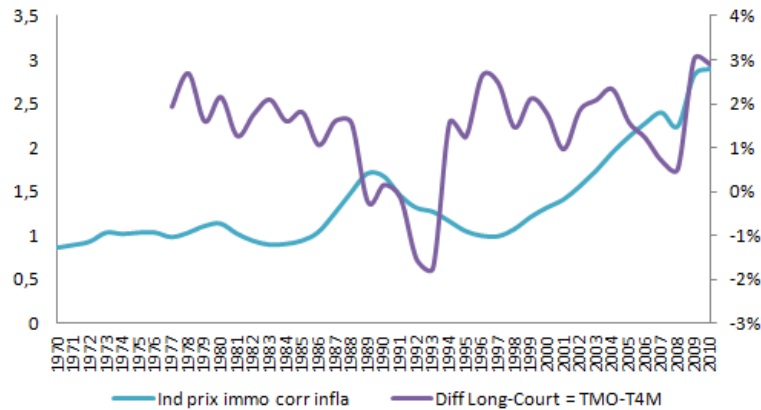


Figure 19 : Indice immo vs. (Taux longs-Taux courts)

On remarque cependant que la crise des taux a commencé avant 1993 et ce situe plutôt autour de 1991. Nous modéliserons donc un début de crise en 1991 pour les acheteurs (dépendant du crédit) et 1993 pour les vendeurs (perte de confiance et début de la couverture médiatique du « Krach immobilier français »).

- 2008 : la crise des subprimes entraîne une incertitude sur la solidité du système financier et une inquiétude pour les épargnants.
- 2010 : crise de la zone euro avec, comme principale réaction de la banque centrale européenne, la mise à disposition des banques d'argent à taux quasi-nul les incitant à produire plus de crédits et empêchant (ou ralentissant) ainsi une récession.

Ces quatre crises majeures vont donc être intégrées au modèle de volatilité des acheteurs/vendeurs.

III.2.2 Principe d'évolution des volatilités Acheteur/Vendeur

Pour chaque groupe, chaque crise va être modélisée par un trapèze modélisant le niveau de base (ou d'équilibre), la montée, le plateau (période d'attente) et la redescende vers le niveau d'équilibre.

Ce qui donne par exemple pour une crise les paramètres suivants :

Années	Tau A base	force	Attaque (nb an)	Traine (nb an)	durée plateau
1979	18%	6%	1	2	1

Ce qui se traduit par la courbe d'évolution de la volatilité (ici Tau A : acheteur) suivante :

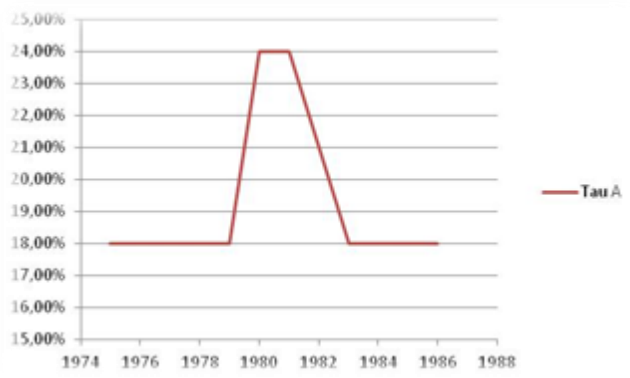


Figure 20 : Exemple de choc de σA en 1979

A titre d'exemple et pour l'ensemble des quatre chocs on obtient la figure suivante :

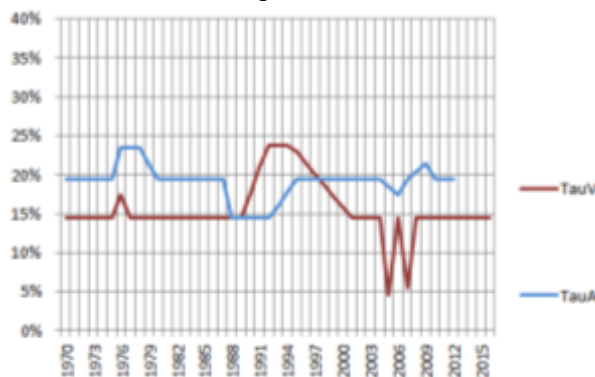


Figure 21 : Exemple des volatilités sur 40 ans à travers les 4 crises

Il est donc possible maintenant de chercher l'ensemble des crises ainsi modélisées permettant de retrouver l'évolution de l'indice immobilier de 1970 à aujourd'hui.

III.2.3 Calibrage des crises de volatilités Acheteurs/Vendeurs

Nous cherchons à trouver les formes des crises qui permettent d'expliquer l'évolution de l'indice immobilier. Nous avons cependant vu qu'il existe une infinité de solution et j'ai pu rapidement constater qu'il y a beaucoup de minimum locaux qui empêche la progression de proche en proche des différentes variables pour converger vers l'optimum. J'ai créé un algorithme qui fait évoluer les variables tant que cela améliore l'erreur entre la courbe modélisée et l'indice historique (critère des moindres carrés sur l'ensemble des années). Cela m'a permis, en partant de différents ensembles de départ, de retenir 2 solutions présentées ci-dessous.

Solution n°1 : On ne considère que 3 crises. 2008 et 2010 sont considérées comme un unique évènement du point de vu du marché immobilier. σA et V « bases » sont respectivement de 30% et 18%. Les crises de volatilités optimisées sont alors :

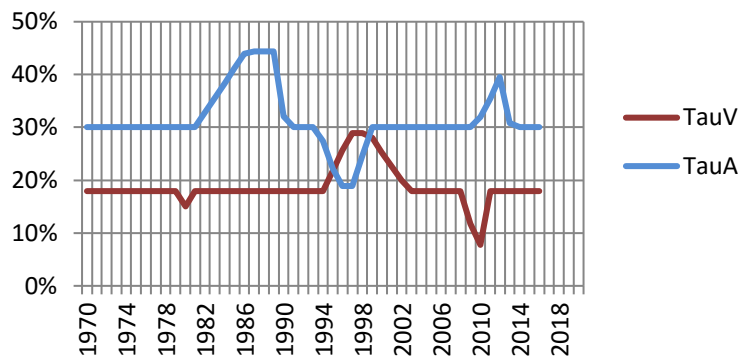


Figure 22 : Volatilités calibrées solution n°1

Et on obtient l'évolution de l'indice immobilier modélisé suivant :

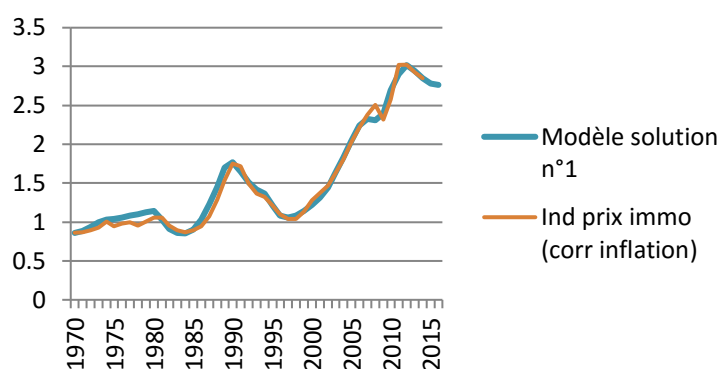


Figure 23 : Indice immobilier vs. modèle solution n°1

Nous pouvons donc, à partir de cette solution optimisée, déduire le scénario correspondant en terme de comportement des groupes acheteurs et vendeurs. Le scénario de la solution n°1 pourrait alors être : la crise pétrolière de 1979 impacte faiblement les vendeurs. Cependant, cela occasionne une amorce de remontée des prix qui, dans un contexte tendu va faire naître une bulle spéculative. En effet, la montée des prix n'est entretenue que par compétition entre acheteurs. Cette phase va durer une décennie. Par la suite, la crise bancaire à partir de 1991 rend difficile l'accès au crédit pour les acheteurs ce qui en diminue le nombre et donc la volatilité. Cela s'améliorera à partir de 1996. Concernant les vendeurs, à partir de 1993, la menace d'un crack immobilier rend incertain les vendeurs et augmente la compétition (leur volatilité) qui va augmenter jusque début 2000 et la bulle internet. En 2008, la crise est brutale et les vendeurs se retirent du marché alors que beaucoup d'acheteurs veulent « investir dans la pierre à tous prix ». A partir de 2012, le marché retrouve la stabilité, les prix reprennent l'oscillation vue dans le chapitre II.2 et commence par une hausse des prix (situation à laquelle nous assistons aujourd'hui).

Concernant l'évolution du nombre de ventes enregistrées comparée à la probabilité de vente issue du modèle :

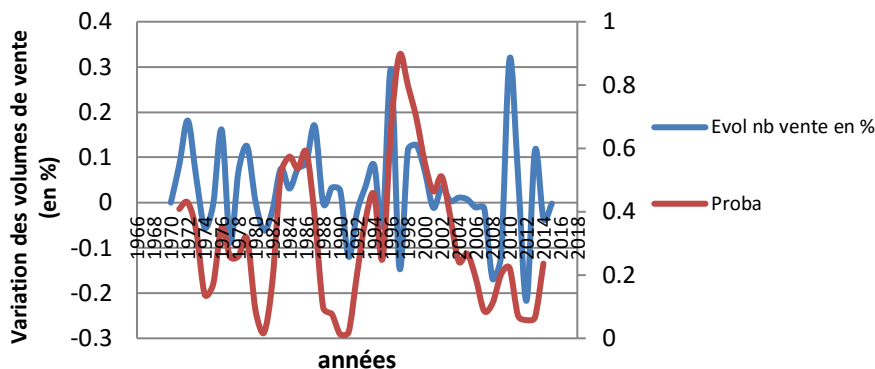


Figure 24 : Probabilité de vente du modèle vs. Variation des volumes historiques : solution n°1

Les nombres de creux et de pics sont de nouveau en majorité respectés. Cependant, il est clair que le modèle ne capte pas l'ensemble des variations. On remarque notamment de réelles incohérences lorsque le modèle détermine une hausse de la probabilité de vente alors qu'il a été enregistré nettement moins de ventes (par exemple 1997 et 2010). Cela peut-être dû à d'autres phénomènes non pris en compte dans cette modélisation comme les aides à l'accession par exemple. Nous pouvons également générer d'autres solutions en modifiant par exemple le nombre de crises.

Solution n°2 : 4 crises dont 2 crises en 2008 et 2010. σA et V « bases » sont optimisés par l'algorithme.

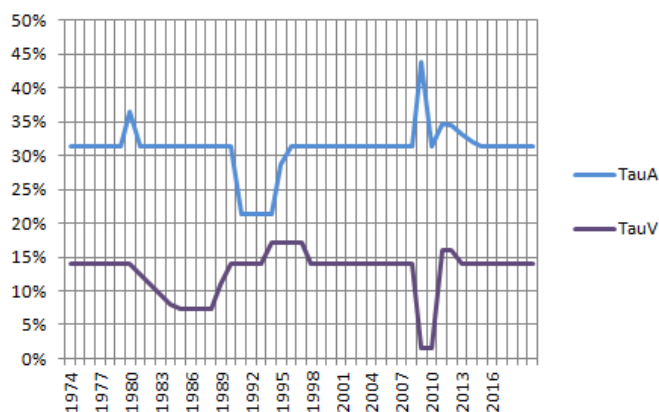


Figure 25 : Volatilités calibrées solution n°2

Ce qui donne pour l'évolution de l'indice de prix modélisé :

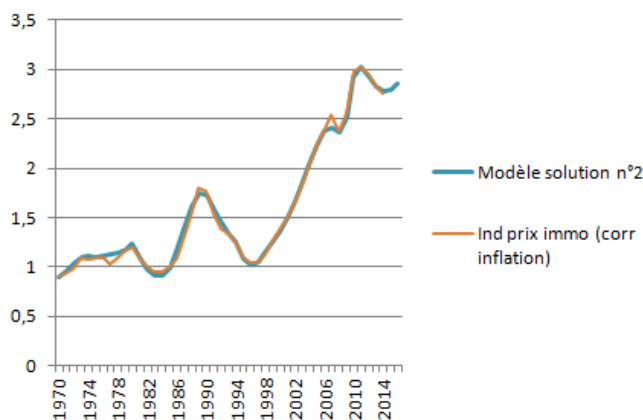


Figure 26 : Indice immobilier vs. modèle solution n°2

Et pour l'évolution du nombre de vente comparée à la probabilité de vente du modèle

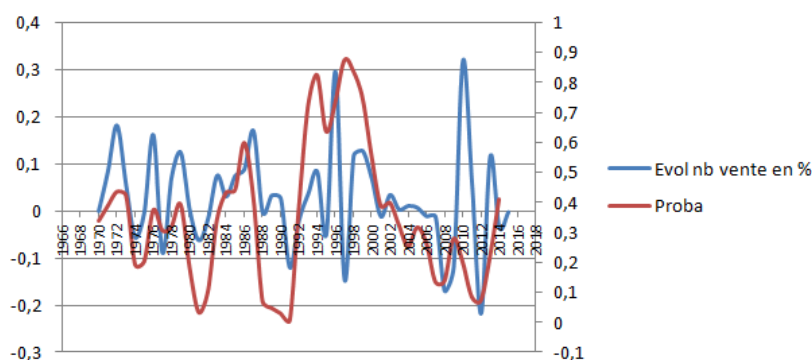


Figure 27 : Probabilité de vente du modèle vs. Variation des volumes historiques solution n°2

L'explication pourrait être alors la suivante :

Certains acheteurs ont voulu rapidement acquérir un bien pendant le choc pétrolier de 1979 anticipant une hausse rapide de taux qui a bien eu lieu. Les vendeurs ont eux voulu conserver leurs biens car ils protégeaient leur patrimoine de l'inflation qu'ils venaient de connaître dans les années 1970. C'est donc une phase longue de conservation pour les vendeurs qui s'est enclenchée à l'occasion du choc pétrolier. De plus, la population est jeune et la demande de biens est en croissance rapide (babyboomers en âge du premier achat immobilier).

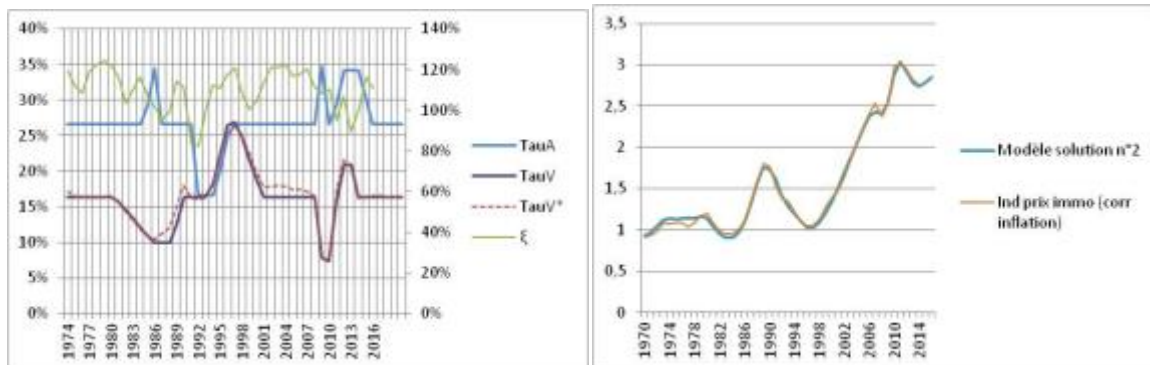
Dès 1990, les acheteurs ont vu l'accès au crédit se durcir pour de nombreuses années et ont également choisi de repousser leur achat en attendant la fin de la baisse. Les vendeurs ont du « courir » après les acheteurs pour réaliser les ventes. De même que pour la solution n°1, cette phase va durer une décennie.

En 2008, et à la manière de 1979, les acheteurs vont rapidement augmenter leur volatilité pour aller chercher un bien en se mettant d'avantage en concurrence. Les vendeurs vont vouloir conserver leurs biens et ne vendre qu'aux prix les plus avantageux et ainsi baisser considérablement la volatilité du groupe. Cependant, à partir de 2010, ils vont concéder un léger effort pour finaliser les ventes car certains redoutent la montée des taux d'intérêts. Toujours à partir de 2010, les acheteurs vont voir progressivement une offre de crédit abondante qui va leur permettre de faire des offres plus compétitives. C'est encore le cas aujourd'hui mais l'impact sur les acheteurs de l'abondance de crédit à bas coût est en train de s'estomper.

Ce dernier scénario est pour moi le plus crédible. Bien qu'il soit possible d'en générer d'autres, ils reprennent tous en général ces grandes lignes. Il est évidemment possible d'ajouter des crises et de ce fait trouver des scénarios encore plus proches de la réalité mais j'ai souhaité avant tout utiliser le moins de variables possible (et donc de crises) afin de garder un pouvoir explicatif simple et direct.

Pour obtenir ces résultats, nous avons dû faire certaines hypothèses sur des paramètres non intuitifs qui sont difficiles à évaluer. Il est donc important de s'assurer que la sensibilité du modèle à ces paramètres soit faible. Cette étude de sensibilité est en annexe 4 mais je présente ici le principal résultat. En effet, on a considéré dans ces scénarios $\xi=1$ mais il est peu probable que l'anticipation des vendeurs ne change pas au cours du temps en fonction des crises et des époques. Pour cela, j'ai tenté de créer une fonction représentant le fait que lorsque l'évolution annuelle des prix se reproduit sur plusieurs périodes, la force de l'anticipation ξ augmente et lorsque les variations annuelles varient fortement, l'incertitude

grandie et ξ baisse. De plus, depuis le début de l'étude du calibrage sur l'indice immobilier, l'hypothèse a été faite qu'il n'y avait pas de corrélation entre les volatilités acheteurs et vendeurs. Afin de s'assurer que ces hypothèses soient justifiées, j'ai refait l'optimisation de la solution 2 avec cette fois le paramètre $\rho=20\%$ et ξ variable (voir annexe 4).



Les résultats de l'optimisation ne changent pas qualitativement et il est intéressant de constater que la solution est robuste vis-à-vis des modifications des paramètres. Cela laisse entendre que beaucoup de scénarios différents auraient conduit à l'évolution des prix de l'immobilier français ou à des courbes très similaires. C'est un autre argument de poids en faveur de la bonne représentativité de cette modélisation.

Nous disposons donc à ce stade d'un outil puissant de modélisation du marché immobilier (et de tout autre marché non substituable). Ce modèle va ainsi permettre de créer des scénarios réalistes de l'évolution future de l'indice des prix (Générateur de Scénario Economique). Il permet également d'avoir une estimation du prix d'équilibre du marché et c'est ce que nous allons voir dans le paragraphe suivant.

III.3 « Effet tunnel » et prévision à court terme

Avant d'intégrer le modèle dans un générateur de scénarios économiques, voyons quels sont les conséquences directes de son utilisation.

III.3.1 « Effet tunnel »

Nous avons vu dans le paragraphe II.1.3.3 que l'indice des prix après stabilisation des oscillations (et toutes choses égales par ailleurs) pouvait s'écrire :

$$\lim_{T \rightarrow \infty} \ln(\tilde{X}_{V_T}) = \frac{\frac{\sigma_V^2}{2} + \mu_{V_{négo}} + \beta}{\alpha} + \ln(X_{A_0})$$

Et que si l'on linéarise α et β en fonction de σ_V et σ_A (Cf annexe 5), on peut écrire :

$$\lim_{T \rightarrow \infty} \ln(\tilde{X}_{V_T}) = \frac{0,5 \cdot \sigma_V^2 - 1,352 \cdot \sigma_V + 0,553 \cdot \sigma_A - 0,048 + \mu_{V_{négo}}}{2,343 \cdot \sigma_V - 1,393 \cdot \sigma_A + 0,229} + \ln(X_{A_0})$$

Valable pour $25\% < \sigma_A < 31\%$ et $14\% < \sigma_V < 21\%$ (Erreur inférieur à 10% sur α et β)

Ce résultat est intéressant car il permet d'établir que l'indice des prix d'équilibre corrigé de l'inflation après un temps long dépend uniquement du pouvoir d'achat immobilier et des volatilités de bases (ou d'équilibre) des acheteurs/vendeurs à la date du marché.

On écrit ainsi le prix d'équilibre « net vendeur » :

$$P_{Equ_i} = X_{A_i} \cdot e^{\frac{\frac{\sigma_{V_{base}}^2}{2} + \mu_{V_{négo}} + \beta}{\alpha}} \cdot e^{-\mu_{fees}}$$

Les frais étant retirés du prix de vente moyen.

Ce qui s'interprète comme étant le produit d'un terme de capacité d'achat moyen des acheteurs (X_A) multiplié par un terme de tension de marché long terme (σ_V et σ_A base)

Il faut d'ailleurs préciser ici que dans toutes les optimisations $\sigma_{V_{base}} < \sigma_{A_{base}}$. Cela peut être expliqué par le fait que les vendeurs sont aiguillés par des professionnels (les agents immobiliers) qui tendent à réduire l'aléa d'appréciation personnelle de la valeur d'un bien. Une deuxième explication possible tiendrait au fait que contrairement aux vendeurs, les acheteurs peuvent changer de type de bien en passant par exemple de l'achat d'un T4 à un T3 ce qui leur permet d'augmenter la distance de leur proposition à la moyenne des T3 (offre plus compétitive) et qui augmente de fait la volatilité du groupe des acheteurs.

On peut donc postuler que le ratio des prix historiques par les prix d'équilibres va osciller autour de un avec une probabilité de se trouver à une certaine distance diminuant rapidement avec cette dernière. C'est un effet tunnel. Cette idée a déjà été proposée pour le marché immobilier français par Jacques Friggit. En effet, la courbe de Friggit est le nom communément donné à la courbe représentant le ratio de l'indice du prix des logements rapporté au revenu par ménage. Le tunnel de Friggit est la description imagée d'une régularité de cette courbe. De 1965 à 2000, la courbe est ainsi restée constante à 10 % près, donnant ainsi l'image d'un tunnel horizontal (voir graphique ci-dessous).



Figure 29 : Tunnel de Friggit

Or, on remarque immédiatement que depuis les années 2000, le ratio « Indice des prix immobilier / revenu par ménage » a fortement divergé ce qui est contre l'idée du tunnel. Et pour cause, les taux d'intérêt ont fortement baissés et ont permis aux acheteurs d'augmenter leur pouvoir d'achat immobilier alors que leur salaire n'a pas augmenté en proportion.

Je propose ici une évolution du travail de Jacques Friggit. A la différence de la courbe de Friggit, je propose de créer le ratio de « l'indice immobilier / indice du prix d'équilibre net vendeur ».

Pour rappel du paragraphe III.1.1, l'évolution de la capacité d'achat immobilier des acheteurs permet de prendre en compte non seulement l'évolution des salaires net d'inflation mais aussi les variations de taux d'intérêts des prêts bancaires et la variation des durées d'emprunts (liées au taux mais aussi à l'allongement de la période de travail et aux pratiques bancaires).

S'il on trace le ratio de l'indice des prix immobiliers à Paris corrigés de l'inflation par l'indice des prix d'équilibre on obtient :

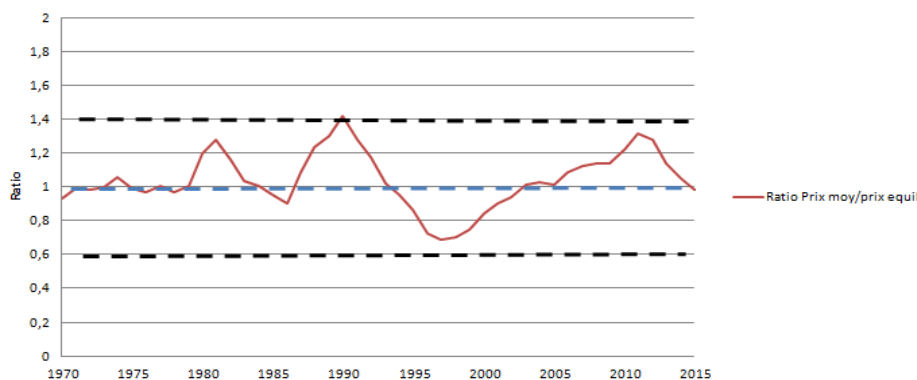


Figure 30 : Tunnel de l'indice immobilier historique / prix d'équilibre

On observe bien un tunnel sur toute la période étudiée. On remarque également que l'on peut y voir les différentes crises avec des cycles de développement d'environ 5 à 10 ans.

Ce tunnel signifie que le prix de l'immobilier est contraint par le taux d'effort (part de l'immobilier dans le budget). Rappelons qu'en France s'applique la règle des 33%. Cette règle des 33% de taux d'endettement fait partie de la loi Neiertz, une loi visant à limiter le surendettement. Etablie en 1989, on peut cependant supposer qu'elle s'applique depuis bien plus longtemps.

C'est donc cette règle qui permet de rester dans le tunnel. D'ailleurs, si elle devait être levée, cela créerait une période transitoire avant de retrouver un nouvel équilibre avec des prix plus élevés (par exemple 50% de taux d'endettement) et un nombre plus important de défauts.

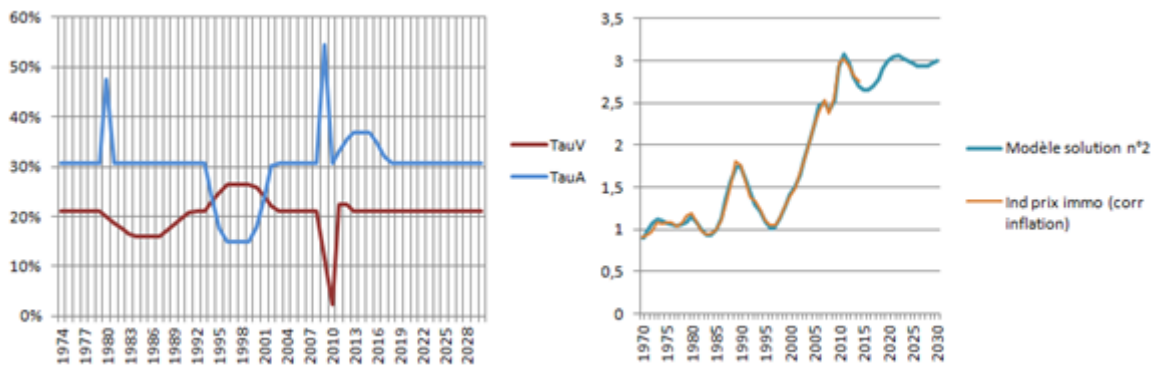
Cela n'apporterait qu'au vendeur et ferait reposer les nouveaux risques d'impayés sur les banques qui les reporteraient sur les taux d'assurances (et donc l'ensemble des acheteurs). En toute logique, il n'est pas souhaitable de faire évoluer cette règle. Cependant une nouvelle règle de solvabilité calculée sur la valeur d'un bien immobilier possédé tel que le système en vigueur aux Etats Unis pourrait arriver prochainement en France. Cela va créer une nouvelle boucle de rétroaction qui, à la vue du modèle, fera augmenter l'amplitude des cycles du marché.

Avant d'intégrer le modèle dans un GSE, voyons si le modèle permet de faire des prédictions court terme cohérentes du marché.

III.3.2 Essai de prévision à court terme

Le modèle ayant une forte dimension explicative peut-il également être prédictif ?

Le modèle est calibré jusqu'à l'année 2014. S'il l'on fait l'hypothèse (en 2014) d'une baisse des taux d'intérêts à 1,5% et leur maintien jusqu'en 2030, on obtient l'évolution modélisée suivante :



Les Taux d'emprunt ont en effet continué à baisser en 2016 pour atteindre 1,45% en moyenne pour une durée d'emprunt de 20 ans. Bien qu'une continuation de ces taux jusqu'en 2030 paraisse encore peu probable aujourd'hui, cela pourrait arriver dans le cas d'une Europe en récession lente (mais sans choc) qui ne pourrait pas remonter ces taux au risque de créer une grave récession. Vu qu'il n'y a pas de mouvement brusque, il n'y aurait pas d'anticipation des acteurs.

On retrouve bien dans ce cas la forte montée des prix constatée en 2016 (+5% à Paris). Le modèle prévoit ensuite des oscillations pour converger vers le prix d'équilibre. Dans ce cas, les prix (hors inflation) retrouveraient leurs niveaux de 2010.

Il est ainsi possible d'utiliser le modèle pour prédire des scénarios futurs possibles. On peut générer ces futurs à partir de deux groupes de variables :

- Les variables constituant la capacité d'achat immobilier (taux d'emprunt, salaires,...)
- Les volatilités implicites qui représentent les crises

En générant ces deux groupes de variables on peut générer des futurs de l'indice des prix immobilier et en étudier la répartition en terme de probabilité d'occurrence. Il est donc possible d'intégrer ce modèle un Générateur de Scénario Economique (GSE).

IV GSE (Générateur de Scénarios Economiques)

Nous allons voir dans cette partie comment le modèle définit précédemment peut-être intégré à un générateur de scénarios économiques

IV.1 Application du modèle dynamique stochastique de l'immobilier à la formation d'un GSE simplifié

IV.1.1 Principe général et simplification

La réalisation d'un GSE demande beaucoup d'attention et d'étude quant au paramétrage des différentes variables. Il existe des modèles de GSE dit intégrés faisant consensus quand à leur cohérence. On peut citer les modèles de Wilkie, Brennan et Xia et enfin le modèle d'Ahlgrim qui sera présenté ici.

Mais avant cela, je vais présenter un modèle simplifié afin d'illustrer les résultats du modèle dynamique stochastique utilisé dans le cadre d'un générateur de scénario.

Par souci de simplification, j'ai choisi d'utiliser un raccourci permettant de générer simplement les crises à partir des 4 crises historiques retenues dans le calibrage du modèle. Je considère que les 4 chocs de volatilités de la solution n°2 (voir §III.2.3) sont des crises types et sont susceptibles de se reproduire dans le futur. J'évite ainsi le tirage aléatoire des crises en amplitude et en forme (voir §III.2.1) des volatilités des acteurs. Il ne reste donc que le tirage des années de crise ainsi que l'évolution des taux d'intérêts.

IV.1.2 Tirages et résultats intermédiaires

L'évolution des taux d'intérêt d'emprunt est modélisée par un processus Ornstein-Uhlenbeck¹⁷ dont l'objectif pourra varier selon les modèles de GSE. Concernant les crises de volatilité (=de confiance) choisissons arbitrairement qu'il n'y aura que 4 crises entre 2017 et 2050. Soit une crise tous les 8 ans en moyenne. Leurs distributions dans le temps seront générées uniformément. C'est-à-dire que toutes les années entre 2017 et 2050 vont être tirées avec la même probabilité et indépendamment de la proximité d'une autre crise. Ce modèle n'est évidemment pas le plus réaliste mais il a le mérite d'être simple et intuitif quand aux résultats.

Nous pouvons alors tirer des scénarios (3000 dans ce cas) comportant à chaque fois un déroulement des taux d'intérêts ainsi qu'une succession de 4 crises entre 2017 et 2050 aléatoirement réparties (les crises de volatilité, en forme et puissance, sont les 4 crises définies dans la solution n°2, §III.2.3).

Ces 3000 scénarios permettent de définir des courbes de VaR (Value at Risk) représentant le % de chance que l'indice immobilier soit, une année donnée, en dessous de la courbe. Soit le graphique suivant :

¹⁷ Wikipedia : https://en.wikipedia.org/wiki/Ornstein-Uhlenbeck_process

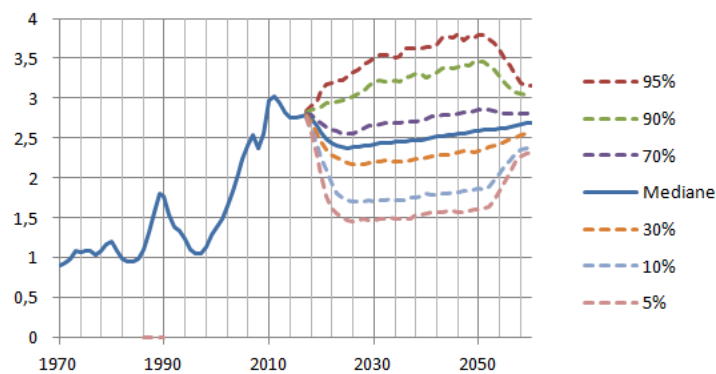


Figure 32 : VaR de l'indice immobilier, Taux objectif des taux d'emprunt à 4%

Le résultat est celui qui pouvait être attendu à savoir :

- Une baisse de l'indice due à l'augmentation des taux
- Puis une reprise de la hausse due à l'augmentation des salaires hors inflation
- Une convergence des VaR après 2050 car il n'y a plus de crises modélisées

Si l'on regarde la VaR des courbes de ratio indice des prix / indice du prix d'équilibre

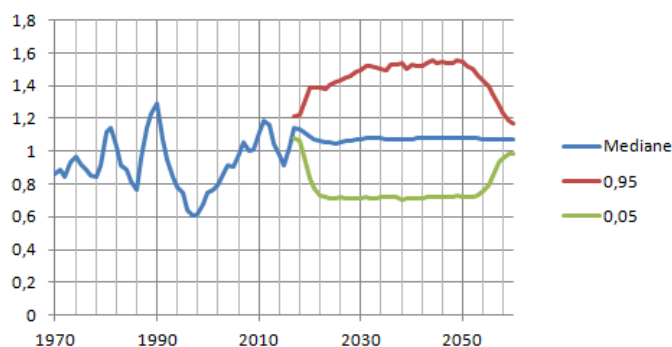


Figure 33 : VaR du ratio Prix moyen / Prix d'équilibre : Tunnel

On peut constater que:

- un tunnel est bien respecté
- la médiane (50% de chance d'être en dessous de la courbe) est légèrement au dessus la valeur 1. L'équation différentielle exacte (non celle linéarisée) qui est modélisée dans le tableur est à coefficients non constants. Il y a donc une erreur sur le prix d'équilibre particulièrement lorsque les volatilités sont élevées (en période de crise). Mais on voit que cette erreur reste acceptable.

Il faut tout de même rappeler que les crises de volatilités (en terme de forme et puissance) sont celles issues du recalage de la solution n°2 (voir §III.2.3). Elles peuvent probablement être plus fortes, ainsi les résultats présentés sous-estiment les VaR. Cependant cela ne retire rien à la pertinence qualitative des résultats. Désormais muni de cet outil, il est possible d'étudier l'influence d'autres variables qui ont été considérées comme constante précédemment.

Voyons par exemple l'impact d'un changement d'objectif de taux à partir de 2030. Ce dernier passerait alors de 4% à 8% dans le processus Ornstein-Uhlenbeck. On pourrait par exemple imaginer qu'une nouvelle crise des dettes souveraines pousse la zone euro à laisser courir l'inflation afin de faciliter le remboursement des dettes ou encore une sortie de l'euro

avec une nouvelle monnaie inflationniste (l'indice immobilier présenté ici étant net d'inflation). Dans ce cas, le graphique des VaR serait :

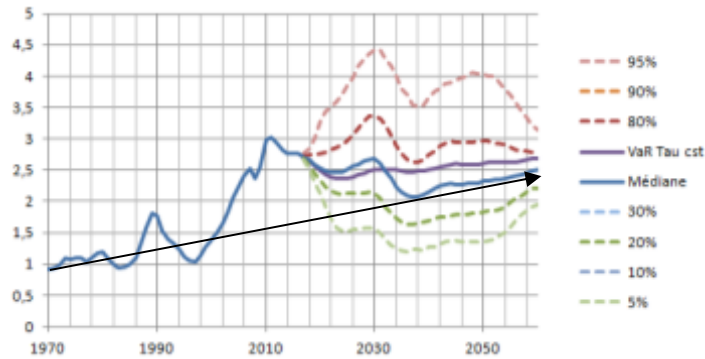


Figure 34 : VaR pour une modification de l'objectif des taux à partir de 2030

On remarque dans ce cas une période de transition à la baisse pendant une dizaine d'année avant de reprendre une croissance identique à ce qu'elle était auparavant. Les taux d'emprunts ayant retrouvés ceux des années 1970, seule la croissance des salaires nets d'inflation (ici imposée à 1% par an) fait croître l'indice immobilier (flèche noire sur la Figure 34).

Les résultats préliminaires semblent cohérents et encourageant pour l'utilisation du modèle immobilier présenté dans un GSE exploitable par les professionnels. Pour s'en assurer, il faut le comparer à un GSE « standard » en l'intégrant dans une structure d'actifs. Je vais utiliser pour cela une structure qui comporte l'actif immobilier et qui fait référence dans le milieu actuariel à savoir le modèle d'Ahlgrim.

IV.2 Comparaison du modèle immobilier à celui utilisé dans un GSE type Ahlgrim

IV.2.1 Intégration du modèle dynamique stochastique pour l'immobilier

La réalisation du GSE type « Ahlgrim » est entièrement reprise du travail de F.Planchet, P.Théron et A.Kamega qui ont mis à disposition un GSE calibré en langage R¹⁸. Les modèles mathématiques d'actifs utilisés sont en annexe 1. De plus et par chance, le GSE d'Ahlgrim calibré par Planchet utilise les mêmes données que pour mon étude (à savoir les données de J.Friggit) et ce calibrage sur les données de 1955 à 2005 produit les résultats suivant :

Matrice de corrélation des résidus :

	Tx_Inflation	Rdt_Immobilier	Tx_reel_long	Tx_reel_court	Rdt_exces_actions
Tx_Inflation	1,0000	0,2019	-0,8576	-0,6514	-0,3242
Rdt_Immobilier	0,2019	1,0000	-0,3213	-0,2688	0,1034
Tx_reel_long	-0,8576	-0,3213	1,0000	0,8019	0,1761
Tx_reel_court	-0,6514	-0,2688	0,8019	1,0000	0,1120
Rdt_exces_actions	-0,3242	0,1034	0,1761	0,1120	1,0000

¹⁸ <http://www.ressources-actuarielles.net/r>

La formule discrète utilisée pour le développement des scénarios et utilisant le modèle Ornstein-Uhlenbeck est la suivante :

$$\frac{\tilde{X}_{i+1}}{\tilde{X}_i} = e^{\ln\left(\frac{\tilde{X}_i}{\tilde{X}_{i-1}}\right) \cdot e^{-k_{Im}} + \mu_{Im} \cdot (1 - e^{-k_{Im}}) + \sigma_{Im} \cdot \sqrt{\frac{1 - e^{-2 \cdot k_{Im}}}{2 \cdot k_{Im}}} \cdot \varepsilon_{Im}}$$

Avec :

- \tilde{X}_i : la valeur de l'indice des prix immobiliers l'année i
- k_{Im} : la vitesse de retour à la moyenne
- μ_{Im} : le taux de rendement continu moyen issu de la variation de l'indice des prix immobiliers
- σ_{Im} : l'écart type de l'erreur
- ε_{Im} : un nombre aléatoire avec une distribution normale centrée réduite

Dont le calibrage sur les données de J.Friggit est :

Paramètres du processus. Ornstein-Uhlenbeck pour l'immobilier :

Vitesse de retour à la moyenne	0,152
Taux de rendement réel moyen	0,094
Ecart type de l'erreur	0,037

Ainsi muni du GSE type Ahlgrim calibré, on peut constituer une base de scénarios (10 000 ici) sur 30 ans d'évolution du taux d'intérêt nominal et de l'immobilier (modèle Ahlgrim). Le modèle dynamique stochastique de l'immobilier nécessite, de son côté, de constituer sur le même horizon une base de volatilités acheteur et vendeur comme définie dans le paragraphe IV.1.1. Il faut enfin que les deux modèles aient la même tendance d'évolution à long terme. C'est dire, le taux de rendement réel moyen pour la modélisation Ornstein-Uhlenbeck et l'augmentation des salaires hors inflation pour le modèle dynamique-stochastique. Ils sont tout deux fixés arbitrairement à 1 % dans le cadre de cette comparaison. On obtient alors le graphique des rendements moyens et quantiles 10% et 90% suivant :

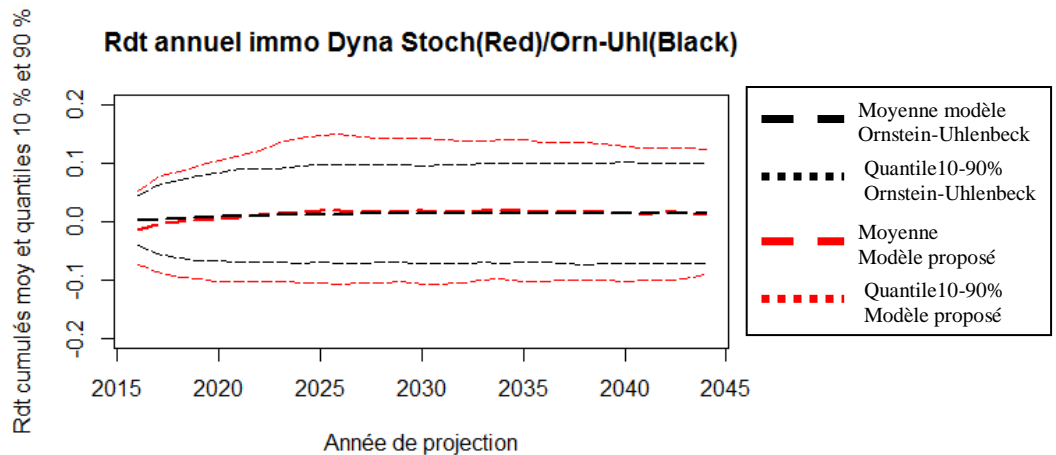


Figure 35 : VaR comparées des rendements annuels des 2 méthodes d'évolution des prix immobiliers (Dynamique Stochastique / Ornstein Uhlenbeck)

Cela montre que les deux modèles ont bien le même taux de croissance moyen annuel de 1% mais que la volatilité des rendements annuels de l'indice des prix de l'immobilier est plus importante sur le modèle dynamique-stochastique que sur le modèle Ahlgrim alors même que l'on a vu que la volatilité du modèle dynamique-stochastique était sous-estimée. Cela ferait donc penser que le modèle présenté ici n'apporte pas de gain dans l'utilisation d'un GSE. Cependant, si l'on regarde les rendements cumulés, on obtient le graphique suivant :

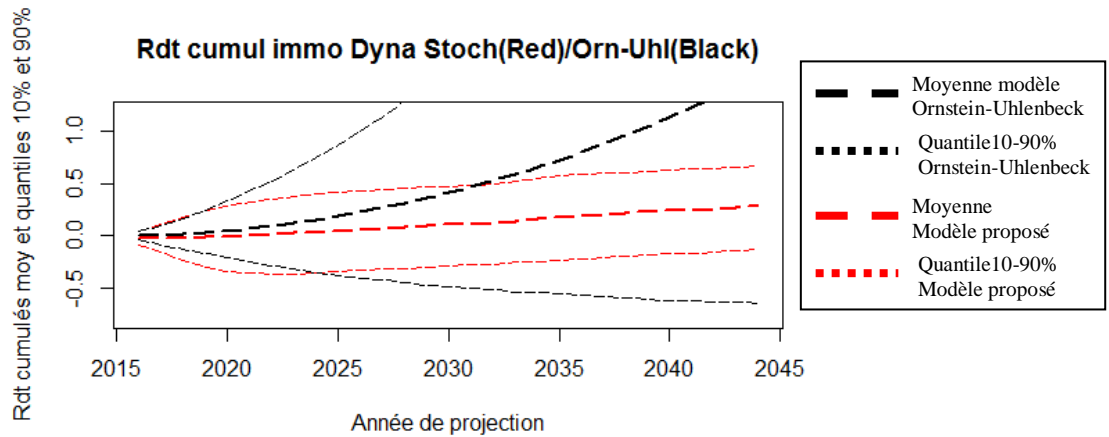


Figure 36 : VaR comparées des rendements cumulés des 2 méthodes (Dynamique Stochastique / Ornstein Uhlenbeck)

Ce résultat est extrêmement intéressant car bien que le rendement annuel soit plus incertain dans le modèle « dynamique-stochastique », le rendement cumulé fait apparaître une plus grande stabilité que le modèle Ornstein-Uhlenbeck du GSE type Ahlgrim. En effet, dans le processus Ornstein-Uhlenbeck, le rendement annuel, lorsqu'il s'écarte, est ramené à la moyenne à l'aide du coefficient de « vitesse de retour » alors que dans le modèle « dynamique-stochastique », c'est directement le rendement cumulé qui est ramené au niveau moyen du pouvoir d'achat des acheteurs. Cela laisse penser que lorsque l'on cherche à évaluer le rendement d'un placement futur >1 an (rendement cumulé de plusieurs périodes), le modèle classique surévaluerait fortement ce dernier avec une marge d'incertitude trop importante (VaR 10% et 90% trop éloignées). Le modèle « dynamique-stochastique » donnerait une valeur plus réaliste des rendements futurs car l'évolution de l'indice des prix est contrainte par la possibilité financière des acheteurs. De plus, l'incertitude (= écartement des quantiles) est plus faible dans le modèle proposé car les oscillations contraignent l'indice des prix à l'intérieur « d'un tunnel » suivant le pouvoir d'achat immobilier (voir §III.3).

Rappelons que les volatilités des acheteurs et des vendeurs sont pour chaque scénario une répétition aléatoire des 4 crises modélisées dans la solution 2 (voir §III.2.3). On peut penser que ces crises ne sont pas extrêmes et sous évaluent l'incertitude autour de la moyenne. Il faudrait en effet tirer des crises plus importantes. Cela, bien qu'augmentant la zone de VaR (quantiles 10%-90%), ne modifierait pas l'avantage du modèle dynamique-stochastique sur le modèle standard de type Ahlgrim. L'évolution de l'indice immobilier est accrochée au pouvoir d'achat des acheteurs et ne peut pas, même en cas de forte crise, s'en éloigner très longtemps. C'est la force de ce modèle et c'est pour une vision long terme des rendements qu'il trouve tout son intérêt.

Si l'on veut enrichir le modèle dynamique stochastique, on peut créer des modèles d'évolution pour des autres variables constitutives. Par exemple, on peut créer un modèle

d'évolution de la durée de l'emprunt moyen et utilisant à nouveau le modèle Ornstein-Uhlenbeck et en calibrant sur les données historiques. En incluant ces nouveaux scénarios on produit le résultat suivant :

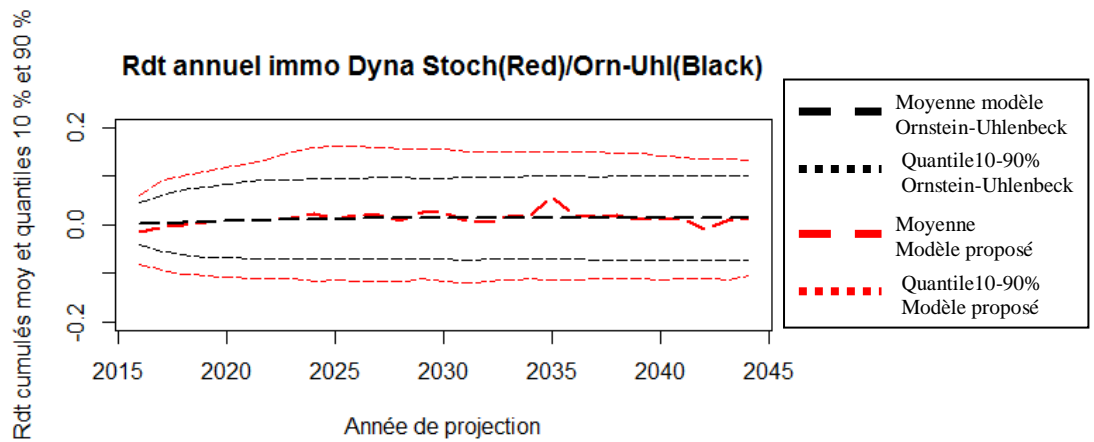


Figure 37 : Comparaison des rendements annuels avec scénarios sur la durée d'emprunt

La zone entre les quantiles 10% et 90% s'est « épaissie ». En effet, en augmentant le niveau de variation dans le modèle on augmente la volatilité des résultats. On remarque également qu'une instabilité commence à apparaître. Cela indique que le nombre de scénarios (10 000) est insuffisant pour lisser les résultats extrêmes et qu'il faudrait donc augmenter le nombre de tirage. Cependant un passage au résultat cumulé permet de lisser les résultats :

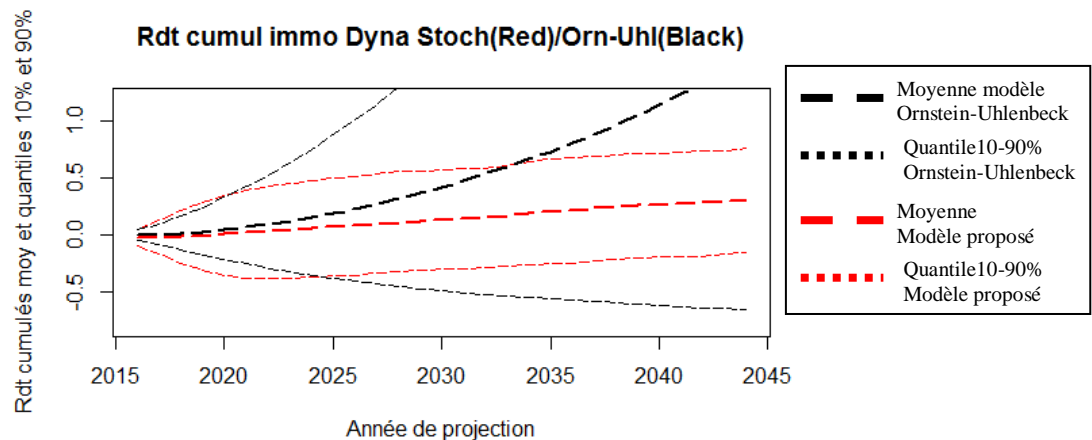


Figure 38 : Comparaison des rendements cumulés avec scénarios sur la durée d'emprunt

Le résultat correspond bien à ce que l'on pouvait attendre : pas de changement en tendance mais l'augmentation de la variabilité des paramètres avec l'ajout de la variation de la durée d'emprunt moyen implique une augmentation de la volatilité de l'indice modélisé.

Le modèle dynamique stochastique permet, par l'utilisation de paramètres économiques réels, de raffiner davantage le réalisme des scénarios.

Pour achever la validation, il faut enfin confronter le modèle à un cas réel

IV.2.2 Backtesting

Le backtesting est un test historique qui évalue un modèle comme s'il avait été effectivement utilisé pendant des périodes passées et dans les mêmes conditions du marché. Je vais donc placer le GSE dans les conditions de l'année 1980 et voir si le GSE permet bien de générer

des scénarios qui incluent l'évolution de l'indice des prix immobiliers enregistrée sur cette période. Il est d'usage de considérer que le passé n'est pas « extrême » au sens des scénarios tirés. Cela se traduit graphiquement par le fait que la courbe historique doit rester la plupart du temps à l'intérieur des quantiles 10% et 90% et être centrée sur la médiane des scénarios. En reprenant le modèle du précédent paragraphe et en comparant le modèle pour l'immobilier d'Ahlgrim au modèle dynamique stochastique avec les paramètres d'entrés suivant :

- Taux long moyens hors inflation = 2,9%
- Inflation moyenne = 5,1%
- Moyenne rendement annuel de l'immobilier hors inflation = 1,5% (pour le modèle Ahlgrim)
- Augmentation annuelle des salaires hors inflation = 1,5% (modèle dynamique stochastique)

Les autres paramètres de vitesses de retour à la moyenne et écarts types sont également repris du travail de Planchet, Théron et Kamega.

On obtient le graphique :

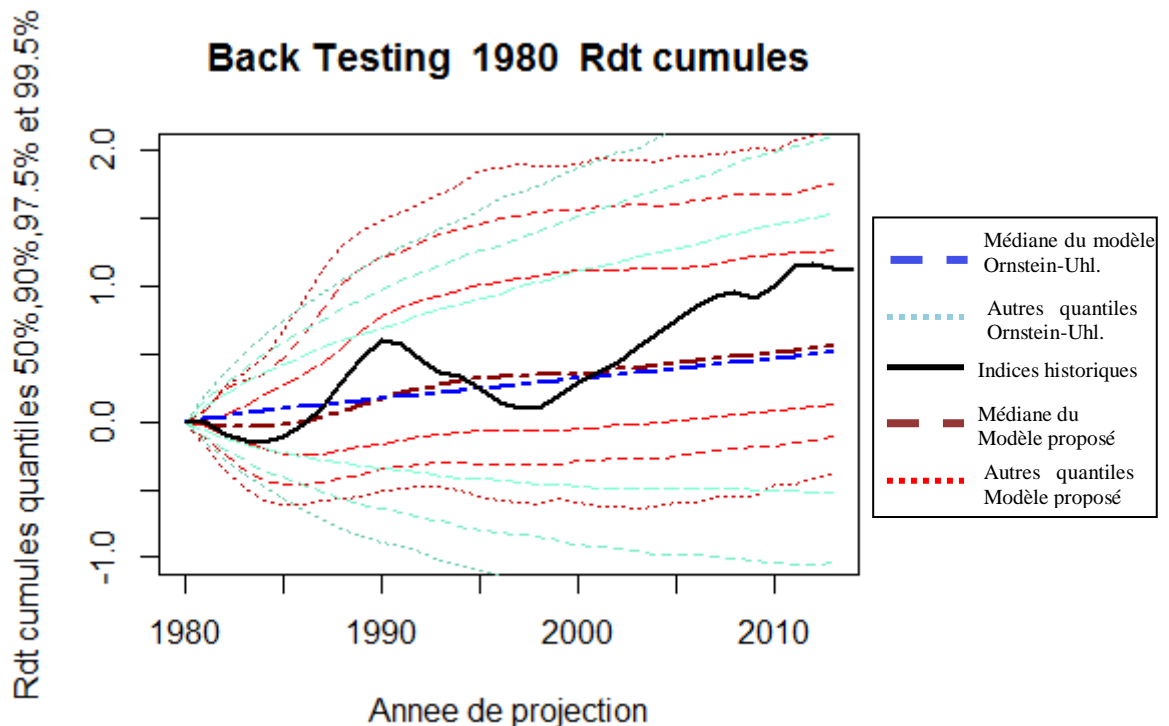


Figure 39 : BackTesting 1980 des rendements cumulés

Concernant le modèle d'immobilier utilisé par Ahlgrim, on remarque que les quantiles 10% et 90% (pointillés bleus, Figure 39) incluent bien la courbe historique (courbe noire, Figure 39). Cependant, les quantiles divergent fortement pour les rendements défavorables et ne sont plus exploitables après dix ans de développement. Comme vu précédemment, ce n'est pas un bon modèle pour estimer l'évolution de l'immobilier sur une longue période (>15 ans) et le backtesting le confirme.

Concernant le modèle dynamique stochastique, on remarque une forme de vague (courbes rouges, Figure 39). Cela provient du fait qu'en 1980 les prix pratiqués sur le marché sont supérieurs au prix d'équilibre défini par le modèle. Cela va donc donner au modèle une

inertie à la baisse de 1980 à 1985 alors que la baisse de l'inflation et des taux d'emprunts va pousser les prix à la hausse par la suite. L'équilibre va alors se rétablir après une oscillation comme nous avons pu le voir lors de l'étude du modèle mathématique. Ces anticipations vont bien se retrouver dans la courbe des rendements de l'indice historique de l'immobilier d'habitation parisien.

Si l'on observe l'évolution des rendements annuels :

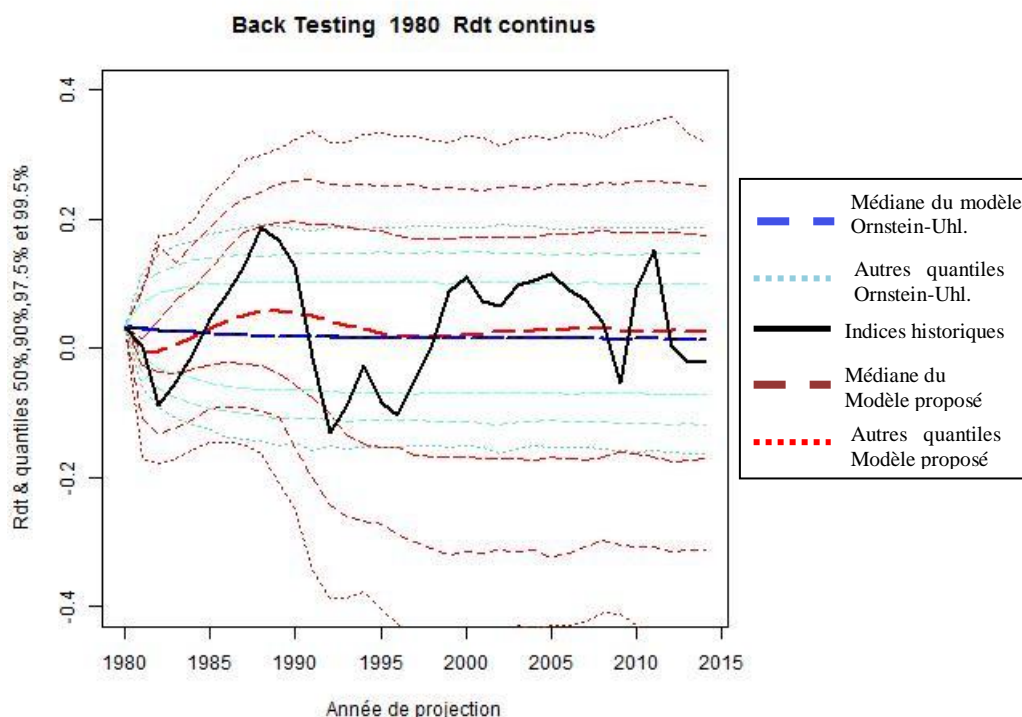


Figure 40 : BackTesting 1980 des rendements annuels

Concernant le modèle utilisé par Ahlgrim, les rendements annuels historiques sont régulièrement de part et d'autre des quantiles (en bleu, Figure 40). Cela montre une sous-estimation de la volatilité des rendements annuels alors que la volatilité des rendements cumulés était surestimée. De nouveau, ce modèle ne semble pas adapté au besoin.

Concernant le modèle dynamique stochastique, la courbe historique est incluse dans les quantiles. On retrouve bien la stabilisation par oscillation du modèle (en rouge, Figure 40) suite au déséquilibre engendré par l'initialisation du modèle avec un indice des prix en 1980 loin de son indice des prix d'équilibre. Cependant, la période d'oscillation du modèle semble trop longue par rapport au cycle d'évolution historique.

Au vu de ces résultats, le modèle dynamique stochastique semble répondre d'avantage à la problématique des GSE. Le modèle utilisé dans les GSE « classiques » comme celui d'Ahlgrim ne représente pas correctement le marché de l'immobilier et ne permet pas un bon niveau de confiance lors d'une prise de décision à long terme consécutive à l'utilisation du GSE. Le modèle dynamique stochastique proposé ici souffre quand à lui d'un niveau de complexité et de sensibilité aux paramètres initiaux qui demande beaucoup d'attention lors de sa mise en œuvre. Il permet en revanche de prendre en compte beaucoup de paramètre d'importance du marché (taux d'intérêt, durée d'emprunt moyen, frais de notaires, intégration du passé récent) et il respecte une dynamique qui semble bien correspondre à ce que l'on observe sur ce marché (effet d'auto-entraînement, sensibilité aux crises financières).

Conclusion

Que cela soit dans le cadre prudentiel, pour la détermination des provisions et du capital de solvabilité, pour la communication vers les tiers (Embedded Value et états comptables) ou pour ses besoins de pilotage technique (choix d'allocation stratégique, stress test), l'organisme assureur doit disposer d'un cadre rigoureux et cohérent prenant en compte l'ensemble des actifs de son bilan et les risques associés. Il utilise pour cela des générateurs de scénarios économiques dont les modèles de projections d'actifs doivent ainsi respecter des contraintes :

- De non sous estimation des rendements défavorables
- Intégrer les contraintes économiques long-terme (plus de 30 ans)
- Intégrer l'inflation et les dépendances entre les actifs

La quasi-totalité des modèles actuels s'appuie sur des structures de dépendance dans lesquels la corrélation tient une place centrale. En pratique, l'intensité de la dépendance augmente dans les situations défavorables, ce qui limite l'efficacité des mesures de diversifications calibrées avec des structures ne prenant pas en compte cet effet. Concernant l'immobilier, actif peut liquide et peu substituable, cette dépendance est usuellement imposée de façon exogène mais n'est pas intrinsèque au modèle.

J'ai, dans ce mémoire, tenté de réaliser la double tâche de :

- Proposer un modèle explicatif de l'immobilier basé sur des données économiques qui peuvent être générées par GSE en créant de fait une corrélation qui se trouve être non linéaire
- Utiliser ce modèle dans le cadre d'un générateur de scénarios économiques (GSE) type Ahlgrim en démontrant son avantage sur le modèle « standard de l'immobilier » (Ornstein-Uhlenbeck)

Concernant le modèle mathématique, les différences conceptuelles entre le modèle classique et le modèle proposé sont nombreuses et pourtant, les formules sont très proches. En effet, le modèle classiquement utilisé pour le développement des scénarios de l'actif immobilier est le modèle Ornstein-Uhlenbeck dont la formule d'évolution du rendement R_i est :

$$R_{i+1} = R_i \cdot e^{-k_{Im}} + \mu_{Im} \cdot (1 - e^{-k_{Im}}) + \sigma_{Im} \cdot \sqrt{\frac{1 - e^{-2 \cdot k_{Im}}}{2 \cdot k_{Im}}} \cdot \varepsilon_{Im}$$

- k_{Im} : la vitesse de retour à la moyenne (constant)
- μ_{Im} : le taux de rendement continu moyen issu de la variation de l'indice des prix immobilier (constant)
- σ_{Im} : l'écart type de l'erreur (constant)
- ε_{Im} : un nombre aléatoire avec une distribution normale centrée réduite

Alors que la proposition de modélisation de ce mémoire, dite dynamique stochastique est:

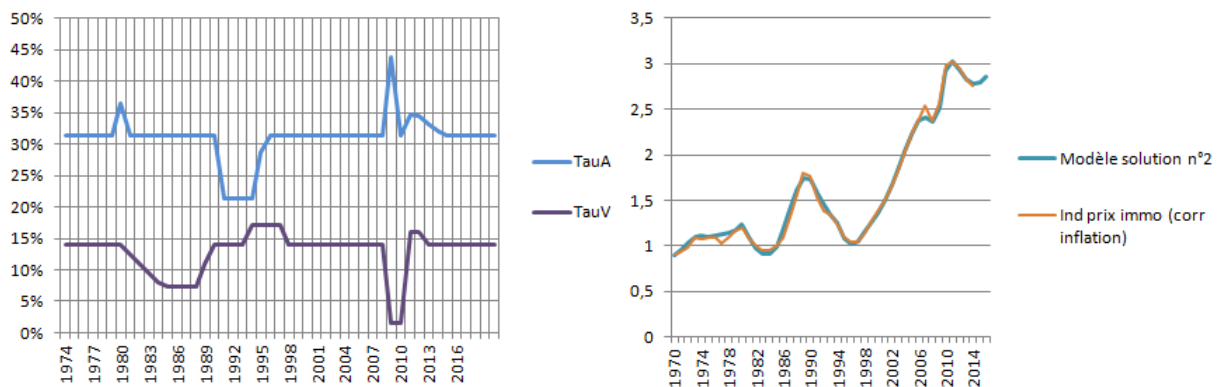
$$R_{i+1} = R_i \cdot \xi + \frac{\sigma_{V_{i+1}}^2}{2} + \mu_{V_{négo}} + \ln \left(\frac{N(d_2)_{i+1}}{N(d_1)_{i+1}} \right)$$

- σ_{V_i} : la volatilité des vendeurs de l'année i
- ξ : le coefficient d'anticipation des acheteurs
- $\mu_{V_{négo}}$: la marge de négociation incluse dans le prix de mise en vente du bien en %

- $N(d_1)$ et $N(d_2)$: fonctions des volatilités et terme de dérive des modèles stochastiques représentant les groupes d'acheteur et de vendeur. Ces formules sont très proches de celles définies pour les options d'échange de Margrabe.

Il est intéressant de constater que la principale différence porte sur le dernier terme des équations. Le dernier terme du modèle dynamique stochastique est un terme non linéaire dépendant des autres variables économiques à chaque instant alors que dans le modèle classique, ce dernier terme est une variable aléatoire corrélée de façon constante. Le modèle proposé répond ainsi au principal défaut du modèle classique. Le caractère aléatoire de la formule dynamique stochastique provient de l'évolution des volatilités des acheteurs et des vendeurs qui doivent être vues comme des réponses à des chocs d'informations sur le marché de l'immobilier.

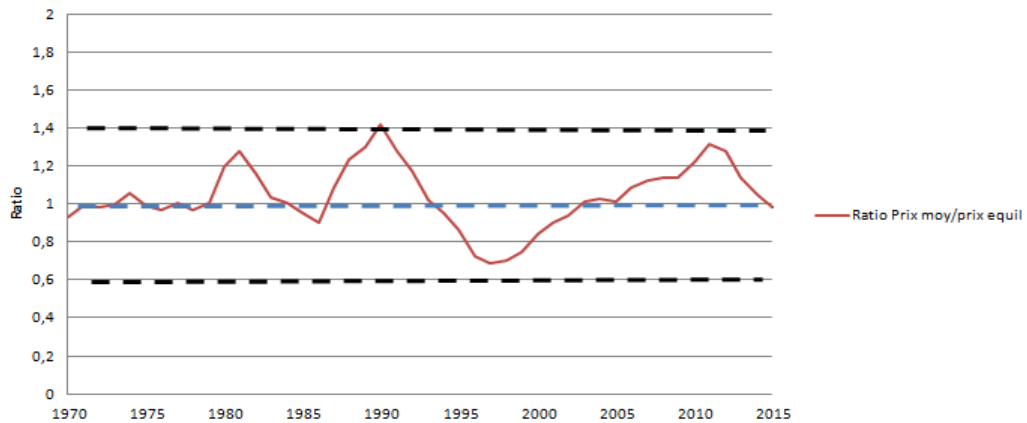
Le modèle proposé est également plus complexe à mettre en œuvre que le modèle classique. Néanmoins, étant fondé sur la description du marché immobilier (plus précisément sur la description d'un marché de biens non substituables), il possède des qualités de réalisme que le modèle classique n'a pas. Il permet par exemple, en retenant 4 grandes crises (1979, 1991-1993, 2008 et 2010) et en recalant la force et la forme des chocs de volatilité, de reproduire l'évolution de l'indice immobilier depuis 1970.



La figure de gauche étant l'évolution calibrée des volatilités acheteur (Tau A) et vendeur (Tau V). La figure de droite superposant l'indice des prix immobilier corrigés de l'inflation au résultat du modèle proposé.

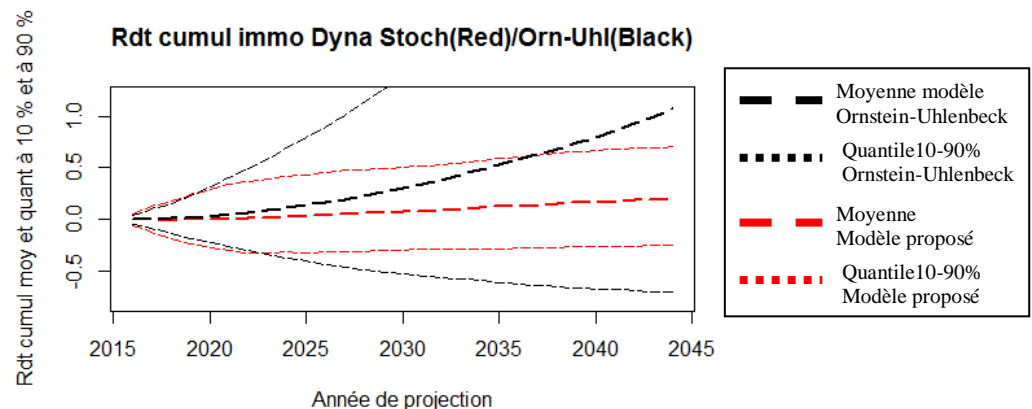
Le modèle calibré correspond bien à l'évolution historique et la modification des paramètres annexes ne change pas significativement les résultats.

De plus, le modèle proposé permet d'établir une notion de Tunnel, qui vient compléter les travaux de Jacques Friggit. En effet, le ratio obtenu en divisant l'indice de prix immobilier par l'indice du prix d'équilibre défini par le modèle doit rester théoriquement dans un Tunnel et c'est bien ce qu'il s'est produit sur la période 1970-2015.



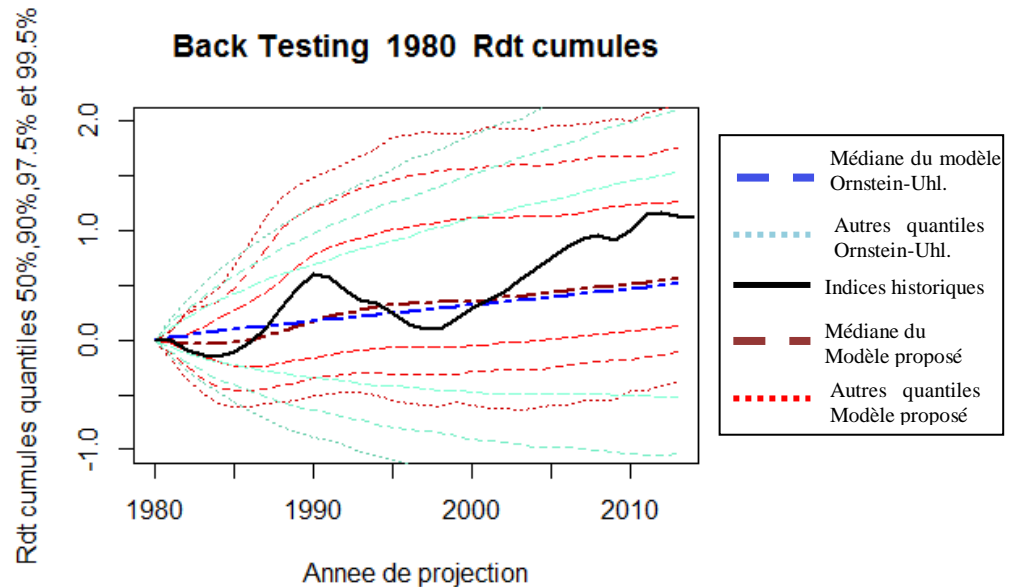
Enfin le dernier avantage de ce modèle est qu'il « accroche » l'évolution de l'indice des prix de l'immobilier à l'évolution du pouvoir d'achat immobilier des acheteurs. En effet, et comme le dit l'adage : « les arbres ne poussent pas jusqu'au ciel ». Cette dernière propriété manque d'ailleurs cruellement au modèle classique.

Si on intègre le modèle dynamique stochastique proposé à un GSE type Ahlgrim et que l'on compare les projections de l'indice des prix de l'immobilier modélisé par le processus Ornstein-Uhlenbeck (standard de ce type de GSE), on obtient le graphique suivant :



On remarque que le niveau d'incertitude est bien moindre avec le modèle présenté ici car les prix évoluent en cohérence avec l'évolution du pouvoir d'achat immobilier des acheteurs. Les prix se corrigeant via des oscillations pour retrouver l'équilibre du marché.

Enfin, pour tester le modèle dans un cas réel, il a été fait un backtesting. Cela consiste à paramétrer le modèle dans une année passée pour ensuite vérifier que les prédictions sont bien conformes à ce qui a été réellement enregistré. Dans le cas d'un backtesting sur l'année 1980, on obtient le résultat suivant :



Sans être exempt de défauts, le modèle dynamique stochastique (pointillés rouges) proposé donne de meilleurs résultats que le modèle utilisé classiquement dans le GSE (pointillés bleus). L'avantage est surtout notable pour les prédictions à très long terme (> 10 ans). Il permet en effet d'avoir un niveau d'incertitude moindre que dans le modèle classique, notamment sur les rendements défavorables. La diminution de l'incertitude et donc du risque pourrait clairement faire revenir cet actif dans les appétits des « Risk adverse » que sont les assureurs.

Il faut rappeler que l'efficacité opérationnelle du modèle dynamique stochastique proposé dépend, dans une très large mesure, de la pertinence des paramètres retenus pour l'alimenter. La détermination de ces paramètres est complexe et fait appel à la fois à des considérations d'ordre statistique (exploitation et choix des historiques considérés), économique (cohérence des valeurs fondamentales) et financières (cohérence avec les prix observés). La prise en compte de ces différentes composantes nécessite une réflexion spécifique qui fait partie intégrante des choix qu'un organisme d'assurance doit faire dans le cadre de sa gestion des risques.

Pour conclure, les résultats sont extrêmement encourageants quant au fait que le modèle capture bien une réalité sous-jacente à l'évolution des prix immobiliers du marché d'habitation parisien de 1970 à aujourd'hui. La démonstration de la pertinence du modèle doit encore être réalisée sur différents marchés mais son application au travers d'un GSE exploitable et réaliste montre l'intérêt d'un tel modèle pour les organismes d'assurance. A ce stade, les travaux de certification pour une utilisation professionnelle restent nombreux. Ce modèle nécessite désormais le soutien d'un acteur reconnu pour le valoriser et le crédibiliser.

Sébastien Gallet

gallet.sebastien2@wanadoo.fr

Bibliographie

Ouvrages imprimés

Planchet F., Théron P. & Kamega A. Scénarios économiques en assurance Modélisation et simulation. Economica collection Assurance Audit Actuariat, 2009. 235 P.

Husson F., Lê S., Pagès J., Analyse de données avec R. Presses Universitaires, 2009. 158P.

Egon H. & Porée P. Statistique et probabilités en production industrielle. HERMANN collection méthodes, 2004. 325 P.

Portait R. & Poncet P. FINANCE DE MARCHE. Dalloz. 2009, second edition. 1101 P.

Travaux universitaires

FALEH A., PLANCHET F., RULLIERE Didier. « Les Générateurs de Scénarios Économiques : quelle utilisation en assurance ? » . Article universitaire: ISFA- Université Lyon. 33 P.

MONBET V. Modélisation de séries temporelles. Cours en ligne 2012-2013. Disponible sur : < https://perso.univ-rennes1.fr/valerie.monbet/ST_M1/CoursST2012.pdf>. (consulté le 27 juillet 2016).

GEEROLF F. UNE ECONOMIE DES BULLES IMMOBILIERES. La découverte, « Regards croisés sur l'économie ». 2011. 45P.

ORLEAN A. Efficience informationnelle versus finance comportementale :éléments pour un débat. Axes de la recherche en sciences économiques CEPREMAP, Presses de la Rue d'Ulm, 2007. 13P.

GALLETI M., PALESTRINI A., BARKLEY ROSSER. The Period Of Financial Distress in speculative markets: interacting heterogeneous agents and financial constraints. Macroeconomic Dynamics, vol. 15, no. 1 , 2011. 22P.

STADNIK B. Spring oscillations within financial markets. *Procedia-Social and behavioral sciences* 110 (2014). Disponible sur : < www.sciencedirect.com>. (consulté le 20 juillet 2016).

BOUCHAUD J.-P., CONT R. A Langevin approach to stoch market fluctuations and crashes, *The European Physical Journal B* 6, 543-550 (1998). 8P. Disponible sur : < www.proba.jussieu.fr/pageperso/ramacont/papers/langevin.pdf >. (consulté le 22 juillet 2016).

BARKLEY ROSSER J. *Econophysics and Economic complexity*. 2008. 26P. Disponible sur : < <http://cob.jmu.edu/rosserjb> >. (consulté le 26 juillet 2016).

ARMEL K. Structure de dépendance des générateurs de scénarios économiques - Modélisation et Simulation, *Mémoire d'actuariat* (2010), EURIA. Disponible sur : < [http://www.ressources-actuarielles.net/EXT/ISFA/1226-02.nsf/8d48b7680058e977c1256d65003ecbb5/59f3eff9eace9a3cc12576fd00345977/\\$FILE/memoire_ARMEL_27_05_2010.pdf](http://www.ressources-actuarielles.net/EXT/ISFA/1226-02.nsf/8d48b7680058e977c1256d65003ecbb5/59f3eff9eace9a3cc12576fd00345977/$FILE/memoire_ARMEL_27_05_2010.pdf)>. (consulté le 22 juillet 2016).

Divers

Jacques FRIGGIT, support de cours ENPC disponible sur <http://www.cgedd.fr/prix-immobilier-presentation.pdf>. (consulté le 9 septembre 2016)

Table des annexes

Annexe 1 :Modèle d'actifs dans le modèle d'Ahlgrim	62
Annexe 2 : Méthode des Densité de probabilité Acheteur/vendeur et densité conjointe	64
Annexe 3 : Résolution numérique du prix de vente moyen par la méthode des fonctions de densités	67
Annexe 4 : Tests sur les simplifications de σV la volatilité implicite des vendeurs et ξ le coefficient d'anticipation.....	70
Annexe 5 : Linéarisation de $Nd2/Nd1$ par la méthode des moindres carrés.....	73
Annexe 6 : Etude multivariée de l'évolution des prix de l'immobilier au regard des données macroéconomiques	78

Annexe 1 : Modèles d'actifs dans le modèle d'Ahlgrim

Le modèle d'Ahlgrim et al. [2005] accorde une importance particulière à l'interdépendance entre les différentes séries économiques et financières étudiées mais cette interdépendance reste limitée car elle se base sur la corrélation. L'objectif est de se mettre en capacité de déterminer, en partant du modèle, les lois du taux d'inflation, du taux d'intérêt nominal à long terme, du taux d'intérêt nominal à court terme, du rendement d'un investissement en actions et du rendement d'un investissement en immobilier. Je ne présente ici que les variables relative au sujet du mémoire à savoir Le taux d'inflation, les taux d'intérêts (réels, nominaux, court terme, long terme) et le taux de rendement immobilier. J'invite le lecteur à consulter le mémoire d'actuariat de Kamal Armel pour une description complète¹⁹.

1. Le taux d'inflation

Le taux d'inflation à l'instant t , noté q , est supposé suivre un processus d'Ornstein-Uhlenbeck (modèle de type Vasicek à un facteur) :

$$dq_t = k_q(\mu_q - q_t)dt + \sigma_q dB_{q,t}$$

- k_q est la vitesse de retour à la moyenne.

- μ_q est le taux moyen à long terme.

- σ_q est l'écart type de l'erreur.

- $B_{q,t}$ est un mouvement brownien.

Afin de calibrer et projeter la dynamique du taux d'inflation, Ahlgrim et al. [2005] propose une résolution approximative de cette équation reposant sur la discrétisation d'Euler.

$$q_{t+\delta} = q_t \exp[-k_q \delta] + \mu_q (1 - \exp[-k_q \delta]) + \sigma_q \sqrt{\frac{1 - \exp[-2k_q \delta]}{2k_q}} \varepsilon_{q,t}$$

où δ représente le pas de la discrétisation et $\varepsilon_{q,t}$ est une variable aléatoire de loi normale centrée réduite.

2. Les taux d'intérêt

Le modèle Ahlgrim et al. [2005] s'intéresse plus particulièrement au taux d'intérêt réels à long et à court terme. Nous présentons dans cette partie les modèles des taux réels tels qu'ils sont présentés dans le modèle d'Ahlgrim et al. [2005] pour en déduire ensuite les lois des taux nominaux. Ahlgrim et al. [2005] proposent de reprendre un cas particulier du modèle Hull-White [1994] pour décrire la dynamique du taux d'intérêt réel long noté l et le taux d'intérêt réel court, noté r . Les équations différentielles stochastiques que vérifient ces deux variables sont (modèle de type Vasicek à deux facteurs) :

¹⁹ ARMEL K. Structure de dépendance des générateurs de scénarios économiques - Modélisation et Simulation, Mémoire d'actuariat (2010), EURIA

$$dl_t = k_l(\mu_l - l_t)dt + \sigma_l dB_{l,t}$$

$$dr_t = k_r(l_t - r_t)dt + \sigma_r dB_{r,t}$$

Comme précédemment, Ahlgrim et al. [2005] propose une résolution approximative de cette équation reposant sur la discrétisation d'Euler.

$$l_{t+\delta} = l_t \exp[-k_l \delta] + \mu_l (1 - \exp[-k_l \delta]) + \sigma_l \sqrt{\frac{1 - \exp[-2k_l \delta]}{2k_l}} \varepsilon_{l,t}$$

$$r_{t+\delta} = r_t \exp[-k_r \delta] + l_t (1 - \exp[-k_r \delta]) + \sigma_r \sqrt{\frac{1 - \exp[-2k_r \delta]}{2k_r}} \varepsilon_{r,t}$$

La réécriture de ces deux formules montre que les lois du taux réel long et du taux réel court sont, à chaque instant t , des lois normales. Aussi, la forme récurrente des taux réels s'écrit où δ représente le pas de la discrétisation.

Le calibrage du modèle des taux d'intérêt réels pose certaines difficultés. D'abord, les séries historiques de ces taux ne sont pas directement observables sur les marchés financiers. Ahlgrim et al. [2005] propose donc de construire ces séries de données en retranchant les taux d'inflation aux taux d'intérêt nominaux. Aussi, une autre difficulté qui concerne la méthode de calibrage se pose. En effet, au regard de la dépendance entre les taux longs et les taux courts, l'utilisation de la méthode des moindres carrés ordinaires (MCO) pour l'estimation des paramètres n'est pas appropriée. Ahlgrim et al. [2005] propose de retenir une procédure d'estimation s'appuyant sur une application en deux étapes de la méthode des moindres carrés ordinaires. Lors de la première étape, la méthode des MCO est appliquée pour estimer les paramètres du taux d'intérêt réel long. Les valeurs estimées de ce dernier servent comme variable explicative lors de la seconde étape pour estimer les paramètres du taux d'intérêt réel court par la méthode des MCO. Il est à noter toutefois que le calibrage des densités de probabilité des taux nominaux nécessite l'estimation de la corrélation entre les résidus des taux réels et de l'inflation.

3. Le taux de rendement de l'immobilier

Le taux de rendement de l'indice immobilier à l'instant t , noté Im , est supposé suivre un processus d'Ornstein-Uhlenbeck :

$$dIm_t = k_{Im}(\mu_{Im} - Im_t)dt + \sigma_{Im} dB_{Im,t}$$

La forme récurrente de la dynamique du rendement de l'indice immobilier s'écrit :

$$Im_{t+\delta} = Im_t \exp[-k_{Im} \delta] + \mu_{Im} (1 - \exp[-k_{Im} \delta]) + \sigma_{Im} \sqrt{\frac{1 - \exp[-2k_{Im} \delta]}{2k_{Im}}} \varepsilon_{Im,t}$$

Soit si l'on se place dans le cas $\delta=1$ et que l'on note que $Im_i = \ln\left(\frac{\tilde{X}_i}{\tilde{X}_{i-1}}\right)$ avec \tilde{X}_i indice des prix immobilier à la période i :

$$\frac{\tilde{X}_{i+1}}{\tilde{X}_i} = e^{\ln\left(\frac{\tilde{X}_i}{\tilde{X}_{i-1}}\right) \cdot e^{-k_{Im}} + \mu_{Im} \cdot (1 - e^{-k_{Im}}) + \sigma_{Im} \cdot \sqrt{\frac{1 - e^{-2k_{Im}}}{2k_{Im}}} \cdot \varepsilon_{Im}}$$

Annexe 2 : Méthode des Densité de probabilité Acheteur/vendeur et densité conjointe

Il aura été nécessaire, pour s'assurer des résultats des calculs issus de l'évaluation des options d'échange, de développer une deuxième méthode de calcul se basant cette fois sur les densités de probabilité des groupes acheteurs/vendeurs et l'utilisation du calcul intégral. Je présente dans cette annexe les développements de cette méthode.

X_A et X_V sont des v.a. log normalement distribuées du type

$$X_A = X_{A_0} e^{\mu_A T + \sigma_A U}$$

Ou encore

$$\ln \frac{X_A}{X_{A_0}} = \mu_A T + \sigma_A U$$

$$\text{Avec } U \sim \mathcal{N}(0,1)$$

De plus on peut trouver la fonction de densité de X_A par la formule

$$f_{X_A}(x) = f_U[\varphi_U^{-1}(x)] \left| \frac{d\varphi_U^{-1}(x)}{dx} \right|$$

Avec

$$\begin{cases} \varphi_U^{-1}(x) = \frac{\ln \frac{x}{X_{A_0}} - \mu_A T}{\sigma_A} \\ f_U = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{U^2}{2}} \end{cases}$$

$$f_U[\varphi_U^{-1}(x)] = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{\left[\ln \frac{x}{X_{A_0}} - \mu_A T\right]^2}{2 \sigma_A^2}}$$

$$\frac{d\varphi_U^{-1}(x)}{dx} = \frac{1}{x \cdot \sigma_A}$$

D'où la densité des mouvements browniens géométriques (MBG)

$$f_{X_A}(x) = \frac{1}{\sigma_A \sqrt{2\pi}} \cdot \frac{1}{x} \cdot e^{-\frac{\left[\ln \frac{x}{X_{A_0}} - \mu_A T\right]^2}{2 \sigma_A^2}} \quad \text{défini pour } x \geq 0$$

De même pour X_V

$$f_{X_V}(y) = \frac{1}{\sigma_V \sqrt{2\pi}} \cdot \frac{1}{y} \cdot e^{-\frac{[\ln \frac{y}{X_{V0}} - \mu_V T]^2}{2\sigma_V^2}} \text{ défini pour } y \geq 0$$

Afin de simplifier l'écriture on utilisera par la suite

$$\begin{cases} \sigma_{A2} = \sigma_A \\ \mu_{A2} = \ln X_{A0} + \mu_A T \end{cases}$$

De même pour σ_{V2} et μ_{V2}

Soit

$$f_{X_A}(x) = \frac{1}{\sigma_{A2} \sqrt{2\pi}} \cdot \frac{1}{x} \cdot e^{-\frac{[\ln x - \mu_{A2}]^2}{\sigma_{A2}^2}}$$

$$f_{X_V}(y) = \frac{1}{\sigma_{V2} \sqrt{2\pi}} \cdot \frac{1}{y} \cdot e^{-\frac{[\ln y - \mu_{V2}]^2}{\sigma_{V2}^2}}$$

Hypothèse importante : On considère ici que acheteurs et vendeurs sont indépendants

Acheteur et vendeur étant indépendant, la densité jointe peut s'écrire

$$f_{X_A, X_V}(x, y) = f_{X_A}(x) \cdot f_{X_V}(y)$$

Muni de ces fonctions de densités, nous allons pouvoir déterminer la vente moyenne du marché.

Moyenne des ventes

Il y a vente lorsque $X_A \geq X_V$ au prix de X_V . Sinon il n'y a pas vente.

$\tilde{X}_{V_{X_A \geq X_V}}$, le prix de vente moyen sur ce marché peut s'écrire par les densités de probabilités :

$$\tilde{X}_{V_{X_A \geq X_V}} = \frac{\int_{-\infty}^{+\infty} \int_y^{+\infty} y \cdot f_{X_A, X_V}(x, y) \cdot dx \cdot dy}{\int_{-\infty}^{+\infty} \int_y^{+\infty} f_{X_A, X_V}(x, y) \cdot dx \cdot dy}$$

En effet, il y a vente si $X_A \geq X_V$ soit si $x \geq y$ d'où la borne inférieure d'intégration.

Afin de faciliter l'intégration, on réalise le changement de variable de la fonction de densité suivant :

$$f_{z,v}(z, v) = f_{X_A, X_V}(x(z, v), y(z, v)) |J_{z,v}|$$

$$\text{Avec } J_{z,v} = \begin{vmatrix} x'_z & x'_v \\ y'_z & y'_v \end{vmatrix} \text{ le Jacobien}$$

En posant $z = \frac{x}{y}$ et $v = y$, on a $x = vz$ et $J = v = y$

$$\begin{cases} v \geq 0 \text{ (prix de vente)} \\ z \geq 1 \text{ (condition de la vente)} \end{cases}$$

On peut ainsi écrire la fonction de densité: $f_{\frac{X_A}{X_V}, X_V} = f_{X_A, X_V}(vz, y) \cdot v$

$$\tilde{X}_{V_{X_A \geq X_V}} = \frac{\int_0^{+\infty} \int_1^{+\infty} v \cdot f_{X_A, X_V}(vz, v) \cdot v \cdot dz \cdot dv}{\int_0^{+\infty} \int_1^{+\infty} f_{X_A, X_V}(vz, v) \cdot v \cdot dz \cdot dv}$$

La fonction de densité de probabilité conjointe est alors

$f_{X_A, X_V}(vz, v) = f_{X_A}(vz) \cdot f_{X_V}(v)$ car X_A et X_V sont indépendants d'où

$$\tilde{X}_{V_{X_A \geq X_V}} = \frac{\int_0^{+\infty} v^2 \cdot f_{X_V}(v) \int_1^{+\infty} f_{X_A}(vz) \cdot dz \cdot dv}{\int_0^{+\infty} v \cdot f_{X_V}(v) \int_1^{+\infty} f_{X_A}(vz) \cdot dz \cdot dv}$$

Avec le calcul intermédiaire :

$$\int_1^{+\infty} f_{X_A}(vz) \cdot dz = \int_1^{+\infty} \frac{1}{v \cdot \sigma_{A2} \sqrt{2\pi}} \cdot \frac{1}{z} \cdot e^{-\frac{[\ln z + \ln v - \mu_{A2}]^2}{2\sigma_{A2}^2}} dz$$

$$\int_1^{+\infty} f_{X_A}(vz) \cdot dz = \frac{1}{v} \left[1 - \int_0^1 \frac{1}{\sigma_{A2} \sqrt{2\pi}} \cdot \frac{1}{z} \cdot e^{-\frac{[\ln z + \ln v - \mu_{A2}]^2}{2\sigma_{A2}^2}} dz \right]$$

Avec le calcul intermédiaire :

$$\int_0^1 \frac{1}{\sigma_{A2} \sqrt{2\pi}} \cdot \frac{1}{z} \cdot e^{-\frac{[\ln z + \ln v - \mu_{A2}]^2}{2\sigma_{A2}^2}} dz = \int_{-\infty}^1 \frac{1}{\sigma_{A2} \sqrt{2\pi}} \cdot e^{-\frac{[t + \ln v - \mu_{A2}]^2}{2\sigma_{A2}^2}} dt \quad (\text{en posant } t = \ln z)$$

$$= N\left(\frac{\ln(1) + \ln v - \mu_{A2}}{\sigma_{A2}}\right)$$

$$\text{soit } \int_1^{+\infty} f_{X_A}(vz) \cdot dz = \frac{1}{v} \left[1 - N\left(\frac{\ln v - \mu_{A2}}{\sigma_{A2}}\right) \right]$$

En appelant N la fonction de répartition de la loi $\mathcal{N}(0,1)$

$$\tilde{X}_{V_{X_A \geq X_V}} = \frac{\int_0^{+\infty} \frac{1}{\sigma_{V2} \sqrt{2\pi}} \cdot e^{-\frac{[\ln v - \mu_{V2}]^2}{2\sigma_{V2}^2}} \left[1 - N\left(\frac{\ln v - \mu_{A2}}{\sigma_{A2}}\right) \right] \cdot dv}{\int_0^{+\infty} \frac{1}{\sigma_{V2} \sqrt{2\pi}} \cdot \frac{1}{v} \cdot e^{-\frac{[\ln v - \mu_{V2}]^2}{2\sigma_{V2}^2}} \left[1 - N\left(\frac{\ln v - \mu_{A2}}{\sigma_{A2}}\right) \right] \cdot dv}$$

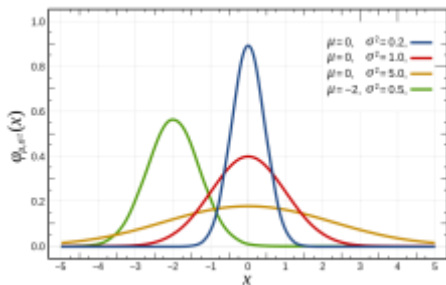
En appelant N la fonction de répartition de la loi $\mathcal{N}(0,1)$ Cette fonction peut ensuite être calculée par intégration numérique.

Annexe 3 : Résolution numérique du prix de vente moyen par la méthode des fonctions de densités

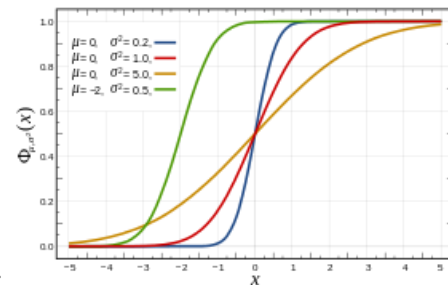
Cette annexe a pour objectif d'expliquer la méthode de résolution numérique qui a été utilisée pour calculer le prix de vente moyen vu au paragraphe II.1.3.2 ainsi qu'en détail dans l'annexe n°1. C'est en fait la résolution de 2 intégrales qui sont des fonctions de fonctions de densité et de répartition de la loi normale centrée réduite.

$$\tilde{X}_{V_{X_A \geq x_V}} = \frac{\int_0^{+\infty} \frac{1}{\sigma_{V2}\sqrt{2\pi}} \cdot e^{-\frac{[\ln v - \mu_{V2}]^2}{2\sigma_{V2}^2}} \left[1 - N\left(\frac{\ln v - \mu_{A2}}{\sigma_{A2}}\right)\right] \cdot dv}{\int_0^{+\infty} \frac{1}{\sigma_{V2}\sqrt{2\pi}} \cdot \frac{1}{v} \cdot e^{-\frac{[\ln v - \mu_{V2}]^2}{2\sigma_{V2}^2}} \left[1 - N\left(\frac{\ln v - \mu_{A2}}{\sigma_{A2}}\right)\right] \cdot dv} = \frac{H}{G}$$

Dans les deux cas, la fonction à intégrer est le produit de la densité (loi normale centrée réduite) par la fonction de répartition.



convolué par



On peut être certain que H et G convergent au vu de la convolution de deux fonctions bornées et dérivables ∞ . Cette forme est localement proche d'une forme conique (polynôme d'ordre 2) et c'est pourquoi j'ai utilisé la méthode de Simpson pour réaliser l'intégration numérique des fonctions H et G.

En effet, la méthode de Simpson est une méthode de calcul numérique d'une intégrale

$$\int_a^b f(x) \cdot dx$$

approchant la fonction à intégrer f par un polynôme P d'ordre 2 en prenant les mêmes valeurs que f aux points d'abscisse a, b et m=(a+b)/2. L'expression de la parabole sur l'intervalle [a,b] est alors définie par l'interpolation Lagrangienne. Le résultat peut-être mis sous la forme :

$$P(x) = f(a) \frac{(x - m)(x - b)}{(a - m)(a - b)} + f(m) \frac{(x - a)(x - b)}{(m - a)(m - b)} + f(b) \frac{(x - a)(x - m)}{(b - a)(b - m)}$$

On approche ensuite l'intégrale de la fonction f , par l'intégrale de la fonction P sur l'intervalle $[a,b]$. On a ainsi la simple formule :

$$\int_a^b f(x)dx \approx \int_a^b P(x)dx = \frac{b-a}{6} \left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b) \right]$$

L'intégrale de chaque intervalle est obtenue en calculant la fonction à intégrer en 3 points $(a,b,(a+b)/2)$

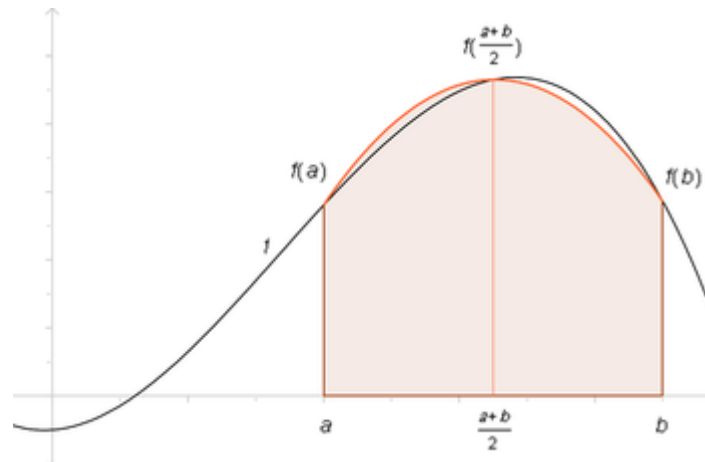


Schéma de la méthode_de_Simpson

Pour H comme pour G, la variable discrétisée est v de 0 à l'infini. Il faut donc pouvoir estimer un point limite à partir duquel il n'est plus intéressant de continuer la discrétisation. J'ai choisi d'utiliser comme limite supérieur d'intégration $10 * X_{V_0}$. Par exemple, si les vendeurs demandent en moyenne 7 000 € / m² à Paris, cela correspond à une limite d'intégration jusqu'à 70 000 € / m². Nous verrons par la suite que cela est suffisant dans notre cas.

Nous allons donc pouvoir comparer le résultat exacte issu du calcul via les options d'échange de Margrabe (cf §II.1.3.1) aux résultats de l'intégration numérique issu du calcul par densité de probabilité. Le tableau ci-dessous reprend différents scénarii :

Scénario	1	2	3	4	5	6	7	8
μ_A	0%	2%	10%	0%	0%	30%	0%	0%
X_{A0}	0,80	0,80	2,00	0,80	0,80	1,00	1,20	0,80
τ_{A0}	30%	30%	30%	30%	30%	30%	20%	30%
μ_V	0%	0%	0%	2%	10%	0%	0%	0%
X_{V0}	1,20	1,00	1,20	1,20	1,20	3,00	1,20	1,00
τ_{V0}	30%	30%	30%	30%	20%	20%	50%	30%
σ	42%	42%	42%	42%	36%	36%	54%	42%
Intégration des densités par la méthode de Simpson								
Proba	0,170	0,316	0,925	0,158	0,080	0,013	0,500	0,299
X moy vente	0,899	0,810	1,208	0,909	1,096	2,290	0,874	0,804
Calcul exacte par les options d'échanges de Margrabe								
d_1	-0,96	-0,48	1,44	-1,00	-1,40	-2,21	0,00	-0,53
d_2	-1,17	-0,69	1,23	-1,21	-1,51	-2,33	-0,46	-0,74
$N(d_1) = \text{Proba}$	0,170	0,316	0,925	0,158	0,080	0,013	0,500	0,299
$N(d_2)$	0,121	0,245	0,890	0,112	0,065	0,010	0,321	0,230
X moy vente	0,899	0,810	1,208	0,909	1,096	2,290	0,874	0,804
Erreurs								
Proba	2,0E-09	-4,5E-09	-1,6E-09	-2,1E-09	-3,4E-09	2,0E-07	-1,4E-06	-4,5E-09
X moy vente	7,8E-09	8,7E-10	-1,8E-10	3,3E-09	-2,6E-09	8,8E-08	-1,3E-06	1,2E-09

Les résultats présentés sont des calculs numériques réalisés avec 100 pas d'intégrations sur l'intervalle $[0 ; 10 * X_{V_0}]$.

Si l'on regarde l'influence du nombre de pas d'intégration sur l'erreur maximum et la moyenne des erreurs des 8 scénarios :

nb pas	Proba		X moy vente	
	max	average	max	average
5	2E+00	4E-01	4E-01	2E-01
10	6E-01	3E-01	1E-01	8E-02
50	5E-04	1E-04	5E-04	1E-04
100	1E-06	2E-07	1E-06	2E-07
500	3E-14	7E-15	4E-14	6E-15
1000	2E-14	3E-15	1E-14	3E-15

Erreur relative moyenne et maximum entre les 2 méthodes sur 8 scénarios

On remarque que les résultats convergent rapidement après 50 pas d'intégrations ce qui est très peu au regard de la puissance de calcul d'un micro ordinateur moderne. Cette étude permet ainsi de conclure sur 2 points :

- La méthode de Simpson est extrêmement efficace pour calculer des intégrales de densités et répartitions (et convoluées) de loi normale.
- Le fait de retrouver les mêmes résultats par deux méthodes différentes nous donne une « quasi » certitude de l'exactitude des deux développements mathématiques. En effet, il paraît impossible que deux méthodes mathématiquement fausses donnent les mêmes résultats.

Annexe 4 : Tests sur les simplifications

Le modèle utilise des paramètres non intuitifs et difficiles à évaluer. Il est donc important de s'assurer que la sensibilité du modèle à ces paramètres soit faible.

σ_V ne prenant pas en compte σ_ξ (volatilité du coefficient d'anticipation)

On a considéré la volatilité implicite des vendeurs σ_V en ne prenant pas en compte σ_ξ et leur corrélation. Il faut donc tester l'impacte de la corrélation de la volatilité des anticipations σ_ξ avec celle du groupe des vendeurs σ_V (Cf chapitre II.1.3.3.1.). Pour voir rapidement l'effet, cette corrélation est fixée à 50%. Cela se comprend en disant que lorsque les vendeurs se dispersent, leurs anticipations se dispersent également. J'ai également considéré que la volatilité des ξ était égale à σ_V (même dispersion autour de la moyenne pour les anticipations que pour les propositions X_V). Cela donne donc avec l'utilisation de σ_V^* :

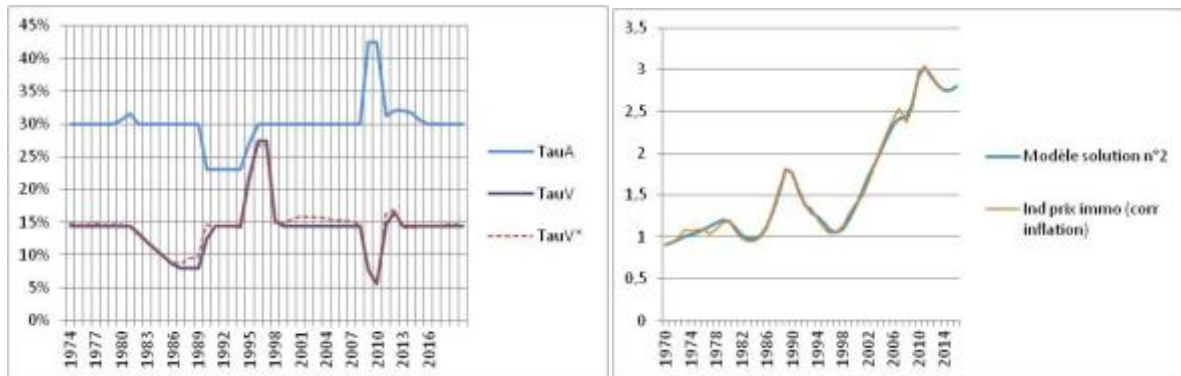


Figure 41 : Solution n°2 avec $\rho_{\xi V} = 50\%$ et $\sigma_\xi = \sigma_V$

Le résultat n'est pas qualitativement modifié. Cela confirme que nous garderons σ_V simplifié dans le reste de l'étude. Il faut également s'attarder sur le paramètre ξ qui a été fixé à 1 dans l'étude et voir si $\xi=f(T)$ change notablement le résultat.

Evaluation de la sensibilité du coefficient d'anticipation $\xi=1$

On a considéré dans ces scénarii $\xi=1$ mais il est peu probable que l'anticipation des vendeurs ne change pas au court du temps en fonction des crises et des époques. Pour cela, j'ai tenté de créer une fonction représentant le fait que lorsque l'évolution annuel des prix se reproduit sur plusieurs périodes, la force de l'anticipation (ξ) augmente et lorsque les variations annuelles varient fortement, l'incertitude grandie et ξ baisse. Soit par exemple la formulation suivante :

$$\xi = 125\% - 25\% * \frac{\text{Ecart Type}_{n-1 \text{ à } n-3} \left(\frac{\Delta \tilde{X}_i}{\tilde{X}_i} \right)}{6\%}$$

Si l'écart type de la variation des prix sur les trois dernières années est faible (voir =0) alors ξ s'approche de 125%. Si elle est égale à 6%, $\xi = 1$ et si elle est supérieure à 6% alors ξ deviendra inférieur à 100%. Notez bien que cette règle est arbitraire et devrait pouvoir être

rattachée à des travaux de finance comportementale. Il faudrait également essayer de la déduire des observations au travers des méthodes statistiques économétriques mais ce n'est pas l'objet de cette étude.

Après nouvelle optimisation des crises on obtient :

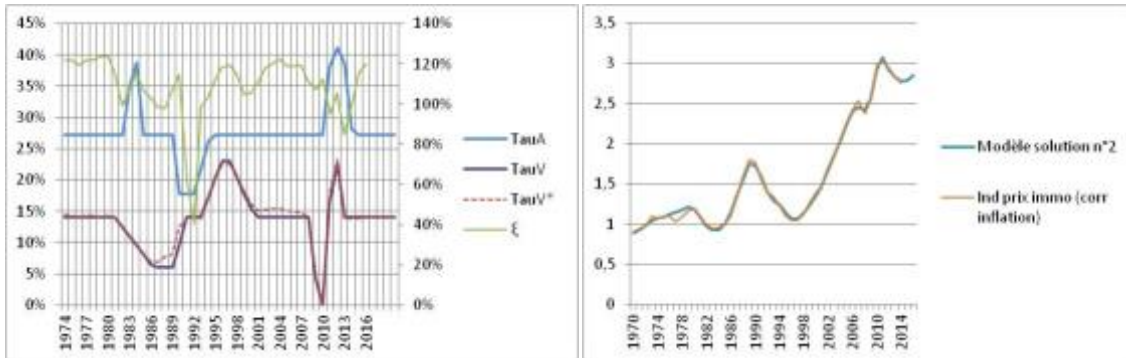


Figure 42 : Solution n°2 avec $\rho_{\xi V} = 50\%$ et $\sigma_{\xi} = \sigma_V$ et $\xi = f(T)$ (Echelle de droite)

Le résultat n'est pas qualitativement modifié. Au vu de ces deux tests, les deux paramètres semblent négligeables mais doivent être gardés à l'esprit car ils pourront devenir prépondérants dans certains cas d'utilisation de la formule. Dans le reste de l'étude nous utiliserons σ_V (et non σ_V^*) ainsi qu' $\xi = 1$.

Test du coefficient de corrélation entre les vendeurs et les acheteurs avec $\rho=20\%$ et -20%

Depuis le début de l'étude du recalage sur l'indice immobilier, l'hypothèse a été faite qu'il n'y avait pas de corrélation entre les volatilités acheteurs et vendeurs. Afin de s'assurer que cette hypothèse soit justifiée, j'ai refait l'optimisation de la solution du § **Error! Reference source not found.** avec cette fois le paramètre $\rho=20\%$ et $\rho=-20\%$

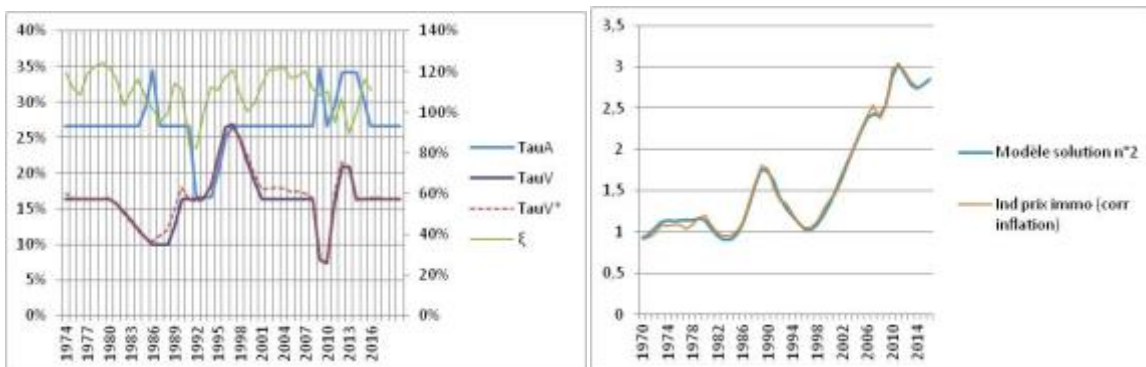


Figure 43 : Optimisation de la solution § **Error! Reference source not found.** avec $\rho=20\%$

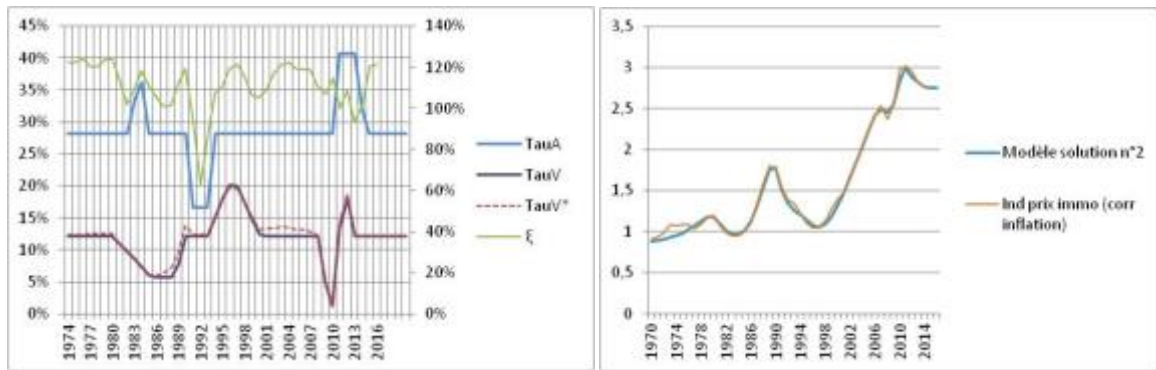


Figure 44 : Optimisation de la solution § **Error! Reference source not found.** avec $\rho=-20\%$

Les résultats de l'optimisation ne changent pas qualitativement et il est intéressant de constater que la solution est robuste vis-à-vis des modifications des paramètres. Cela laisse entendre que beaucoup de scénarii différents auraient conduit à l'évolution des prix de l'immobilier français ou à des courbes très similaires. C'est un autre argument de poids en faveur de la bonne représentativité de cette modélisation.

Annexe 5 : Linéarisation de Nd2/Nd1 par la méthode des moindres carrés

L'objectif de cette annexe est d'expliquer la méthode utilisée pour obtenir une forme linéarisée de l'expression $\ln \frac{N(d_2)}{N(d_1)}$ fonction de σ_A , σ_V les volatilités implicites des groupes acheteurs et vendeurs. En effet, le terme $\ln \frac{N(d_2)}{N(d_1)}$ est non calculable analytiquement. Cependant, afin de simplifier la résolution, nous ferons une approximation de cette fonction par une droite affine fonction de $\ln \left(\frac{X_{A_i}}{\tilde{X}_i} \right)$. Soit

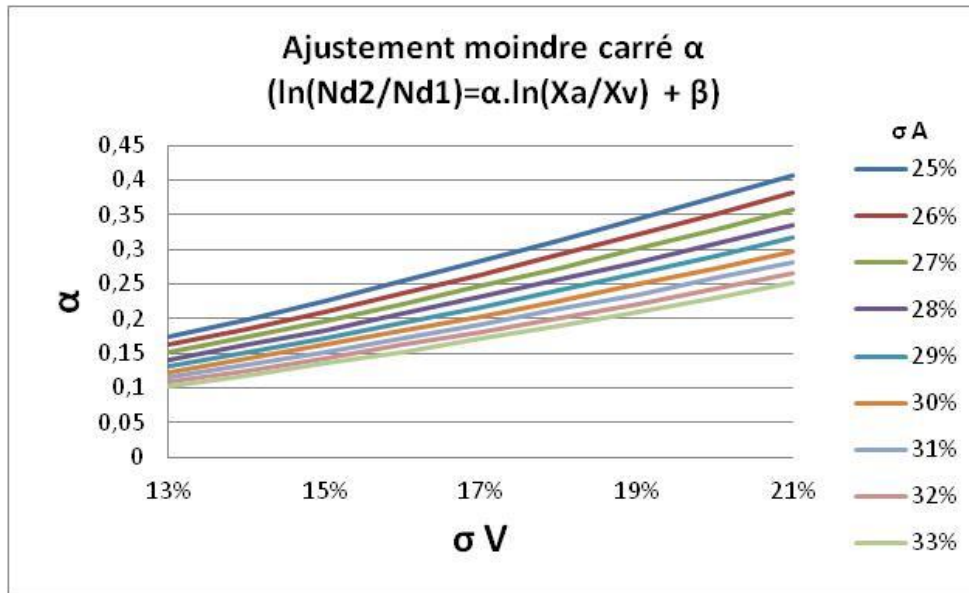
$$\ln \frac{N(d_2)_i}{N(d_1)_i} = \alpha \cdot \ln \left(\frac{X_{A_i}}{\tilde{X}_i} \right) + \beta$$

Nous pouvons donc à l'aide des formules définies §II.1.3.1 calculer pour des ensembles $\{\sigma_A, \sigma_V\}$ et en fixant $\mu_A = 0$ et $\mu_{fees} = 10\%$, une approximation de α et β par la méthode classique des moindres carrés à l'aide du tableur.

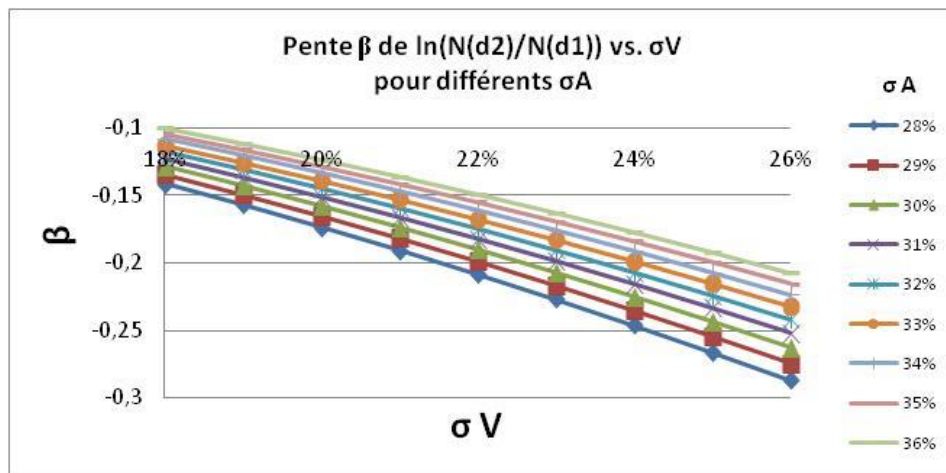
Nous obtenons alors les tableaux suivants pour les différentes valeurs de α (σ_A valeurs horizontales, σ_V verticales)

	25%	26%	27%	28%	29%	30%	31%	32%	33%
13%	0,17367043	0,16151592	0,1505478	0,14062168	0,1316137	0,12341733	0,11594061	0,10910389	0,10283797
14%	0,19911235	0,18533151	0,17287545	0,16158618	0,15132761	0,14198219	0,13344809	0,12563684	0,11847135
15%	0,22580661	0,2103611	0,19637653	0,18368246	0,1721314	0,16159536	0,15196309	0,14313757	0,13503403
16%	0,25364665	0,23651024	0,22096722	0,20683602	0,1939587	0,18219762	0,17143263	0,1615586	0,15248337
17%	0,28252649	0,26368456	0,2465634	0,23097174	0,21674235	0,20372882	0,19180276	0,18085145	0,17077572
18%	0,31234157	0,29179044	0,27308118	0,25601443	0,24041488	0,22612831	0,2130189	0,20096694	0,1898667
19%	0,34298949	0,32073569	0,30043758	0,28188934	0,26490892	0,24933535	0,23502622	0,22185549	0,20971144
20%	0,3743707	0,35043014	0,32855101	0,30852266	0,29015768	0,27328947	0,25776993	0,24346742	0,23026485
21%	0,40638898	0,38078617	0,35734174	0,33584191	0,31609529	0,29793079	0,28119558	0,26575322	0,25148184

Ce qui permet de tracer le graphique suivant :



On remarque immédiatement qu'une linéarisation de $\alpha = f(\sigma_A, \sigma_V)$ donnera de bon résultats si la plage est bien choisie.
De même pour β



On cherche donc à écrire α (respectivement β) en fonction de σ_A , σ_V et un scalaire s (pris =1) en minimisant l'erreur faite. Il faut donc définir les coefficients μ qui minimisent ε tels que

$$\alpha = \begin{pmatrix} \sigma_A \\ \sigma_V \\ s \end{pmatrix}^T \cdot \mu + \varepsilon \text{ pour différence valeurs de } \sigma_A, \sigma_V$$

α (respectivement β), σ_A et σ_V étant ici des vecteurs lignes. La taille de ces vecteurs dépendants du nombre de cas d'estimation.

Avec les hypothèses suivantes concernant le résidu :

$$E[\varepsilon] = 0$$

$$E[\varepsilon\varepsilon^T] = \sigma^2 I_n ; \text{ avec } I_n \text{ la matrice identité.}$$

Dans ce cas, l'estimateur des moindres carrés ordinaire (MCO) est

$$\hat{\mu} = \left[\begin{pmatrix} \sigma_A \\ \sigma_V \\ s \end{pmatrix} \cdot \begin{pmatrix} \sigma_A \\ \sigma_V \\ s \end{pmatrix}^T \right]^{-1} \cdot \begin{pmatrix} \sigma_A \\ \sigma_V \\ s \end{pmatrix} \cdot \alpha$$

Le calcul est ainsi fait pour différentes valeurs de σ_A , σ_V .

Si l'on prend l'exemple de σ_A , σ_V variant de 10% à 50% par pas de 5%, cela fait 81 couples σ_A , σ_V pour lesquels nous avons une valeur de α . Cela permet donc de définir les coefficients

Tau A	-3,242
Tau V	2,834
s	0,846

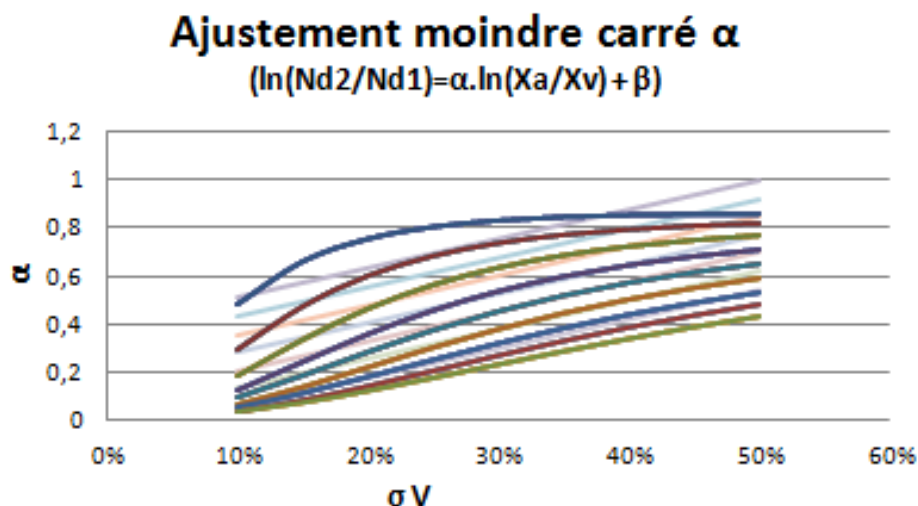
Qui permettent de réécrire α sous la forme linéarisée suivante :

$$\alpha = 2,834 \sigma_V - 3,242 \sigma_A + 0,846$$

Qui lorsque l'on compare aux résultats des α issu de la formule donne le tableau d'erreurs relatives suivant

Erreur alpha	10%	15%	20%	25%	30%	35%	40%	45%	50%
10%	-50%	-131%	-190%	-193%	-106%	109%	487%	1067%	1885%
15%	-2%	-43%	-81%	-98%	-81%	-11%	127%	349%	672%
20%	15%	-11%	-38%	-56%	-57%	-32%	29%	135%	295%
25%	21%	3%	-17%	-34%	-41%	-33%	-4%	52%	141%
30%	23%	9%	-7%	-21%	-30%	-30%	-17%	14%	67%
35%	22%	12%	-1%	-13%	-23%	-27%	-21%	-4%	28%
40%	21%	13%	2%	-9%	-18%	-24%	-23%	-14%	6%
45%	18%	12%	3%	-7%	-16%	-22%	-24%	-19%	-7%
50%	16%	11%	3%	-6%	-14%	-20%	-24%	-23%	-16%

Si l'on superpose graphiquement les α à leurs linéarisation on obtient

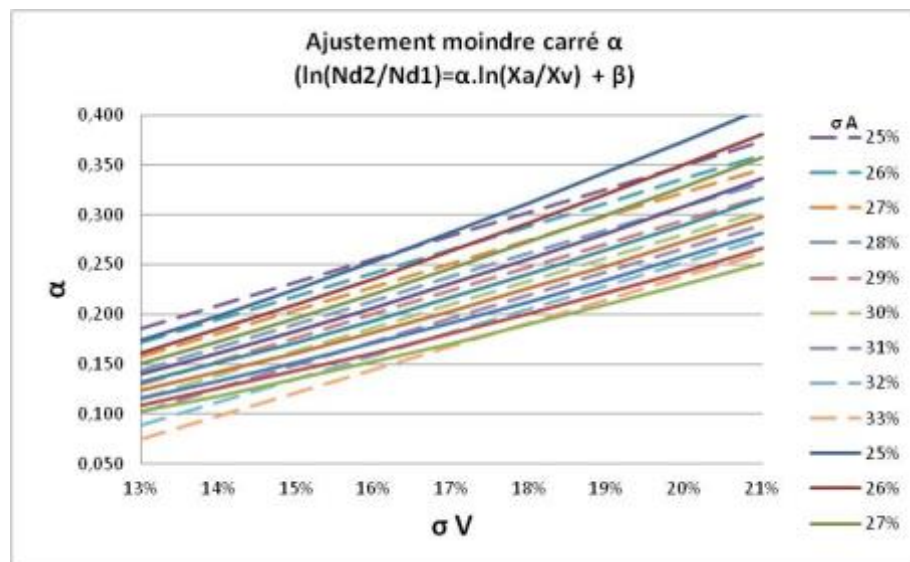


On voit bien que l'ajustement n'est pas suffisamment bon sur l'ensemble des valeurs étudiées.

Il faut donc considérer une plage de linéarisation plus restreinte. Par exemple, si l'on se borne sur l'intervalle [25% ; 31%] pour σ_A et [14% ; 21%] pour σ_V alors on obtient le tableau des erreurs relatives suivant :

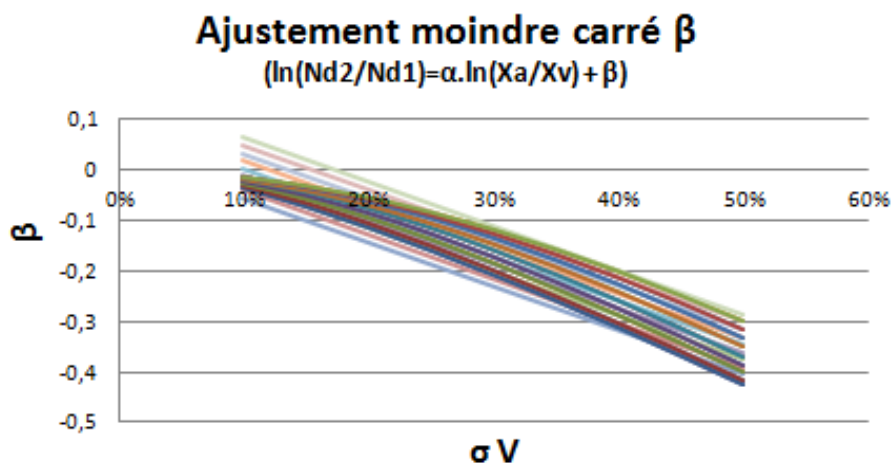
Erreur alpha	25%	26%	27%	28%	29%	30%	31%	32%
13%	-7%	-6%	-5%	-2%	1%	6%	12%	19%
14%	-5%	-5%	-5%	-4%	-1%	2%	6%	11%
15%	-3%	-4%	-4%	-4%	-3%	-1%	2%	6%
16%	-1%	-2%	-3%	-4%	-3%	-2%	-1%	2%
17%	1%	-1%	-2%	-3%	-3%	-3%	-2%	-1%
18%	3%	1%	-1%	-2%	-3%	-3%	-3%	-2%
19%	5%	3%	1%	-1%	-2%	-3%	-3%	-3%
20%	7%	4%	2%	0%	-1%	-3%	-3%	-4%
21%	8%	6%	3%	1%	0%	-2%	-3%	-4%

Soit graphiquement la superposition entre la linéarisation et les valeurs brutes :

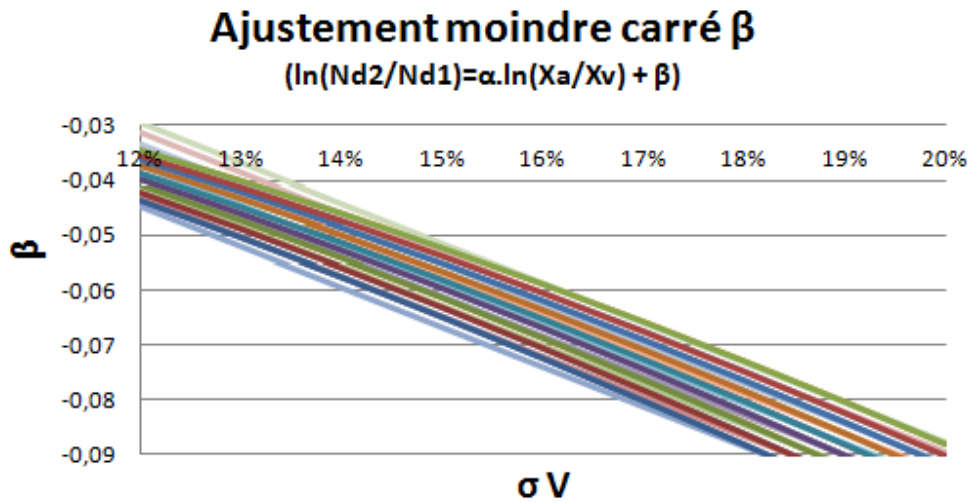


Ce qui est acceptable dans le cadre de cette étude.

De même pour β , le recalage sur un intervalle trop grand conduit à des erreurs relatives non acceptables.



Mais la restriction à l'intervalle [25% ; 31%] pour σ_A et [14% ; 21%] pour σ_V permet de définir une linéarisation acceptable qui est présentée dans le mémoire §II.1.3.3.2. Graphiquement, la superposition pour β est la suivante :



Avec le tableau d'erreurs relatives :

Erreur alpha	25%	26%	27%	28%	29%	30%	31%	32%	33%
13%	-7%	-6%	-5%	-2%	1%	6%	12%	19%	28%
14%	-5%	-5%	-5%	-4%	-1%	2%	6%	11%	17%
15%	-3%	-4%	-4%	-4%	-3%	-1%	2%	6%	10%
16%	-1%	-2%	-3%	-4%	-3%	-2%	-1%	2%	5%
17%	1%	-1%	-2%	-3%	-3%	-3%	-2%	-1%	2%
18%	3%	1%	-1%	-2%	-3%	-3%	-3%	-2%	-1%
19%	5%	3%	1%	-1%	-2%	-3%	-3%	-3%	-2%
20%	7%	4%	2%	0%	-1%	-3%	-3%	-4%	-4%
21%	8%	6%	3%	1%	0%	-2%	-3%	-4%	-4%

Ce qui en résumé conduit à

$$\alpha = 2,343 \cdot \sigma_V - 1,393 \cdot \sigma_A + 0,229$$

Valable pour $25\% < \sigma_A < 31\%$ et $14\% < \sigma_V < 21\%$ (Erreur inférieur à 10% sur α)

De même pour β

$$\beta = 0,553 \cdot \sigma_A - 1,352 \cdot \sigma_V - 0,048$$

Valable pour le même intervalle.

Annexe 6 : Prévisions à moyen terme

Le modèle ayant une forte dimension explicative peut-il également être prédictif ?
Il faut pour cela faire l'hypothèse de futures crises car sur un temps « court » (moins de 20 ans) elles sont prépondérantes vis à vis des changements démographiques.
Tout d'abord, voyons l'évolution de l'indice sans évolution du taux ni des autres paramètres (sauf une augmentation des salaires hors inflation de +0,5% par an)

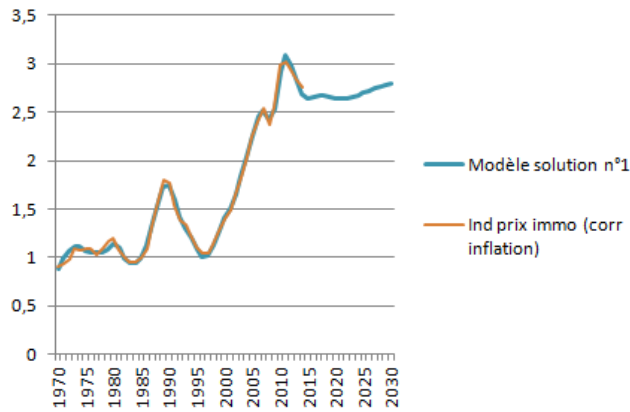


Figure 45 : prévision en statut quo

Remontée des taux d'intérêt des prêts bancaires sans réaction des acteurs

Si on crée une remontée brusque en 2020 des taux d'intérêt (+2%) amenant les taux d'emprunt de 2% actuellement à 4% en 2020 on obtient :

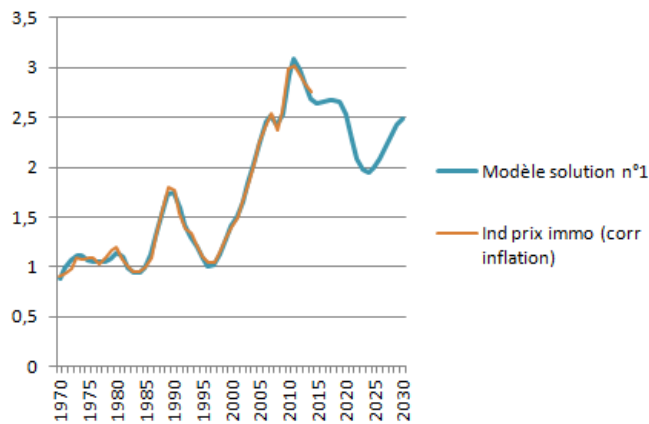


Figure 46 : Taux d'emprunt passe brusquement de 2% à 4% en 2020

Alors qu'avec une remontée lente des taux d'emprunt de 2% en 2020 à 4% en 2029 à raison de 0,2% par an, on obtient :

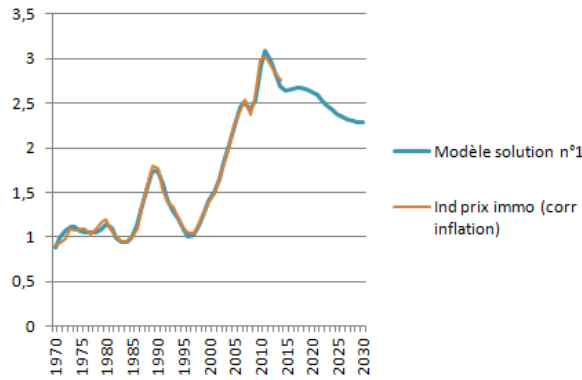


Figure 47 : Taux d'emprunt passe progressivement de 2 à 4% à raison de 0,2% par an

Cela permet une stabilisation du marché qui a le temps de suivre la tendance sans oscillations. Mais en réalité et comme nous l'avons vu avec le recalage du modèle au § III.2.3, les acteurs vont réagir à chaque nouvelle information.

Sans annonce anticipée, réaction à postériori des acteurs

Plaçons-nous dans le cas de l'augmentation brutale des taux d'emprunt mais cette fois avec une réaction des vendeurs voyant les prix baisser brutalement et souhaitant s'adapter. Les vendeurs, voyant les prix baisser, se mettent en concurrence (comme dans les années 1990). Il faut noter ici que les acheteurs réagissent brutalement mais cela ne dure pas, leur budget est diminué et voyant les prix baisser, ils retournent rapidement à leur volatilité « de base ».

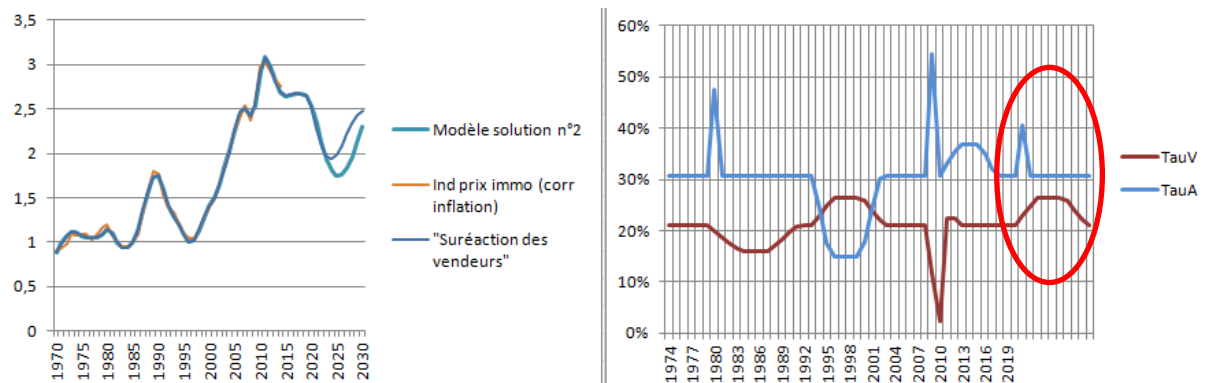


Figure 48 : Relèvement brutal des taux d'emprunt en 2020 et sur-réaction des vendeurs comme pour 1993

Dans ce cas, après un réel effondrement de prix de l'immobilier, une très forte oscillation fait remonter le prix. Mais si on considère un relèvement progressif des taux d'emprunts à partir de 2020 :

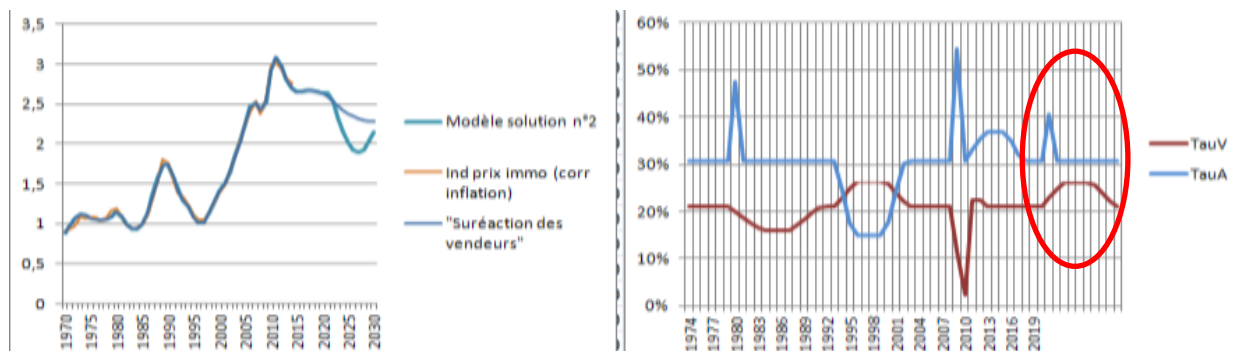


Figure 49 : Un relèvement progressif des taux d'emprunt à partir de 2020 et surréaction des vendeurs comme pour 1993

On constate que la réaction des vendeurs produit une chute rapide des prix suivie par une oscillation. Dans ce cas, comme après 1993, les prix ne reflètent plus le prix d'équilibre mais intègrent la peur de ne pas pouvoir vendre « assez vite ».

Voyons maintenant les conséquences d'une annonce laissant les acteurs anticiper le changement.

Avec annonce anticipée, réaction à priori des acteurs

Considérons qu'en 2019 une annonce soit faite en prévenant d'une remontée des taux d'emprunt de 2% en 2020. On peut imaginer que les acheteurs vont vouloir accélérer leur achat avant de voir leur budget fondre. De plus les vendeurs ne vont pas anticiper de baisse des prix suite à la remontée des taux. Les vendeurs vont se comporter comme si l'annonce n'avait pas été faite et seront malgré tout surpris par la baisse des prix

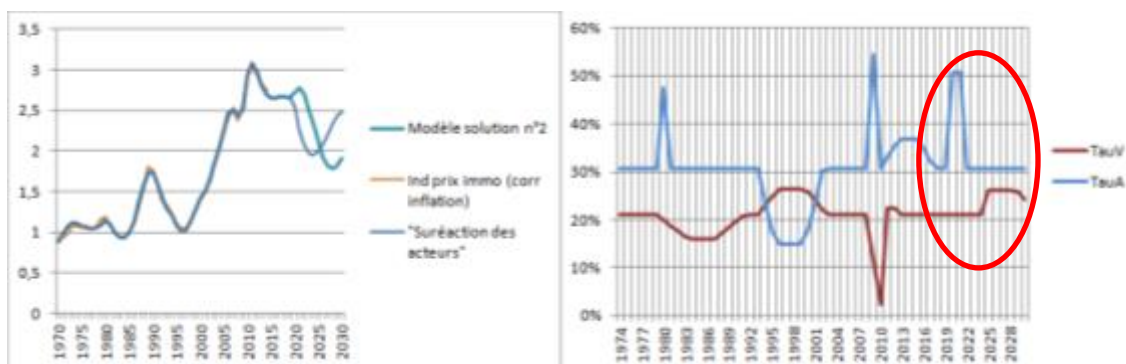


Figure 50 : prévision pour une annonce d'une remontée brusque des taux à horizon 1 an

L'anticipation des acheteurs tend à faire augmenter les prix mais ne repousse que de 2 ans la chute. La profondeur de la baisse est amplifiée ainsi que l'oscillation qui va suivre.

Avec une remontée progressive des taux de 0,2% par an on obtient :

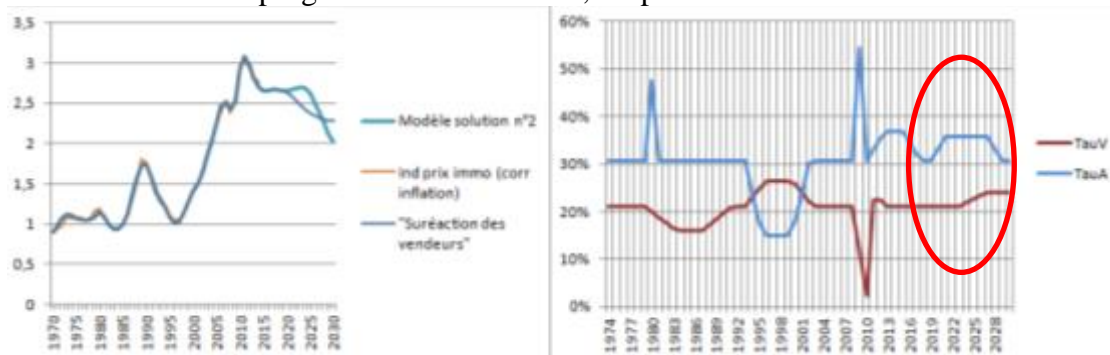
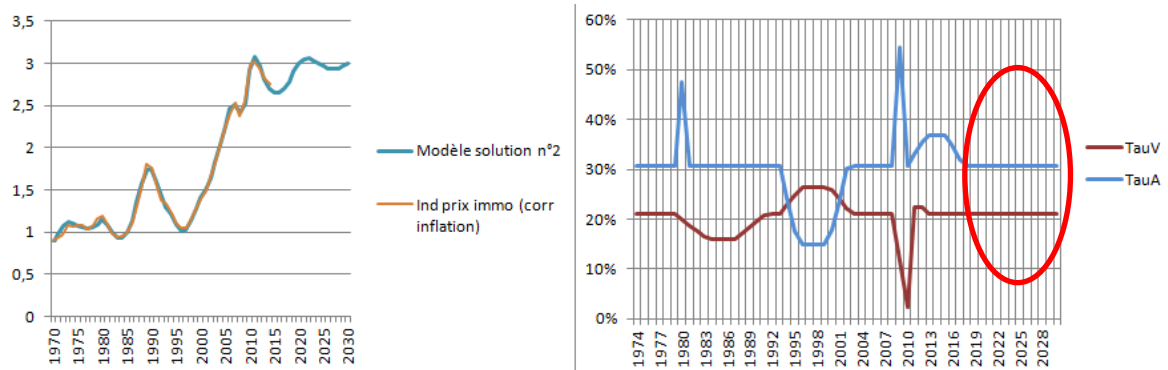


Figure 51 : prévision pour une annonce d'une remontée progressive des taux

On constate qu'après une stagnation des prix prolongée de quelques années, les prix finissent par repartir à la baisse vers leur prix d'équilibre. La force des oscillations dépendra de la réaction des vendeurs à cette baisse.

Baisse des Taux à 1% et maintient jusqu'en 2030

Bien que ce scénario paraisse encore peu probable aujourd'hui, il fallait tout de même évoquer la possibilité que les taux baissent encore jusqu'à 1% de Taux moyen d'emprunt et pour longtemps. Cela pourrait arriver dans le cas d'une Europe en récession lente (mais sans choc) qui ne pourrait pas remonter ces taux au risque de créer une grave récession. Vu qu'il n'y a pas de mouvement brusque, il n'y aurait pas d'anticipation des acteurs. Le modèle prévoit dans ce cas :



En effet, on retrouve bien les oscillations pour converger vers le prix d'équilibre correspondant à un taux de 1%. Dans ce cas, les prix (hors inflation) retrouveraient leurs niveaux de 2010.

Annexe 7 :

Etude multivariée de l'évolution des prix de l'immobilier au regard des données macroéconomiques

1.	Présentation du sujet d'étude	83
1.1.	OBJECTIF DE L'ETUDE.....	83
1.2.	VARIABLE FRANCE DE 1970 A 2010 : F7010	83
1.2.1.	PRESENTATION DES VARIABLES BRUTES	83
1.3.	PREMIERE ACP « RAPIDE » ET PREMIERE REDUCTION DES VARIABLES DE L'ETUDE.....	84
2.	Analyse et choix des variables étudiées et méthodes	88
2.1.	APPROCHE DIFFERENTIELLE POUR LA VARIATION DES PRIX DE L'IMMOBILIER PAR RAPPORT AU TAUX D'EMPRUNT : VF7010.....	88
2.2.	RESUME DES VARIABLES A ETUDIER.....	89
2.2.1.	METHODES UTILISEES	89
3.	Données brutes F7010	90
3.1.	DESCRIPTION MULTIVARIEE	90
3.1.1.	ANALYSE ACP DES DONNEES BRUTES CENTREES REDUITES	90
3.1.2.	ACM SUR DONNEES BRUTES F7010	92
3.2.	PREDICTION	94
3.2.1.	PLS-DA SUR DONNEES BRUTES F7010	95
3.2.2.	DISQUAL	96
3.2.3.	PCR DONNEES BRUTES.....	97
3.2.4.	PLSR DONNEES BRUTE F7010.....	99
4.	Variation des données brutes VF7010	99
4.1.	DESCRIPTION MULTIVARIEE	100
4.1.1.	ACP DE LA VARIATION DES DONNEES BRUTES VF7010	100
1.	ACM SUR DONNEES VF7010	101
4.2.	PREDICTION	102
4.2.1.	PLS-DA SUR DONNEES VF7010	102
4.2.2.	DISQUAL	102
4.2.3.	PCR DONNEES BRUTES.....	103
4.2.4.	PLSR DONNEES VF7010.....	103
4.2.5.	CONCLUSION A L'ANALYSE MULTIVARIEE	104
5.	Analyse temporelle	105
5.1.	TEST DE L'HYPOTHESE D'AUTOREGRESSIVITE ENTRE ANNEES.....	105
5.2.	ANALYSE FONCTIONNELLE	107
A.	F7010.....	107
B.	VF7010.....	110
5.3.	CONCLUSION.....	112

1.Présentation du sujet d'étude

1.1.Objectif de l'étude

Cette étude a pour objectif de mettre en avant les liens entre les mouvements de l'indice immobilier et les variables économiques classiques. Cela constitue une étude préliminaire au projet de formaliser un modèle d'évolution stochastique des cours de l'immobilier en y intégrant les variables issues de cette étude.

1.2.Variable France de 1970 à 2010 : F7010

1.2.1.Présentation des variables brutes

La liste des variables qui vont être étudiées est la suivante :

Années
Salaire moyen net indice base 100 en 1951 (en euro constant)
Nb de ménages
Taux d'intérêt LT
Nb de logements commencés
Dettes immobilières des ménages, en fin d'année
Durée moyenne des prêts pour l'achat d'une résidence principale dans l'ancien
Espérance de vie à la naissance
Prix de l'or (€, base 2000)
Prix du pétrole (en Dollars constant)
Taux de chômage (BIT)
Indicateur de la confiance des ménages
Opinion sur leur capacité d'épargne
Indice du prix de l'immobilier, Paris corrigé de l'inflation
Indice du l'immobilier Paris corrigé de l'inflation divisé par le salaire moyen en euro constant. Cette variable correspond à la part de l'immobilier dans le budget moyen (ou l'effort à l'accession) fait par les ménages. Une variable qualitative qui indique la hausse ou la baisse de l'indice du prix du logement corrigé de l'inflation
Années de Hausse/Baisse (contient uniquement ces deux modalités)

Les variables « indice immo » vont être étudiées en variables supplémentaires car c'est celles que l'on cherche à expliquer. La dernière variable va permettre de différencier visuellement les années de hausse de celles de baisse mais aussi d'appliquer les méthodes discriminantes. Les années seront ici les individus. Les individus sont les années de 1970 à 2010 soit 41 années en ligne.

Les données vont être centrées réduites avant d'être étudiées et ce pour l'ensemble des méthodes.

1.3.Première ACP « rapide » et première réduction des variables de l'étude

On réalise une première ACP dont les résultats sont les suivants :

```
F7010 <- read.table("G:/STA 201/Tabs/F7010.csv", header=TRUE,
  sep=";", na.strings="NA", dec=".", strip.white=TRUE, row.names=1)
#-----
#pca avec package Factominer
#-----
resF7010<-PCA(F7010, ncp=Inf , quanti.sup=c(12,13), quali.sup=c(14:15))
plot.PCA(resF7010, axes=c(2,3), habillage = 14)
```

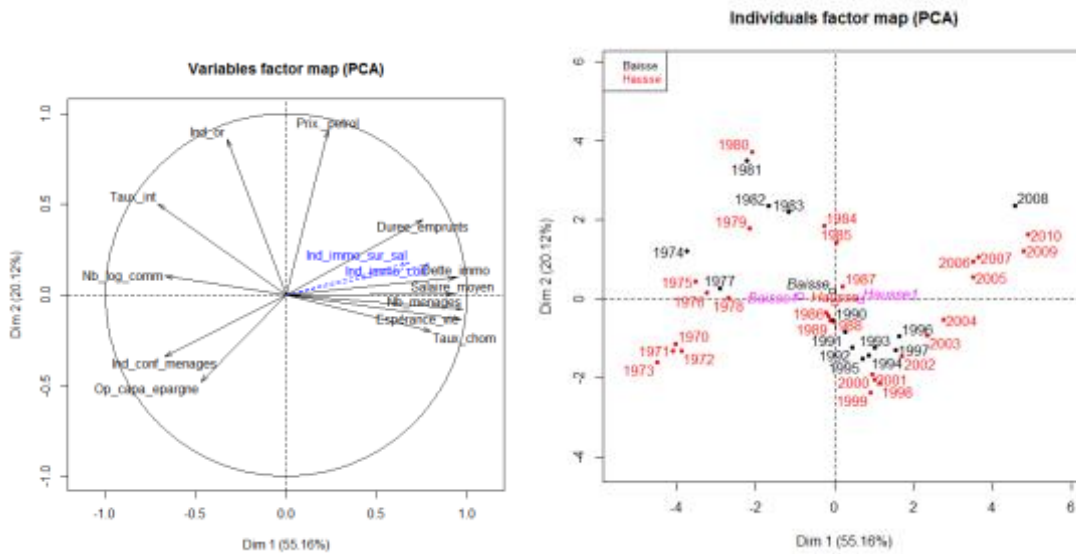
	Eigen value	percentage of variance	cumulative percentage of variance
comp 1	6,6	55,2	55,2
comp 2	2,4	20,1	75,3
comp 3	1,4	11,9	87,2
comp 4	1,0	8,4	95,6
comp 5	0,2	1,6	97,2
comp 6	0,1	0,9	98,1
comp 7	0,1	0,7	98,8
comp 8	0,1	0,6	99,3
comp 9	0,0	0,3	99,6
comp 10	0,0	0,2	99,9
comp 11	0,0	0,1	100,0

La description des axes indique que la dimension1 représente le temps qui passe (Nb de ménage, dette, espérance de vie...) et ne va pas être utile dans le cadre de cette étude.

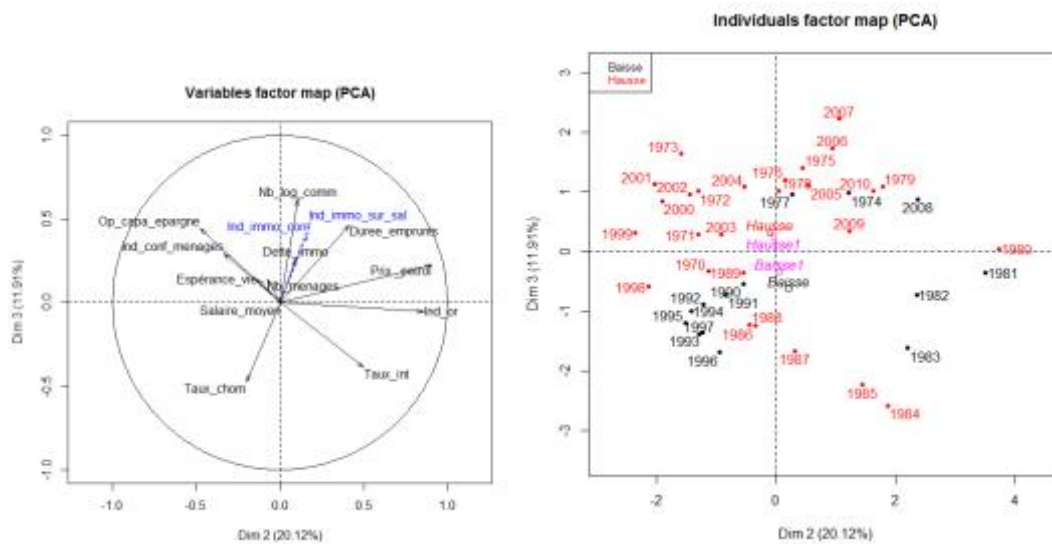
La dimension 2 représente d'avantage les cycles macroéconomiques courts (Pétrol, or, taux d'intérêt).

La dimension 3 est représentative du coût du logement (Nb de logement commencés, durée des emprunts, indices immobiliers) mais également des cycles économique locaux (car en France) tel que le chômage.

On va voir par la suite que cette structure va se retrouver dans l'ensemble des méthodes d'analyse.



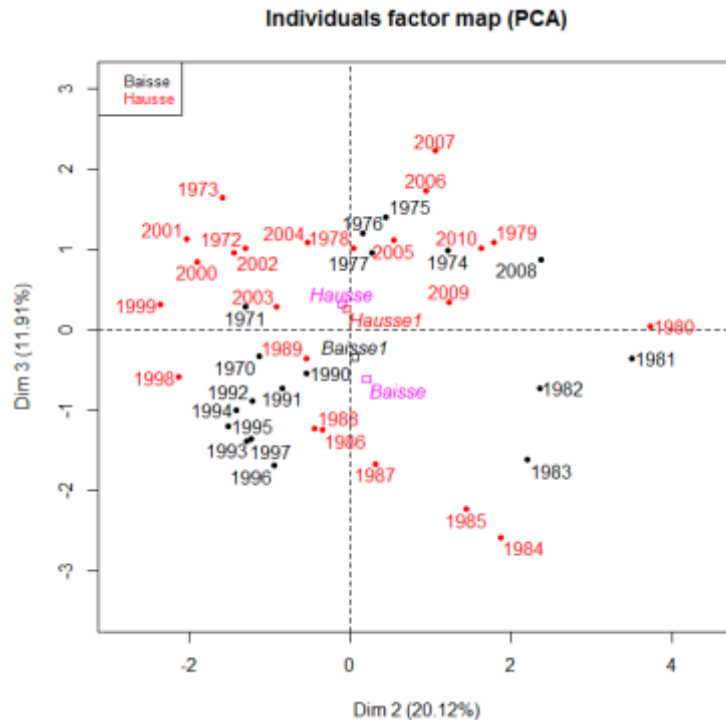
Graph de l'ACP pour les dimensions (1,2)



Graph de l'ACP pour les dimensions (2,3)

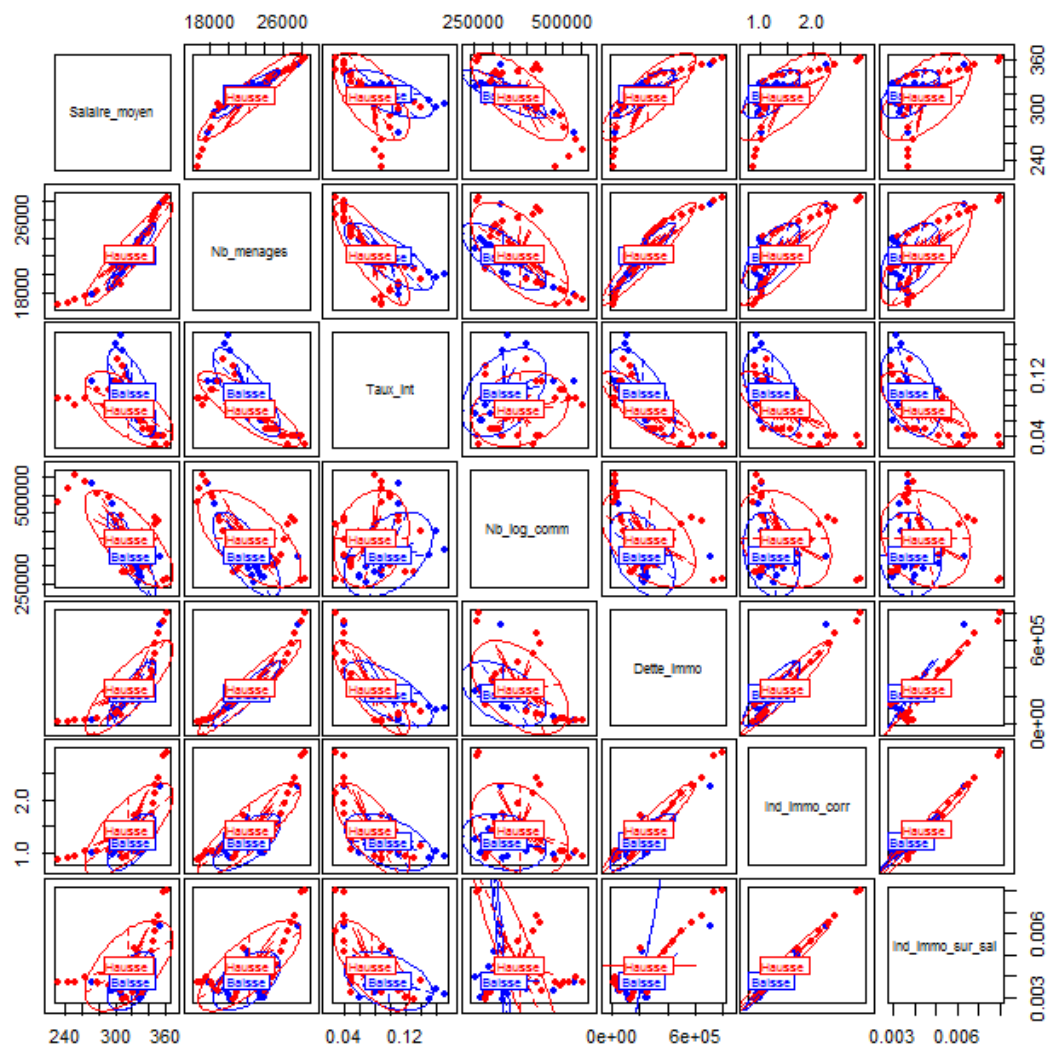
Sur ce dernier graphique, on remarque en effet que les baisses (basées sur l'indice immobilier corrigé de l'inflation) sont nettement en négatif sur la dimension 3. Mais la discrimination est loin d'être parfaite...

Si on utilise un paramètre qualitatif de Baisse/Hausse basé sur la variable Indice immobilier corrigé de l'inflation / par le salaire moyen (euro constant), on obtient



Ce qui n'améliore pas la discrimination sur la dimension 3. Néanmoins cette variable est intéressante car elle correspond à la mesure du pouvoir d'achat immobilier des ménages. Afin de mieux comprendre le lien entre les variables, on va procéder à une étude des analyses des colinéarités. Pour cela on va tracer une variable en fonction d'une autre et dessiner l'intervalle de confiance pour les ensembles d'années de hausse et de baisse.

```
F7010 <- read.table("G:/STA 201/Tabs/F7010.csv", header=TRUE,
  sep=";", na.strings="NA", dec=".", strip.white=TRUE, row.names=1)
library(ade4)
par(mar=c(0,0,0,0))
bivar<- function(x,y,...) {
  xy<- cbind.data.frame(x,y)
  s.class(xy, F7010$Mvt_immo, include.ori = F, add.p = T, clab=1,
  col=c("blue","red"), cpoi=2,csta=0.5) }
pairs(F7010[,c(1:5,13,14)], panel=bivar)
```



2. Analyse et choix des variables étudiées et méthodes

Dans un premier temps, on représente les 5 premières variables par rapport aux deux variables que l'on veut étudier, à savoir Indice immo corrigé de l'inflation (Indice_immo_corr_inf) et l'indice immo corrigé de l'inflation / salaire (eur const (ind_immo_sur_sal)). On remarque tout de suite que la Hausse/baisse n'est pas franchement discriminée par aucune variable (ce que l'on a vu en ACP). De plus les variables Salaire moyen et Nb menage sont très (trop !) corrélées linéairement et ne marquent que le passage du temps. Nous éliminerons le salaire moyen dans la suite de cette étude (également pour le fait qu'il est utilisé pour construire des variables composites).

De plus, le graphe « **Indice du prix logements, Paris vs. Taux d'intérêt long terme** » est très intéressant car il montre une relation vaguement linéaire à tendance décroissante. C'est entre autre cette relation que l'on souhaite étudier. Ce graphique va faire l'objet d'une étude spécifique dans le paragraphe suivant.

Pour les mêmes raisons de colinéarité excessive, nous éliminerons du reste de l'étude l'espérance de vie, la dette immobilière (également pour la raison que cette variable est probablement d'avantage une conséquence qu'une cause) et l'indicatrice « indice immobilier (corrigé de l'inflation) divisé par le salaire moyen » (en EUR const). En effet ce dernier indice n'est pas suffisamment différenciant par rapport à l'indicatrice « indice immobilier corrigé de l'inflation » (cela sous-entend que le pouvoir d'achat immobilier suit simplement le prix de l'immobilier (corrigé de l'inflation) et même plutôt l'inverse, ou dit autrement, les ménages investissent toujours au maximum de leur possibilité en immobilier).

Je souhaite également dans cette étude voir s'il est possible de mettre en forme les données sources de façon plus pertinentes afin d'améliorer les résultats des différentes analyses multivariées. Ces données sont présentées ci-dessous.

2.1. Approche différentielle pour la variation des prix de l'immobilier par rapport au taux d'emprunt : VF7010

On peut s'attendre à ce que le lien entre ces deux variables soit prépondérant dans cette étude et de plus, on peut apporter de l'information sur ce lien. En effet, dans une situation de prix stable et tout étant égal par ailleurs, on peut considérer que le prix immobilier est fait par la capacité d'emprunt des acheteurs elle-même étant conditionnée par la limite des annuités que l'acheteur peut rembourser (règle des 33% du crédit toujours en vigueur chez les banquiers). Si on regarde :

$$dK = 2 \cdot a \cdot \tau \cdot d\tau \quad (\text{tq } a < 0)$$

Cela indique qu'il est préférable d'étudier la variation du taux d'intérêt composite (différence des taux entre n et n-1 moyenné sur les 6 dernières années * le taux d'emprunt annuel n) et la variation des indices immobiliers corrigé de l'inflation. S'inspirant de ce constat, un jeu de données de variation des variables (plutôt que la variable elle-même) va être créé

2.2.Résumé des variables à étudier

Cette première partie conduit donc à l'étude de deux jeux de données :

-les variables brutes

- les variables issues des variations de variables brutes (mais uniquement lorsque cela à un sens. En effet, pour certaines variables qui sont le résultat d'une variation de flux (ex du prix du pétrol) on peut considérer qu'elles sont déjà des variables du type « variation ».

Ce tableau résume les deux jeux de variable qui vont être utilisés par la suite.

A ce niveau de l'étude, on espère que les analyses et prédictions donneront de meilleurs résultats sur les données VF7010 que sur les simples données brutes F7010...

2.2.1.Méthodes utilisées

Pour chaque jeu de données (F7010 et VF7010), il y aura quatre principaux axes d'étude :

1. Description multivariée par ACP et ACM (equivalent à l'AFC sur tableau disjonctif complet).
Les fonctions utilisées seront celles de Factominer
2. Prédiction discriminante par PLS (fonction PLSDA de Factominer) et Disqual (Fonction de Disriminer)
3. Prédiction de variable quantitative par PLS et PCR (Fonction PLSR et PCR du package pls)
4. Autoregressivité temporelle des données et analyse fonctionnelle (fonctions du package fda) qui seront regroupées à la fin de l'étude.

3. Données brutes F7010

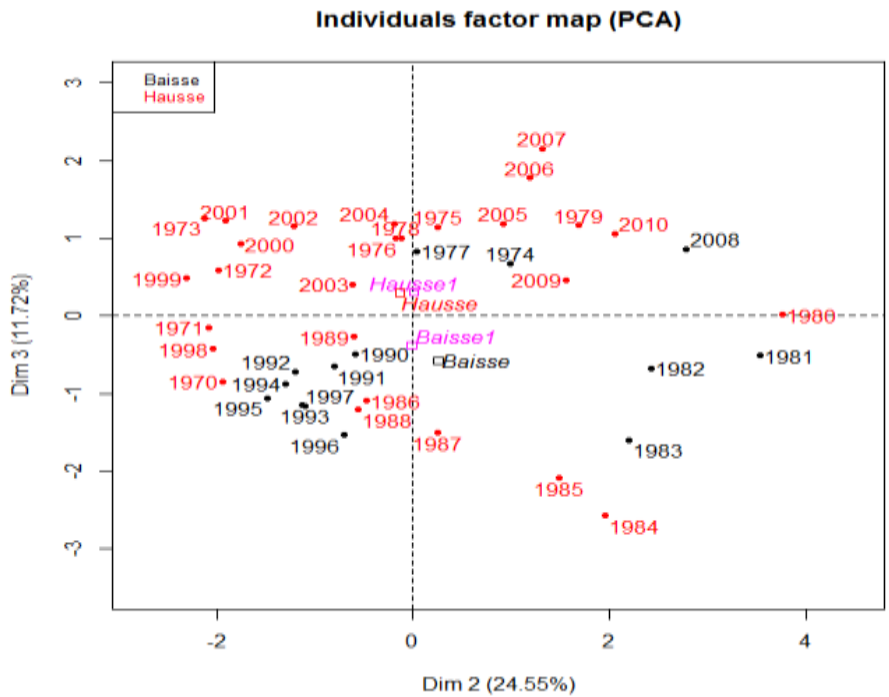
3.1. Description multivariée

3.1.1. Analyse ACP des données brutes centrées réduites

```

###Import données
F7010 <- read.table("G:/STA201/Tabs/F7010.csv", header=TRUE,
  sep=";", na.strings="NA", dec=".", strip.white=TRUE, row.names=1)
summary(F7010)
#-----
#pca avec package Factominer
resF7010<-PCA(F7010, ncp=Inf, quanti.sup=c(12,13), quali.sup=c(14:15))
plot.PCA(resF7010, axes=c(1,3), habillage = 14)
round(resF7010$eig,2)

```



On retrouve une discrimination intéressante le long de la dimension 3 entre les années de hausse et de baisse au sens de l'indice immobilier corrigé de l'inflation. Cependant, on remarque des années différentes des autres années de baisse comme 2008 (1 an après le début du crack boursier) 1974 (1 an après le premier choc pétrolier) et 1977 (une année riche en crise géopolitique mais je n'ai pu identifier un fait particulier). Inversement, il y a également beaucoup d'année de hausse fortement négative selon la dimension 3 comme les années de 1984 à 1989 (peut-être expliquée par la libéralisation et la déréglementation des marchés).

Et pour ce qui concerne l'inertie des axes :

	Eigen value	percentage of variance	of	cumulative percentage of variance
--	-------------	------------------------	----	-----------------------------------

comp 1	5.51	50.09	50.09
comp 2	2.70	24.55	74.63
comp 3	1.29	11.72	86.36
comp 4	1.00	9.10	95.46
comp 5	0.20	1.78	97.24

La moitié de l'inertie est contenue dans l'axe 1 et les 4 premiers axes contiennent donc plus de 95% de la variabilité du nuage d'individu. Nous n'étudierons pas au-delà. Si l'on regarde maintenant les contributions des variables à chaque axe ainsi que la qualité de représentation de ces derniers.

	Contribution				Cos ²			
	Dim.1	Dim.2	Dim.3	Dim.4	Dim.1	Dim.2	Dim.3	Dim.4
Salaire_moyen	14.68	1.05	1.25	12.60	0.81	0.03	0.02	0.13
Nb_menages	16.41	0.25	4.96	0.72	0.90	0.01	0.06	0.01
Nb_logements_com	9.81	0.01	20.37	15.63	0.54	0.00	0.26	0.16
Indice_or_const	3.77	24.53	0.58	6.52	0.21	0.66	0.01	0.07
Prix._petrol_const	0.07	32.63	4.89	0.70	0.00	0.88	0.06	0.01
Taux_chomage	13.59	0.11	9.75	7.71	0.75	0.00	0.13	0.08
Confiance_menages	7.70	6.30	5.21	24.46	0.42	0.17	0.07	0.24
Opinion_capacite_epa	3.77	9.90	14.83	25.68	0.21	0.27	0.19	0.26
Duree_emprunts	7.05	10.34	23.07	0.12	0.39	0.28	0.30	0.00
Inflation_France	13.36	8.43	0.05	0.92	0.74	0.23	0.00	0.01
Taux._int	9.79	6.45	15.03	4.94	0.54	0.17	0.19	0.05

On remarque que l'on retrouve bien les résultats de la PCA préliminaire. La dimension 1 marque l'axe du temps et les modifications en tendances longues de la société, la dim2 marque les variations courts terme (Pétrol et Or), la dimension 3 est l'axe de l'immobilier et le 4ieme axe pourrait être interprété comme un axe de la confiance des ménages. Mais il faut également regarder la représentativité des variables sur les dimensions. Malheureusement, les variables explicatives de la dimension 3 (et 4) sont très mal projetées et on ne peut donc pas considérer que la dimension 3 (respectivement 4) et représentative des variables qui la compose.

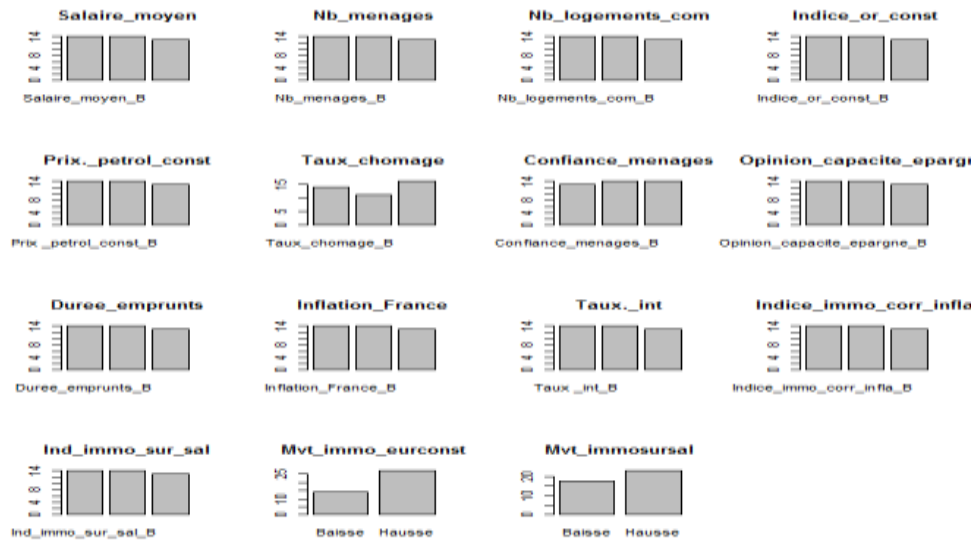
En conclusion, on remarque une discrimination intéressante mais selon un axe peut représentatif (ni en inertie, ni des variables qui le compose). Voyons si l'on peut obtenir d'avantage avec une ACM

3.1.2.ACM sur données brutes F7010

Dans un premier temps, on crée une table des modalités (F7010-ACM). Ces modalités sont définies automatiquement en divisant toutes les variables explicatives en trois classes de nombre d'individus identique (_B pour Bas, _N pour neutre et _H pour haut). Cela assure une bonne répartition des classes (1/3,1/3,1/3) mais il faut rester vigilant car cela rend encore plus dépendant au groupe d'individus choisi.

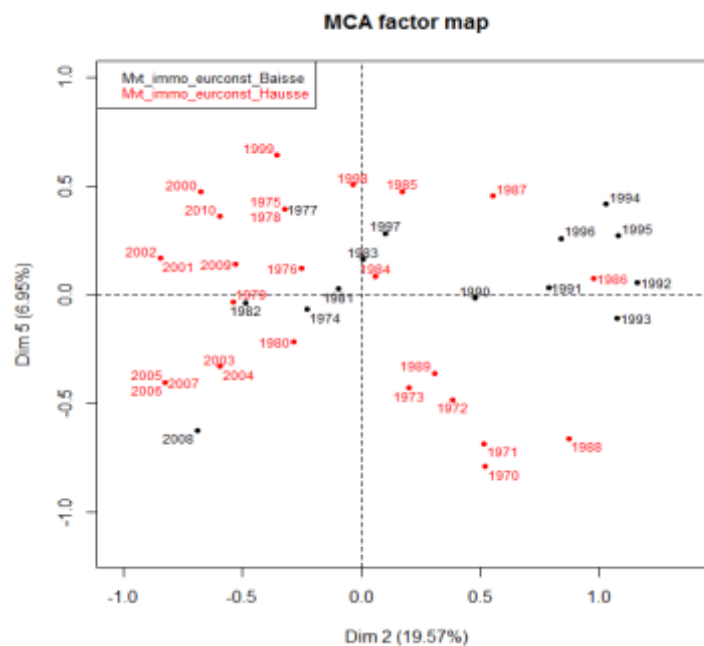
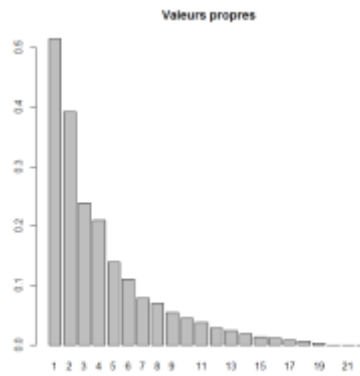
On peut ensuite soit créer le tableau disjonctif (fonction de ade4) et procéder à une AFC ou étudier directement le tableau des modalités par ACM (fonction MCA). C'est ce second choix que nous retiendrons ici. Voyons d'abord la répartition des modalités dans chaque variable:

```
# ACM
F7010 <- read.table("G:/STA201/Tabs/F7010-ACM.csv", header=TRUE,
  sep=";", na.strings="NA", dec=".", strip.white=TRUE, row.names=1)
par(mfrow=c(4,4))
for (j in 1:16) {plot(F7010[,j],main=colnames(F7010)[j])}
layout(1)
```



Les graphiques représentant la répartition des valeurs propres issues de l'ACM ainsi que le Factor's Map sont obtenus par :

```
F7010ACM<-MCA(F7010,ncp=3, quali.sup=c(1,11,12,14:17), graph=F)
barplot(F7010ACM$eig[,1], main="Valeurs propres", names.arg=1:nrow(F7010ACM$eig))
plot(F7010ACM,cex=0.75,choix="ind",invisible=c("var", "quali.sup"), axes=c(2,3), habillage=14)
```



On remarque une bonne discrimination des années de baisse (dans le coin en haut à droite) dans le plan (2,5). On remarque également toujours la particularité de l'année 2008. On peut cependant remarquer que l'inertie de la dimension 5 est faible et qu'il ne va pas être aisé de discriminer sur si peu de variance...

	Eig Val	% of variance	Cumulative	% of variance
dim	1	0.51	25.71	25.71
dim	2	0.39	19.57	45.28
dim	3	0.24	11.94	57.22
dim	4	0.21	10.46	67.68
dim	5	0.14	6.95	74.62

On a donc pour l'ACM sur données brutes la même conclusion que pour l'ACP : une discrimination intéressante (mais loin d'être parfaite) et une inertie faible qui pousse à la méfiance quand à l'utilisation des résultats.

3.2.Prédiction

Dans cette partie, on va mettre en œuvre deux méthodes de prédiction discriminante (PLSDA et DisQual package Discriminer) et deux méthodes de prédiction quantitative (PCR et PLSR, package pls).

On a mis en évidence lors de la description que l'année 2008 est particulière. Afin d'améliorer la cohérence des données, nous l'éliminerons cet individu dans la partie prédiction.

3.2.1. PLS-DA sur données brutes F7010

Nous allons réaliser une PLS discriminante. C'est une PLS sur composante principale lorsque la variable cible est qualitative (ici deux modalités : Baisse et Hausse). Nous allons devoir utiliser une partie des données pour réaliser la PLS et le reste pour réaliser un test et voir la performance de la prédiction. La technique de base pour réaliser cela est le LOO (leave one out) mais je souhaite être plus général et couvrir d'avantage de cas de figure. J'ai développé pour cela un code permettant de tester différentes configurations de liste Learn et Test. Ces listes sont générées aléatoirement. La liste Test sera composée de 4 à 10 années (sur 41 années au total pour F7010 et 40 pour VF7010). Le reste sera la liste Learn. Je procède à 100 tirages différents pour chaque taille de liste Test et je fais cela pour un nombre de composantes principales allant de 2 à 10. Enfin, je récupère l'erreur calculée pour chaque tirage et j'en calcul la moyenne et la variance.

```
F7010<-F7010[-c(38),-c(1,12,13,15)] #Mvt_immo_euroconst devient colonne 11
mat_err<-matrix(nrow=10,ncol=3)
for (j in 2:10) { #boucle sur le nb de composantes de la PLS DA
  list_err<-c()
  for (x in 4:10) { #nb de Leave out -> test
    for (k in 1:10) { #boucle pour realiser plusieurs tirages des listes
      learnbench<-c()
      testbench<-c()
      for (i in 1:nrow(F7010)){ #création de la liste des numero de ligne
        learnbench<-c(learnbench,i)}
      testbench<-sample(learnbench,x) #distribution des test et learn
      learnbench<-setdiff(learnbench,testbench)
      F7010Learn<-F7010[learnbench,]
      F7010Test<-F7010[testbench,]
    }
    #calcul PLS DA
    resplsda<-plsDA(F7010[,-c(11)],F7010[,11],autose1=FALSE, comps=j,
    validation="learntest", learn=learnbench, test=testbench)
    #Création de la liste des erreurs succécives
    list_err=c(list_err,resplsda$error_rate)
  } }#boucle for nb de tirage test et boucle for plusieurs tirages
#remplissage de la matrice des resultats par composante
mat_err[j,1]<-j
mat_err[j,2]<-mean(list_err)
mat_err[j,3]<-var(list_err)
} #fin boucle sur composantes j
round(mat_err,2)
```

On obtient alors :

	nb composantes	Moy du ratio d'erreur	Var du ratio d'erreur
[1,]	NA	NA	NA
[2,]	2	0.32	0.03
[3,]	3	0.28	0.03

[4,]	4	0.26	0.03
[5,]	5	0.27	0.03
[6,]	6	0.26	0.02
[7,]	7	0.26	0.03
[8,]	8	0.25	0.03
[9,]	9	0.26	0.02
[10,]	10	0.26	0.03

On remarque que pour 3 composantes le ratio d'erreur (error_rate) est de 28% et la variance est très faible. C'est un résultat encourageant au vu du ratio « aléatoire » de 50% auquel on aurait pu s'attendre.

Par exemple, avec 3 composantes et la liste des années choisies en test suivante :

Testbench (aléatoire)

[1] 34 31 39 19 16 2 23 37 9 30

On obtient la Matrice de confusion suivante :

	predicted	
original	Baisse	Hausse
Baisse	1	0
Hausse	1	8

Soit un error_rate de 10% seulement. Nous allons voir s'il est possible d'obtenir une prédiction plus précise avec une autre méthode.

3.2.2. DisQual

La méthode DisQual (G.Saporta) est une discrimination linéaire faite sur la décomposition en vecteur de l'ACM.

On retire comme précédemment de l'étude l'année 2008 ainsi que la variable « salaire », indice_immo et Mvt_immo car se sont toujours les variables que l'on cherche à prédire.

Dans la méthode DisQual de Discriminer, on ne choisit pas le nombre de composante conservée pour la discrimination. En faisant, comme précédemment varier les Leave K out (de 1 à 10 années) et en réalisant 100 tirages pour chaque K, on obtient en moyenne l'erreur d'affectation suivante :


```

list_err<-c()
for (x in 1:10) { #nb de Leave out -> test
  for (k in 1:100) { #boucle pour realiser plusieurs tirages des listes
    learnbench<-c()
    testbench<-c()
    for (i in 1:nrow(F7010)){ #création de la liste des numero de ligne
      learnbench<-c(learnbench,i)}
    testbench<-sample(learnbench,x) #distribution des test et learn
    learnbench<-setdiff(learnbench,testbench)
  }#calcul DisQual
  resdisq<-disqual(F7010[,c(1:10)],F7010[,c(10)], validation="learntest" , learn=learnbench,
  test=testbench)
  #Création de la liste des erreurs succécives
  list_err=c(list_err,resdisq$error_rate)
} }#boucle for nb de tirage test et boucle for plusieurs tirages
mean(list_err)
var(list_err)

```

```

> mean(list_err)
[1] 0.1129217
> var(list_err)
[1] 0.04927242

```

Ce qui est encore mieux qu'avec la PLS DA. 11% d'erreur sur l'affectation des années dans l'un des deux groupes Hausse/ Baisse est vraiment un résultat étonnamment bon. En y réfléchissant, la définition des modalités lisse les différences au sein des variables et met en avant les grandes tendances ainsi que des liens non linéaires. On avait en effet remarqué une bonne discrimination des années de baisse en ACM selon les dimensions 2 et 5 et cela corrobore bien les résultats de la méthode DisQual.

Après avoir fait de la discrimination selon les deux modalités Hausse et Baisse on peut essayer de voir si l'on peut prédire quantitativement la hausse ou la baisse par les deux méthodes les plus communes de prédiction que sont la PCR et la PLS.

3.2.3.PCR données brutes

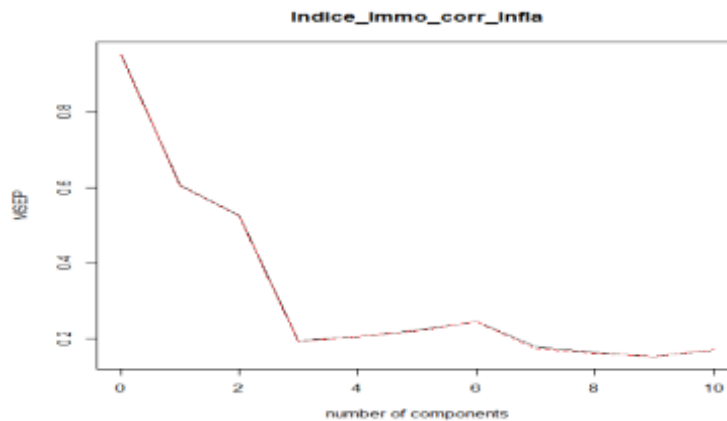
La méthode PCR se déroule en trois phases :

5. Analyse en composantes principales sur les variables explicatives
6. Régression linéaire de la variable indépendante sur la partie des composantes la plus corrélée à celle-ci
7. Calcul des paramètres de la régression en fonction des variables d'origines

Nous utiliserons ici la fonction PCR du package pls.

```
#on retire l'année 2008 n°38 et les colonnes qualitative des groups (Hausse/baisse)et la
colonne des années
F7010<-F7010[-c(39),-c(1,13:15)]
#avec LOO
respcr.loo<-pca(Indice_immo_corr_infla ~ .,data=F7010, validation="LOO")
validationplot(respcr.loo,val.type="MSEP")
```

On obtient le graphique suivant. On remarque une convergence rapide par cross validation LOO de l'erreur moyenne quadratique pour 3 composantes. Cela correspond bien à ce que l'on a pu voir lors de l'étude ACP où la dimension 3 était la variable importante concernant l'immobilier.



Si on applique de nouveau la technique de cross validation vue précédemment (tirage aléatoire de 4 à 10 années en test, tiré 100 fois pour chaque) on obtient de nouveau une erreur moyenne quadratique ainsi que la variance.

Résultats :

	[,1]	[,2]	[,3]
[1,]	1	0.61	0.16
[2,]	2	0.53	0.10
[3,]	3	0.24	0.03
[4,]	4	0.21	0.02
[5,]	5	0.24	0.02
[6,]	6	0.25	0.02
[7,]	7	0.21	0.02
[8,]	8	0.17	0.01
[9,]	9	0.17	0.01
[10,]	10	0.20	0.01

On confirme bien la convergence pour 3 dimensions. L'erreur de 0,24 centré-réduit correspond à une erreur de 1,5 sur l'indice immobilier d'origine alors qu'il était de

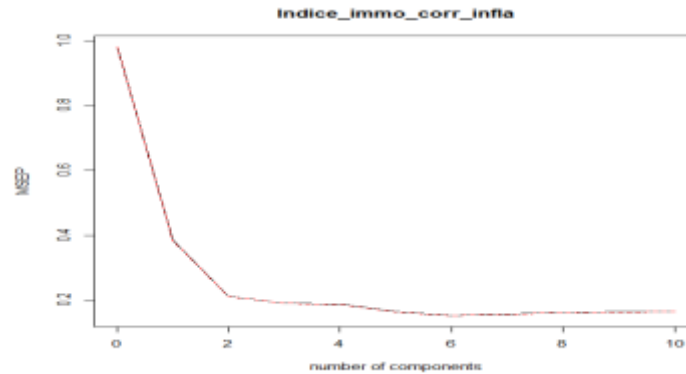
2009 2,15
2010 2,37

Soit une erreur de plus de 60% et quasiment systématique vu la faiblesse de la variance. Cette estimation n'est donc pas exploitable.

3.2.4. PLSR données brute F7010

PLSR est un algorithme itératif cherchant à maximiser la colinéarité de variables latentes et de la variable à prédire. Les paramètres de la fonction sont similaires à ceux de la PCR et nous ferons donc la même étude.

Si l'on trace le plot de validation MSE de la crossvalidation LOO pour la PLS, on constate bien une convergence rapide dès 2 composantes.



Comme précédemment, on utilise la même boucle pour créer notre propre cross validation.

La fonction PLSR est mise en œuvre comme la PCR.

La cross-validation « maison » donne :

	[,1]	[,2]	[,3]
[1,]	1	0.39	0.05
[2,]	2	0.22	0.01
[3,]	3	0.19	0.01
[4,]	4	0.19	0.01
[5,]	5	0.19	0.02
[6,]	6	0.17	0.01
[7,]	7	0.17	0.01
[8,]	8	0.19	0.02
[9,]	9	0.18	0.01
[10,]	10	0.19	0.01

On retrouve des résultats du même ordre de grandeur qu'avec le PCR. En effet, pour 3 composantes, on fait une erreur moyenne de prédiction de 1,47 point d'indice immo (de 2,37 en 2010) ce qui est au vu des variances et du nombre de tirage équivalent en résultat à la méthode PCR et donc également non exploitable dans l'optique de faire une prédiction du cours de l'immobilier futur. Nous allons alors tenter d'étudier les données VF7010 de la variation de variable en espérant améliorer le niveau prédictif des différentes méthodes.

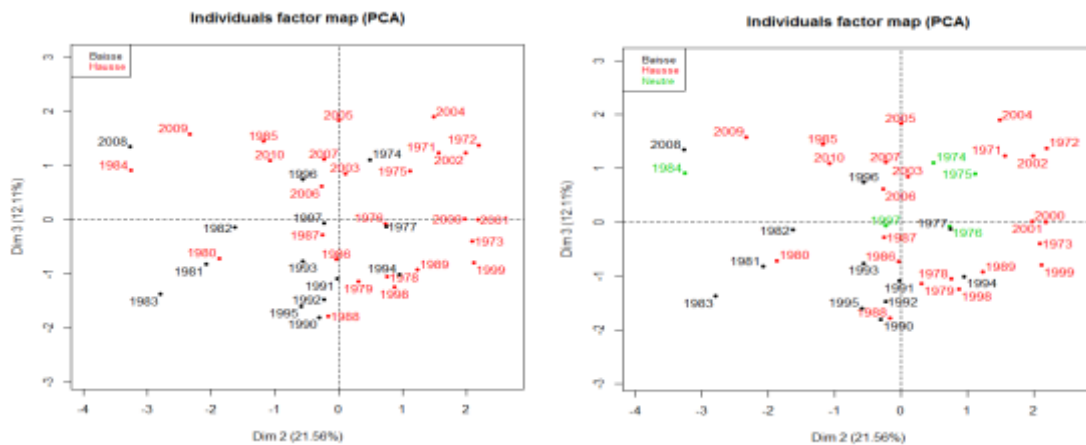
4. Variation des données brutes VF7010

Cette partie reprend entièrement la forme et les codes R de l'analyse des données brutes. Seuls les résultats seront présentés.

4.1. Description multivariée

4.1.1. ACP de la variation des données brutes VF7010

Visuellement, on retrouve la meilleure discrimination pour les dimensions (2,3). Cependant elle n'apparaît pas meilleure que celle de l'étude sur les données brutes. On teste également la création d'une classe Neutre pour les années où la hausse/Baisse n'a pas été franche (graph de droite)



La classe neutre ne se retrouve pas suffisamment près du zéro des axes et ne change donc pas réellement la qualité de discrimination. Nous ne l'utiliserons plus dans le reste de l'étude. S'il on regarde les contributions aux axes :

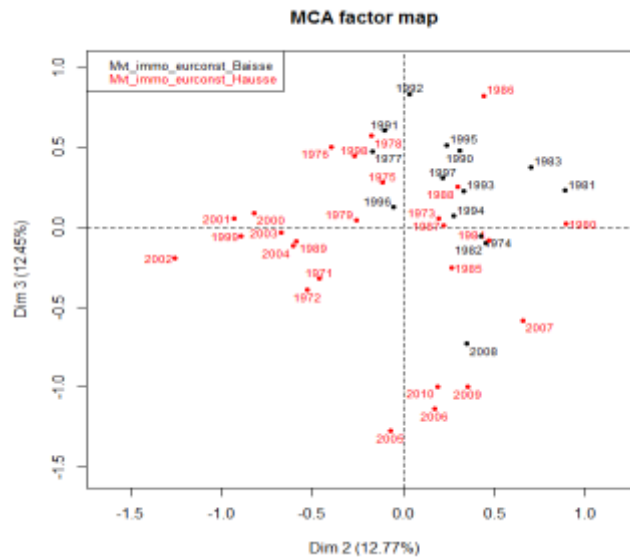
	Dim. 1	Dim. 2	Dim. 3	Dim. 4	Dim. 5	Dim. 6	Dim. 7	Dim. 8	Dim. 9
VNb_menages	12.16	1.01	19.74	3.35	25.86	3.31	5.58	0.02	21.65
VNb_logements_com	1.85	15.91	0.01	24.08	11.06	44.13	1.15	0.35	1.28
VIndice_or_const	13.45	13.22	0.98	1.78	13.55	0.59	7.43	24.32	20.75
VPrix._petrol_const	0.25	1.42	30.13	47.16	8.92	8.52	2.37	0.79	0.00
VTaux_chomage	21.19	2.56	1.53	0.00	11.73	1.19	9.65	0.01	17.03
VConfiance_menages	6.94	20.33	14.23	4.21	2.23	0.01	2.98	28.50	20.13
VOpinion_capacite_epargne	4.28	25.38	9.39	6.13	0.11	3.12	11.51	37.37	1.13
VInflation_France	24.35	1.27	0.66	0.64	2.48	0.35	7.09	1.38	16.28
Vtaux_int_compo	15.34	6.32	0.77	0.59	11.71	0.02	50.89	7.22	1.74
VDuree_emprunts	0.20	12.58	22.57	12.06	12.35	38.76	1.36	0.03	0.00

On remarque que le taux d'intérêt composite (que l'on souhaiterait voir discriminer la hausse et baisse) contribue peu aux axes 2 et 3. Il contribue beaucoup à l'axe 1 qui n'est pas discriminant et à d'autres axes qui représentent peu d'inertie et qui ne sont pas discriminants non plus.

Il semble donc que la présentation des données sous la forme de la variation des données sources (sorte de dérivée) et la création de la variable composite du taux d'intérêt n'apporte pas à l'analyse discriminante des années de hausse/baisse de l'immobilier corrigé de l'inflation.

1. ACM sur données VF7010

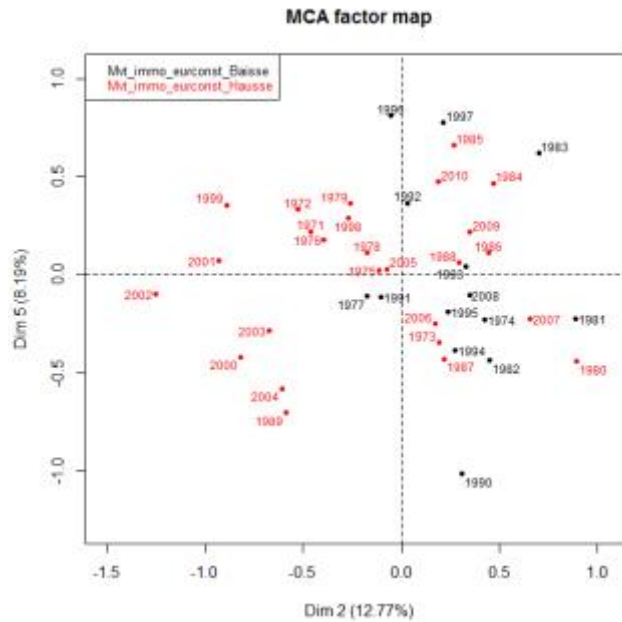
Est réalisé sur les données V7010 le même traitement automatique en décile de 33% pour créer les modalités de la table V7010-ACM.



S'il on regarde les contributions et les qualités de représentations, les axes 2 et 3 sont fortement discriminants. Leurs principaux contributeurs sont la confiance et l'opinion sur la capacité d'épargne ainsi que la variation du taux d'intérêt composite lorsqu'il est neutre.

	Dim1	Dim2	Dim3	Dim4	Dim5
VnbMenage_B	10.90	0.09	4.86	1.01	0.01
VnbMenage_H	1.47	3.15	10.38	3.24	0.61
VnbMenage_N	4.18	4.61	1.30	8.27	0.54
Vtauxintcompo_B	0.59	1.23	6.33	2.68	12.86
Vtauxintcompo_H	0.23	2.68	0.00	2.79	18.63
Vtauxintcompo_N	0.08	7.88	6.22	0.01	0.80
Conf_men_B	0.50	0.02	11.12	2.19	2.99
Conf_men_H	4.86	6.30	1.20	3.29	0.60
Conf_men_N	2.49	7.58	4.83	0.16	0.85
Op_capa_epar_B	1.01	4.15	7.70	3.61	6.97
Op_capa_epar_H	2.83	9.46	1.89	1.58	0.58
Op_capa_epar_N	0.55	1.33	1.82	0.35	11.78

On trouve également que Taux composite contribue fortement à la 5ieme dimension.



Malheureusement il n'est pas plus discriminant en Baisse/Hausse que l'axe 2. Cependant il faut souligner qu'il « normalise » des années exotiques dans les autres études (comme 2008). Voyons si le choix de cette présentation des variables (VF7010) améliore le niveau de prédiction.

4.2.Prédiction

Même si l'analyse graphique des données VF7010 n'a pas montré de réelles avancées quant à la discrimination des années de Baisse par rapport aux années de Hausse, nous allons malgré tout dérouler l'analyse discriminante des 4 méthodes appliquées lors de l'étude des données F7010.

4.2.1.PLS-DA sur données VF7010

On réalise de nouveau l'étude de sensibilité au nombre d'axes utilisé pour réaliser la PLS discriminante.

On obtient le ratio d'erreur moyen suivant (par nombre de composantes)et la variance :

```
> round(moy,2)
```

```
[1] NA 0.31 0.31 0.29 0.26 0.26 0.29 0.25 0.25 0.22
```

```
> round(sd,2)
```

```
[1] NA 0.18 0.15 0.15 0.15 0.17 0.14 0.18 0.15 0.16
```

On a par exemple 29% d'erreur si on utilise 4 composantes et une variance moyenne de l'erreur de 0,15. Ce qui n'est pas meilleur que lors de l'utilisation des données brutes.

4.2.2.DisQual

Le même code que pour les données F7010 est utilisé ici. 1000 tirages sont réalisés pour chaque répartition Learn/Test (de 4 à 10 années misent en test).

On obtient :

```
> round(mean(list_err),2)
```

```
[1] 0.13
> round(sd(list_err),2)
[1] 0.17
```

Soit en moyenne un ratio d'erreur de 13% sur la table de confusion avec une variance importante. Encore une fois, ce résultat est moins bon que celui obtenu avec les variables brutes F7010.

4.2.3.PCR données brutes

A nouveau, convergence de 100 tirages pour chaque distribution de 4 à 10 années dans le panier des tests. Toutes les erreurs quadratiques (MSE) sont conservées puis on en calcul la moyenne et la Var pour chaque nombre de composantes ayant servi pour le calcul de la PCR.

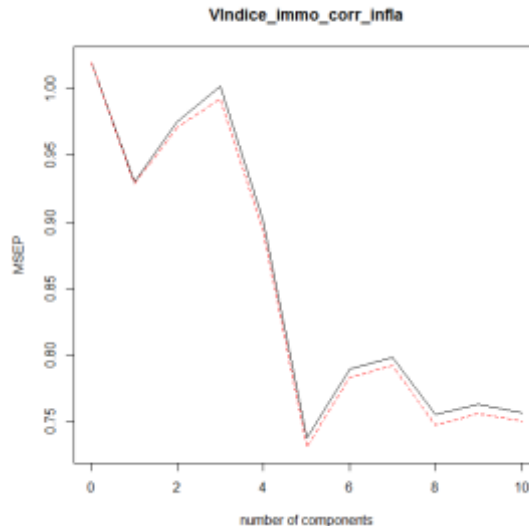
```
> round(matMSE,2)
```

	Nb composante	Mean	Var
[1,]	1	0.97	0.23
[2,]	2	0.88	0.24
[3,]	3	0.92	0.33
[4,]	4	1.10	0.39
[5,]	5	1.00	0.32
[6,]	6	1.27	0.71
[7,]	7	1.07	0.48
[8,]	8	1.03	0.63
[9,]	9	0.90	0.61
[10,]	10	0.86	0.25

On remarque que le nombre de composante pour le calcul de la prédiction PCR influence peu l'erreur quadratique moyenne (et la var). D'après ce résultat, on peut donc garder 2 composantes. Une erreur moyenne de la variable centré-réduite de 0,88 correspond à une erreur de +-11,8% sur l'estimation du mouvement de l'indice immobilier (équivalent du prix de l'immobilier). Par exemple, si la PCR prédit un mouvement de l'immobilier de 7%, il faut s'attendre à ce que cela soit en moyenne entre -3.8% et 18.8%. Ce qui est bien évidemment une erreur beaucoup trop importante pour pouvoir exploiter la prédiction.

4.2.4.PLSR données VF7010

L'étude LOO de la PLS des données VF7010 nous donne l'erreur quadratique selon le nombre de composantes utilisées suivant :



La recherche par l'algorithme « maison » de l'erreur quadratique (par nb de composantes) converge difficilement. C'est le signe que le modèle n'est pas stable. Il est extrêmement dépendant des vecteurs Learn et Test. On ne peut rien conclure si ce n'est que l'approche pls quantitative ne donne rien sur les données VF7010.

4.2.5. Conclusion à l'analyse multivariée

Cette étude aura permis de tirer des conclusions concernant les données macro-économiques et la possibilité de prédire le cours futur de l'immobilier en faisant des hypothèses sur ces variables.

1. Le résultat de prédiction sur les données brutes est étonnamment bon. Le taux d'erreur de 11% du classement en année de Hausse/Baisse sur les données brutes obtenu par DisQual est positif et permet de mieux mettre en avant le côté non linéaire des relations. En effet, le résultat sur les modalités est bien meilleur que sur les variables quantitatives.
2. Les prédictions quantitatives ne sont pas exploitables. D'autres facteurs sont à l'œuvre dans l'établissement des cours de l'immobilier (effet spéculatif, faits ponctuels déclenchant une inversion de la tendance...).
3. La tentative d'ajout d'information aux données n'a pas été fructueuse. Cela a rendu les liaisons entre variables moins claires (et linéaires) sans apporter la linéarité dans d'autres relations (taux d'intérêt vs indice immo). Les résultats sur les données VF7010 sont très décevants mais j'imagine que cela fait partie de la vie d'un analyste de données...

J'ai également essayé d'étudier des variables moyennées sur 3 ans mais cela n'a pas donné de meilleurs résultats. Il pourrait-être intéressant de faire de nouveau l'étude avec une plage de donnée plus étendue (1950 à aujourd'hui) et sur d'autres pays ...

Pour finir, on peut tout de même dire que la variation des prix de l'immobilier ne dépend pas linéairement des variables macro économiques y compris du taux d'intérêt et de la confiance des ménages. On peut dans ce cas je pense dire que le marché immobilier est hautement inefficace (ne serait ce que par manque de liquidité) et que les variations long terme doivent venir en grande partie d'une inertie autoregressive qui surpasse en magnitude les contraintes réelle du marché de l'emprunt et de la richesse réelle des individus.

5. Analyse temporelle

5.1. Test de l'hypothèse d'autoregressivité entre années

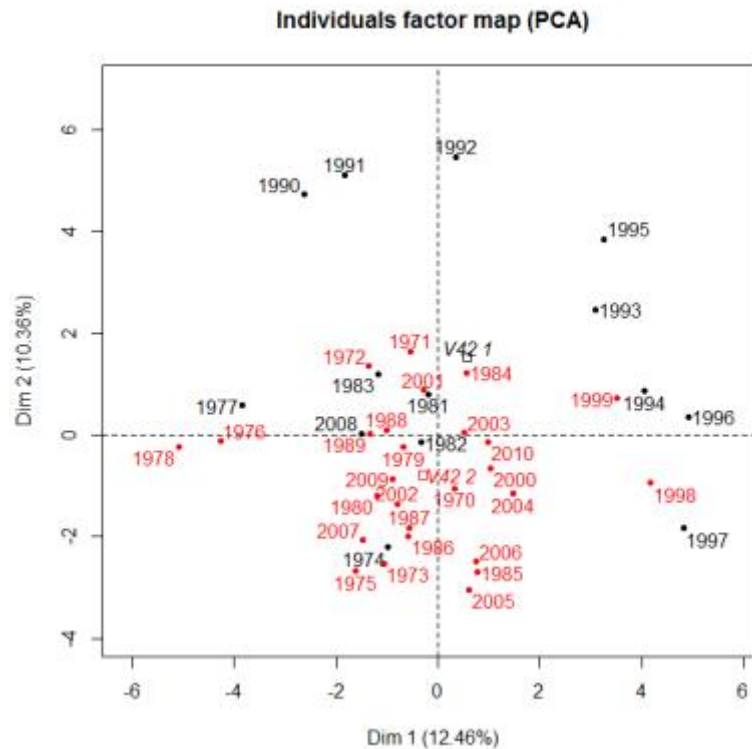
L'idée est d'appliquer la méthode MMDS (ou PCoA) pour représenter les distances entre années et tenter de mettre en avant la plus grande proximité des années successives par rapport aux années plus éloignées. Cela mettra en avant le caractère autorégressif et le fait que les variables d'une année sont toujours proches de celles des années qui la précède ou la suit.

En pratique, on calcule dans un premier temps la distance de Mahalanobis d'une année à toutes les autres (utilisation de la fonction mahalanobis). Cela produit un vecteur de la distance de l'année i à toutes les années. Puis on recrée une matrice de distance en agrégeant (rbind) le résultat pour chaque année i . Enfin on réalise une ACP sur la matrice des distances pour visualiser les distances entre années.

En reprenant les données F7010 (et en retirant la colonne Indice_immo) et en calculant les matrice de corrélation. On peut appliquer le code R suivant :

```
F7010 <- as.matrix(F7010, rownames.force = TRUE)
F7010InvMat <- as.matrix(F7010InvMat, rownames.force = TRUE) #Mat de corrélation Inverse
# MMDS
#-----
#Calcul Mat
TabMaha<-c()
for(i in 1:41) {
  TabMaha<-rbind(TabMaha,mahalanobis(F7010, F7010[i,], F7010InvMat, inverted =
  TRUE, check.margin=TRUE))}
Id<-diag(41)
E<-Id-(1/41)
TabDistMaha<-(-1/2)* E %*% TabMaha %*% E
#on remet la colonne des années en tête
TabDisMaha<-cbind(F7010[,0],TabDistMaha)
#on remet la colonne de hausse/baisse pour l'habillage des graph
TabDisMaha<-cbind(TabDisMaha,F7010origine[,14])
summary(TabDisMaha)
#-----
###ACP par FactoMineR
TabDisMahaPCA<-PCA(TabDisMaha,ncp=Inf, quali.sup=c(42))
plot.PCA(TabDisMahaPCA, axes=c(1,2), choix="ind", habillage=42)
```

Il faut noter que l'on retrouve le même résultat pour les jeux de données F7010 et VF7010



Graph obtenu pour les données F7010

On remarque immédiatement que les années se suivent dans le plan ce qui tend à confirmer le lien autoregressif entre les années. On peut également remarquer que les années de baisses (ici en noir) sont plus loin du centre (en général) ce qui indique que ce sont des années particulières. Enfin, on remarque que certaines années sont en Baisse alors qu'elles semblent peu différentes des années de Hausse qui les entourent.

On peut donc résumer cette information ainsi :

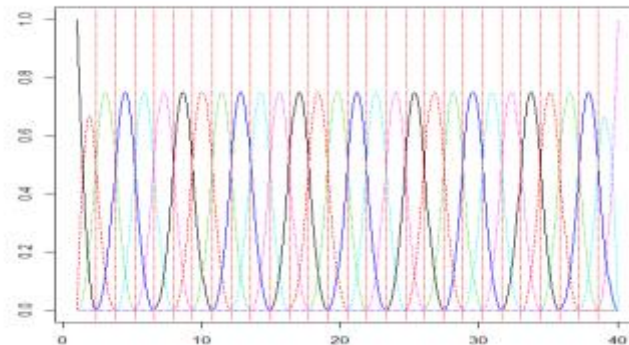
- Les années sont en générales autorégressives et à la hausse
- L'histoire a montré des périodes de Baisse (années 1990) due à un environnement économique particulier mais retournant à la normale à partir de 1998
- Certaines années apparaissent comme des accidents suite à des chocs de confiance (1974 choc pétrolier, 2008 crise des subprime puis crack boursier). Pour les années 1981-1983 je n'ai pas pu identifier un événement en particulier.
- La distance de mahalanobis accorde moins de poids aux variables les plus discordantes. Cela marque les différences d'ensemble et on remarque ainsi que l'année 2008 redevient une année « classique ». En effet elle différait fortement pour le prix du pétrole et l'opinion sur la capacité d'emprunt des ménages ce qui est mitigé ici.

Le cours de l'immobilier est donc plus complexe que les autres types d'actifs. Il dépend bien dans une certaine mesure des sous-jacent macro-économique (taux intérêt notamment) mais il est également autorégressif. Il est intéressant dans ce cas de modéliser les variables par des fonctions continues qui représentent mieux l'évolution temporel d'un phénomène.

5.2. Analyse fonctionnelle

L'analyse fonctionnelle consiste à définir chaque variable comme une courbe constituée d'une base de fonctions (trigonométrique, polynômes...) et de coefficient permettant de « coller » au mieux à la courbe d'origine. C'est ensuite sur le tableau de ces coefficients que nous ferons les analyses multicritères déjà vu dans cette étude.

```
#création d'une base de fonctions (package FDA)
bspl<-create.bspline.basis(c(1,max(t)),nbasis=30, norder=3)
plot(bspl)
```



Nous utilisons ici 30 splines d'ordre 3 (polynôme d'ordre 3 par morceau) ce qui donne par optimisation des moindres carrés, 30 coefficients par variable.

a. F7010

Dans un premier temps, on ajuste au jeu de données F7010 (données brutes) 30 splines par variables par la méthode des moindres carrés. Cette méthode est codée dans le package FDA de R

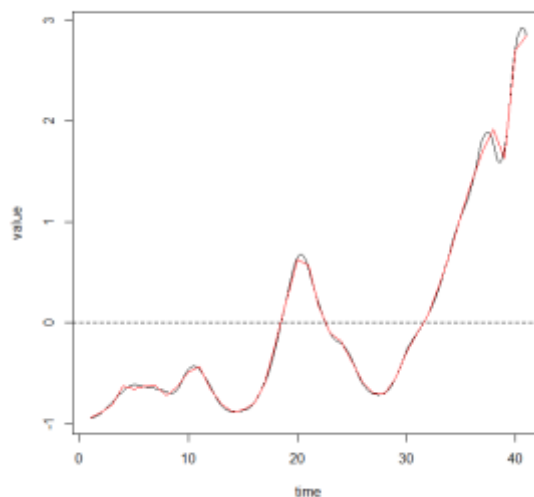
i. FDA (Functional data analysis)

```

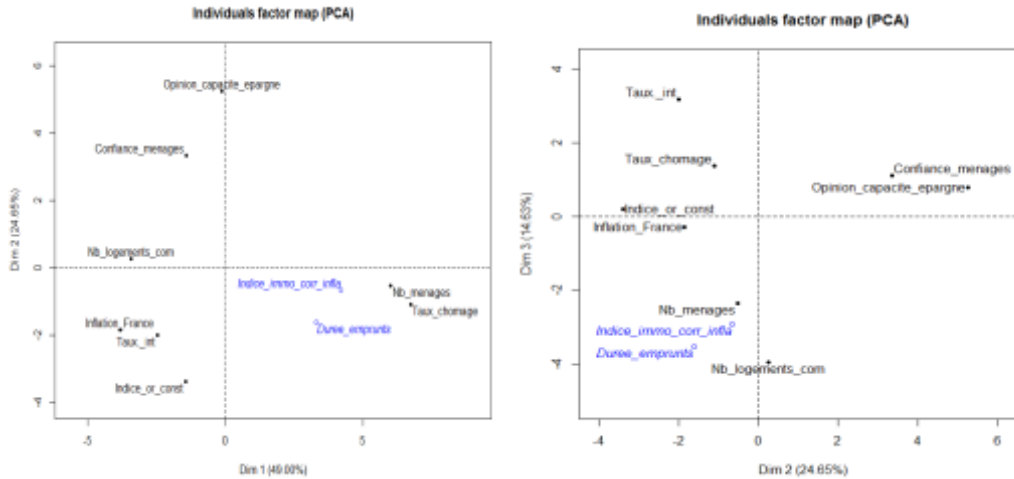
library(fda)
library(splines)
library(Matrix)
#-----
#          FDA analyse fonctionnel Données brutes F7010
###Import données
F7010 <- read.table("G:/CNAM/STA201/Tabs/F7010.csv", header=TRUE,
  sep=";", na.strings="NA", dec=".", strip.white=TRUE, row.names=1)
x=as.matrix(F7010[,-c(1,5,13,14,15)]) #transformation des données en matrice
t=1:41 #def de la dimension en abscisse
#-----
#création d'un objet "donnée fonctionnelle" : data2fd
FDF7010<-Data2fd(x,argvals=t, basisobj=bspl)
matplot(FDF7010$coef, type="l", pch=46)
plot(FDF7010[10], lty=1, col="black")
lines(t,x[,10], col="red")
#-----
#pca avec package Factominer
F7010<-FDF7010$coef
F7010<-t(F7010)
resF7010<-PCA(F7010, ncp=Inf, ind.sup= c(7,10)) # , quanti.sup=c(10))
plot.PCA(resF7010, axes=c(2,3))

```

Ce qui nous donne le modèle de la variable indice immo corr de l'inflation suivante :

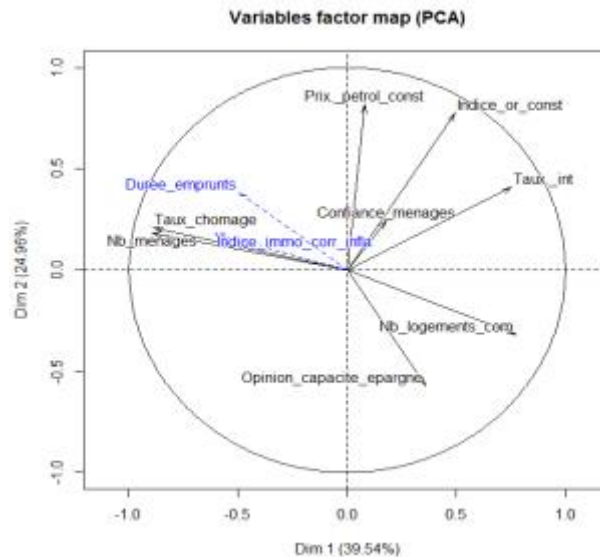


La courbe noire est le modèle à 30 splines. On voit qu'il « Fit » bien la courbe rouge des relevés de l'indice immo. Ces coefficients définis pour l'ensemble des variables vont donc nous permettre de réaliser les analyses multivariées « classiques ». On fait ici une PCA en transposant la matrice. C'est-à-dire en mettant les coefficients d'ajustement en variables et les variables économiques en individus.



L'étude PCA sur les coefficients issu de la table F7010 ne montre pas de tendance ou de proximités de la variable indice immo corrigée de l'inflation.

On peut également reprendre la transposée de la table des coefficients pour avoir les coefficients en individus. Cela permet d'avoir le variable factor's map suivante:

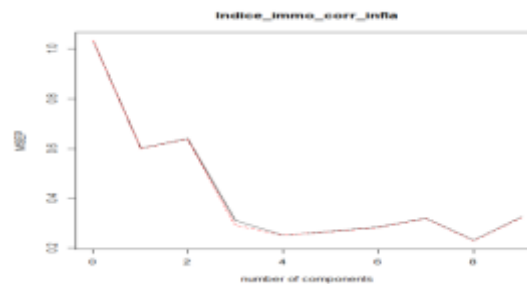


L'indice immo est corrélé à la dimension 1 qui est également corrélée au taux de chômage et au nombre de ménage qui, comme on l'a déjà vu, est représentatif du temps. On n'obtient donc pas d'information intéressante sur un lien entre l'indice immo et les autres variables économiques. Pour être le plus exhaustif possible dans la démarche, j'ai réalisé une prédiction des coefficients par PCR.

ii. Prédiction par PCR

```
#pcr avec package pls
F7010<-FDF7010$coef
F7010<-scale(F7010, center = TRUE, scale = TRUE)
F7010<-as.data.frame(F7010)
#avec LOO
respcr.loo<-pcr(Indice_immo_corr_infla ~ .,data=F7010, validation="LOO")
summary(respcr.loo)
validationplot(respcr.loo,val.type="MSEP")
```

Soit pour l'étude LOO donne le graphique de convergence de l'erreur quadratique en fonction du nombre de composantes conservées :

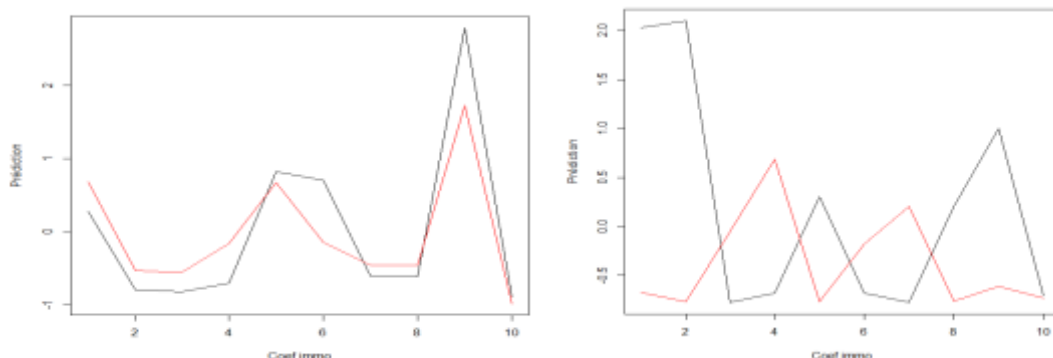


Soit en conservant 8 composantes et en réalisant une prédiction par PCR sur un échantillon de coefficients tirés au sort, le graphique d'erreurs suivants :

```
#Graphique en conservant 8 composantes
j<-8 #nb de composantes de l'ACP
x<-10 #nb de coef fonctionnels en test
learnbench<-c()
testbench<-c()

for (i in 1:nrow(F7010)){ #distribution des test et learn
  learnbench<-c(learnbench,i)}
testbench<-sample(learnbench,x)
learnbench<-setdiff(learnbench,testbench)

F7010Learn<-F7010[learnbench,]
F7010Test<-F7010[testbench,]
respcr<-pcr(Indice_immo_corr_infla ~ ., data=F7010Learn, ncomp=j)
#prediction
respcr.pred=predict(respcr,F7010Test,ncomp=j)
matplot(cbind(F7010Test[,10],respcr.pred), lty=1, type="l", xlab="Coef immo",
ylab="Prédiction")
```



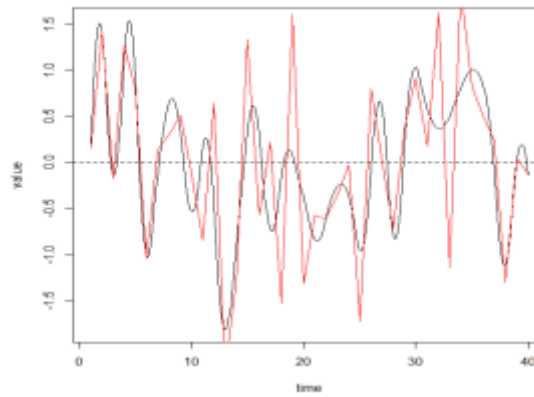
Exemple de prédiction (en noire) par rapport au coefficient calculé (en rouge) pour différents tirages aléatoires de variable « test » et « learn » (20 cas learn et 10 cas test). On remarque que la prédiction n'est pas utilisable. Le test de cross validation montre que l'erreur moyenne est de l'ordre des coefficients recherchés et ce quelque soit le nombre de composantes conservées lors de l'ACP.

b. VF7010

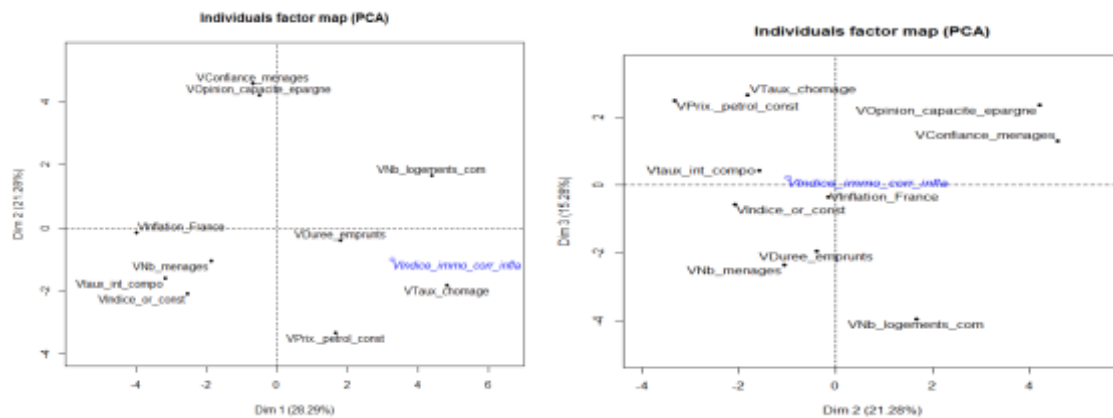
On réalise le même traitement que précédemment mais cette fois sur les données VF7010 (variation des indices de 1970 à 2010).

i. FDA (Functional data analysis)

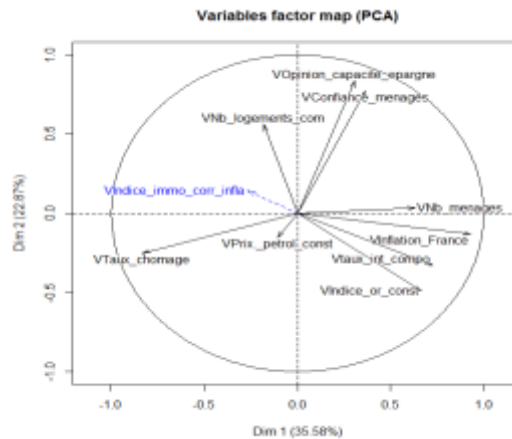
En ajustant la 30-splines à la variation de l'indice immobilier corrigé de l'inflation on obtient :



On remarque que les variations de l'indice immo sont plus chaotiques et que la 30-splines lisse beaucoup de variations qu'il est difficile d'imaginer comme continues. La continuité étant dans la FDA l'hypothèse de base.



Sur la dimension 1 (de plus forte inertie) on voit que l'indice immobilier est opposé à « taux d'interet composite » (quand les taux montent, les prix baissent comme attendu dans l'étude en annexe) et sur le plan (2,3) on retrouve bien une proximité qui était recherchée. Cela confirme le liens entre ces deux variables mais qui n'est cependant pas suffisant pour expliquer les variations de l'immobilier. Si on transpose la matrice on obtient :



Le vecteur « indice immo » est très court sur le plan (1,2). Cela indique qu'il est très mal représenté dans le plan d'inertie le plus fort. Dit autrement, l'information de la variation variables économiques (dans le cadre des variables temporelles) n'explique pas les variations de l'indice immobilier. On peut malgré tout remarquer que le vecteur du taux d'intérêt composite est diamétralement opposé à l'indice immo et c'est effectivement ce que l'on s'attendait à voir. On retrouve également le taux d'intérêt composite corrélé avec l'inflation et l'indice de l'or. A elles trois, ces composantes peuvent-être interprétées comme une dimension de cycle économique court. On voit également que la variation de confiance des ménages est à l'opposé du prix du pétrole (on pouvait également s'y attendre). Mais ce qui est suprenant ici c'est que ces deux ensembles (cycle court-immo et confiance-pétrole) sont décorrélés (à 90°) dans le plan de plus forte inertie.

5.3.Conclusion

Sur le choix de jeux de variable, on a vu que les données brutes donnaient en général de meilleurs résultats car plus régulières. Les données en variations amplifient les irrégularités mais cela a pu être lissé par une méthode fonctionnelle et qui a permis de retrouver des résultats cohérents. En ce qui concerne les liens entre les variables « immobilières » et les autres variables économiques, ils sont de plusieurs ordres :

- Un lien « non linéaire » où l'on retrouve une bonne corrélation avec des méthodes lissant les années telles que l'ACM.
- Pas de liens linéaires notables entre la variation de l'indice « immobilier » et les variations des variables économiques
- Les années sont normalement en hausse mais les baisses peuvent suivre deux modes :
 - o Choc de confiance
 - o Modification des conditions économiques globales
- Variables autorégressives
- En intégrant le caractère autorégressif (modèle fonctionnel), la variation de l'indice immobilier est anti-corrélé au taux intérêt composite (voir définition p.7) mais non corrélé à la confiance des ménages (dans un modèle fonctionnel et donc continu).

A noter que l'immobilier semble corrélé à des chocs de confiance mais non corrélé à une variation continue de cette même confiance des ménages.

L'ensemble de ces informations permettent de mieux appréhender les particularités du cours de l'immobilier sans entrer dans la granularité des villes, des quartiers, étages... A noter

qu'il n'aura pas été possible dans le cadre de cette étude de proposer une prédiction fiable. La prédiction la plus fiable ayant été celle de la méthode DisQual sur données brutes qui lisse énormément les résultats et n'est donc pas capable de prédiction hors de la zone d'apprentissage.

Dans l'optique de proposer un modèle stochastique d'évolution des prix qui puisse être intégré à un générateur de scénario économique, il sera important de retrouver toutes les caractéristiques et particularités qui ont pu être mises en avant lors de cette étude.

Liste des figures

Figure 1 : Structure du modèle d'Ahlgrim.....	7
Figure 2 : Schéma du cube des tirages du GSE	7
Figure 3 : Evolution de l'indice de prix immobilier France / Paris (Base 2000).....	10
Figure 4 : Comparaison des indices immobiliers corrigés de l'inflation	10
Figure 5 : Evolution comparée des taux d'intérêts, inflation et indices immobiliers	11
Figure 6 : Courbe des différences d'indice sur chaque période.....	12
Figure 7 : schéma de principe de l'approche stochastique.....	17
Figure 8 : $\ln Nd2Nd1$ vs. $\ln XAiXi$	24
Figure 9 : Solutions d'une équation différentielle du second ordre à coefficients constants.....	26
Figure 10 : effets du coefficient d'anticipation.....	29
Figure 11 : effets des volatilités avec $\xi = 0,75$	29
Figure 12 : exemple de choc de volatilité des vendeurs	30
Figure 13 : graphique humoristique sur les bulles spéculatives.....	30
Figure 14 : Evolution historique des propositions moyennes des acheteurs.....	33
Figure 15 : 1ere application simple du modèle	34
Figure 16 : Probabilité de vente du modèle vs. Variation des volumes de vente historique.	34
Figure 17 : Indice immobilier calibré solution n°0	36
Figure 18 : Volatilités calibrées solution n°0.....	36
Figure 19 : Indice immo vs. (Taux longs-Taux courts)	37
Figure 20 : Exemple de choc de σA en 1979.....	38
Figure 21 : Exemple des volatilités sur 40 ans à travers les 4 crises	38
Figure 22 : Volatilités calibrées solution n°1	39
Figure 23 : Indice immobilier vs. modèle solution n°1	39
Figure 24 : Probabilité de vente du modèle vs. Variation des volumes historique : solution n°1	40
Figure 25 : Volatilités calibrées solution n°2.....	40
Figure 26 : Indice immobilier vs. modèle solution n°2	40
Figure 27 : Probabilité de vente du modèle vs. Variation des volumes historique solution n°2	41
Figure 28 : Optimisation de la solution 2 avec $\rho=20\%$ et ξ variable	42
Figure 29 : Tunnel de Friggit.....	44
Figure 30 : Tunnel de l'indice immobilier historique / prix d'équilibre.....	44
Figure 31 : Baisse des Taux à 1,5% et maintient jusqu'en 2030 sans réaction des acteurs.....	45
Figure 32 : VaR de l'indice immobilier, Taux objectif des taux d'emprunt à 4%.....	47
Figure 33 : VaR du ratio Prix moyen / Prix d'équilibre : Tunnel	47
Figure 34 : VaR pour une modification de l'objectif des taux à partir de 2030	48
Figure 35 : VaR comparées des rendements annuels des 2 méthodes d'évolution des prix immobiliers (Dynamique Stochastique / Ornstein Uhlenbeck).....	49
Figure 36 : VaR comparées des rendements cumulés des 2 méthodes (Dynamique Stochastique / Ornstein Uhlenbeck)	50
Figure 37 : Comparaison des rendements annuels avec scénarios sur la durée d'emprunt	51
Figure 38 : Comparaison des rendements cumulés avec scénarios sur la durée d'emprunt	51
Figure 39 : BackTesting 1980 des rendements cumulés.....	52
Figure 40 : BackTesting 1980 des rendements annuels.....	53

Proposition d'une nouvelle modélisation de l'actif immobilier pour les générateurs de scénarios économiques

Mémoire d'actuariat C.N.A.M., Paris 2017

RESUME

Il est possible de définir un modèle d'évolution des prix sur un marché non substituable. Bien que la non substituabilité d'un marché donné –ici l'immobilier- empêche d'évaluer les prix et les rendements comparativement à tout autre marché, le modèle proposé se base sur un calcul stochastique et définit une dynamique comportementale associée. Les vendeurs et les acheteurs sont alors vus comme deux groupes distincts ayant des contraintes différentes. Le groupe des acheteurs est contraint par l'économie réelle alors que celui des vendeurs prend comme référence les ventes passées en créant ainsi une dynamique type masse-ressort amorti (inertie, force de rappel et dissipation).

L'étude démontre l'intérêt d'un tel modèle par son intégration dans un générateur de scénario économique et son application au marché immobilier d'habitation Parisien de 1970 à aujourd'hui. Le modèle proposé se montre alors plus réaliste et performant que le modèle aujourd'hui couramment utilisé.

Mots clés : modèle dynamique stochastique, marché non substituable, prix d'équilibre, tunnel, GSE, VaR, immobilier français

SUMMARY

It's possible to design a price model for a non substitutable market. Even though the non substitutability of a specific market –real estate- prevent from comparing its prices and yields to any other market, the presented model is based on a stochastic approach and on a linked behavioural dynamic. Buyers and sellers are seen as two distinct groups having different constraints. Buyers are constrained by the economy but sellers uses past prices as reference creating a dynamic loop (damped mass-spring system).

This study shows possibilities of the model with an application trough an economical scenario generator concerning the Parisian real estate market. The proposed model appears to be more realistic and more efficient than the model commonly used today.

Key words : stochastic dynamical model, non substitutable market, equilibrium price, tunnel effect, ESG, VaR, French real estate market