

Mémoire présenté le :

**pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA
et l'admission à l'Institut des Actuaires**

Par : **Simon GASQUET**

Titre **Reverse Stress Test : un outil d'aide à la décision du
pilotage financier appliqué aux risques action, taux et
immobilier**

Confidentialité : NON OUI (Durée : 1 an 2 ans)

Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

*Membre présents du jury de l'Institut
des Actuaires*

signature

Entreprise :

Nom : Caisse des Dépôts

Signature :

Directeur de mémoire en entreprise :

Nom : **Nicolas BABY**

Signature :

Invité : BIOLAY YANNOIS

Nom :

Signature :

**Autorisation de publication et de mise
en ligne sur un site de diffusion de
documents actuariels (après expiration
de l'éventuel délai de confidentialité)**

Signature du responsable entreprise



Signature du candidat



Membres présents du jury de l'ISFA



Mémoire d'Actuariat

Reverse Stress Test : un outil d'aide à la
décision du pilotage financier appliqué aux
risques action, taux et immobilier

Auteur : Simon GASQUET

Tuteur entreprise : Nicolas BABY

Tuteur académique : Nicolas LEBOISNE

Mars 2022

Table des matières

Résumé	4
Abstract	5
Remerciements	6
Introduction	7
1 Cadre du mémoire	9
I Contexte	9
1 Caisse des Dépôts et Consignations	9
2 Structure de la Caisse des Dépôts	9
3 Métiers et filiales de la Caisse des Dépôts	10
4 Sensibilités de la Caisse des Dépôts aux facteurs de risque	11
II Introduction aux Reverse Stress Tests	12
1 Définition des Stress Tests	12
2 Définition des Reverse Stress Tests	13
2 Etude des données	14
I Choix du jeu de données	14
1 Profondeur d'historique	14
2 Granularité des données	15
II Choix des facteurs de risque	15
1 Risque Action	16
2 Risque Taux	17
3 Risque Immobilier	18
3 Méthode théorique pour déterminer une relation d'ordre de probabilité	19
I Réflexions unidimensionnelles	19
1 Etude en valeur absolue	20
2 Etude en variation absolue	21
3 Etude en variation relative	22
4 Choix de la modélisation des facteurs de risque	24
5 Prise en compte des tendances	27
II Réflexions bidimensionnelles pour Excel	29
1 Structure de corrélation	29

2	Définition du coefficient de corrélation linéaire r	30
3	Mise à la même échelle	31
4	Choix de l'horizon de probabilité	31
5	Illustrations de la fonction de score	33
6	Fonction de score en deux dimensions avec un 3 ^{me} facteur de risque	34
III	Fonction de score tridimensionnelle	36
4	Reverse Stress Test via Excel	39
I	Objectif	39
1	Cadre général	39
2	Application à la Caisse des Dépôts	40
3	Application à ce mémoire	40
II	Sensibilités des facteurs de risque	41
III	Solveur Excel	42
IV	Définition des métriques de score	43
V	Solution du problème	44
VI	Sensibilités du solveur	45
VII	Contribution des facteurs de risque à l'impact cible	46
5	Alternatives aux hypothèses simplificatrices du module Excel	47
I	Alternatives aux lois marginales gaussiennes	48
1	Tests de normalité	48
2	Alternatives à la loi normale	50
2-A	Lois théoriques alternatives	50
2-B	Statistiques de comparaison	51
2-C	Proposition de procédure de sélection des marginales	54
2-D	Lois marginales retenues	55
3	Estimateurs à noyau	56
4	Bilan sur les marginales retenues	56
II	Alternatives à la copule gaussienne	57
1	Définition d'une copule	57
2	Définitions des copules envisagées	58
2-A	Copules elliptiques	58
2-B	Copules archimédiennes	59
3	Choix des copules adéquates	60
4	Bilan sur les copules retenues	63
III	Alternatives à la mesure de corrélation brute	64
1	Mesures de corrélation alternatives	64
2	Pistes de calcul du coefficient de Pearson	64
2-A	Modification de l'historique	64
2-B	Modification de la pondération	67
3	Bilan sur les coefficients de corrélation retenus	67
IV	Synthèse des modèles de Reverse Stress Test	68
6	Reverse Stress Test via R	69
I	Méthodologie par tirage	69
1	Optimisation en temps	69
2	Suppression des tendances	70
3	Prise en compte des <i>drifts</i>	70
4	Résumé des étapes du tirage du Reverse Stress Test via R	71
II	Sélection des <i>outputs</i> adéquats	72

7	Comparaison des modèles implémentés sur R	73
I	Analyse en composantes principales	74
1	Théorie	74
2	Analyse en composantes principales sur les 25 variables représentatives des modèles	76
2-A	Détermination du nombre d'axes à observer	76
2-B	Représentation des variables	77
2-C	Représentation des modèles	78
2-D	Représentation des 25 variables et des 18 modèles	79
3	Analyse en composantes principales sur le jeu des 12 variables <i>outputs</i> des modèles	81
3-A	Détermination du nombre d'axes à observer	81
3-B	Représentation des 12 variables et des 18 modèles	82
II	Partitionnement en k-moyennes	83
1	Théorie	83
2	Application sur le jeu des 25 variables représentatives des modèles	84
2-A	Détermination du paramètre k	84
2-B	Application du partitionnement en 2-moyennes	85
2-C	Application du partitionnement en 6-moyennes	86
3	Application sur le jeu des 12 variables <i>outputs</i> des modèles	87
3-A	Détermination du paramètre k	87
3-B	Application du partitionnement en 4-moyennes	88
4	Synthèse des trois partitionnements en k-moyennes	89
III	Analyse qualitative des <i>outputs</i> des modèles	90
1	Analyse qualitative de l'impact à - 3 Md €	90
2	Analyse qualitative de l'impact à - 5 Md €	91
3	Analyse qualitative de l'impact à - 7 Md €	92
4	Analyse qualitative de l'impact à - 10 Md €	92
5	Bilan des comparaisons des 18 modèles	93
	Conclusion	95
	Glossaire	98
	Bibliographie	99
	Table des figures	103
	Annexes	104
	Annexe 1 : <i>Packages</i> R utilisés	105
	Annexe 2 : Compléments sur le choix de la copule théorique adéquate	109
	Annexe 3 : Valeurs des paramètres de calibrage des RST via R	112
	Annexe 4 : Jeux de données des 18 modèles	114
	Annexe 5 : Coefficients de corrélation de Spearman et de Kendall	117
	Annexe 6 : Fonction d'autocorrélation partielle	119
	Annexe 7 : Autres lois marginales alternatives	120

Résumé

Mots-clefs : Reverse Stress Tests, test de résistance inversé, action, taux, immobilier, marginale, copule, corrélation, k-moyennes, analyse en composantes principales

Les tests de résistance inversés, ou plus communément appelés Reverse Stress Tests, sont des Stress Tests particuliers qui permettent de déterminer le scénario le plus probable provoquant un impact cible défini en amont. Ces Reverse Stress Tests figurent dans les lignes directrices de l'Autorité Bancaire Européenne (ABE) et tendent à émerger dans l'ensemble de l'environnement bancaire.

La mise en place de Reverse Stress Tests s'inscrit aussi, pour la Caisse des Dépôts, dans le cadre de la loi PACTE. En effet, la Caisse des Dépôts est depuis 2020 placée sous la supervision de l'Autorité de Contrôle Prudential et de Résolution (ACPR), adossée à la Banque de France, augmentant ainsi les exigences prudentielles auxquelles la Caisse des Dépôts doit répondre.

La Direction Financière du Groupe Caisse des Dépôts a donc décidé d'introduire des Reverse Stress Tests comme outil d'aide à la décision du pilotage du bilan. Ce mémoire a pour objectif d'étudier les différentes réflexions induites par la mise en place de Reverse Stress Tests. Le Reverse Stress Test de ce mémoire est appliqué à trois des principaux risques financiers du Groupe Caisse des Dépôts : action, taux et immobilier.

Le mémoire se découpe en quatre principales étapes : la première étape consiste à réfléchir à une méthodologie de Reverse Stress Test implémentable sur Excel, la seconde étape porte sur l'implémentation d'un premier module Excel et la formalisation des résultats, la troisième étape se concentre sur l'élaboration et l'implémentation de modèles plus robustes statistiquement sur R et la dernière étape compare ces différents modèles et estime la pertinence de maintenir ou non le module Excel.

Abstract

Key words : Reverse Stress Tests, equity risk, interest rate risk, estate risk, marginal, copula, correlation, k-means, principal component analysis

Reverse Stress Tests are specific Stress Tests calculating the most likely financial event that causes a target impact defined upstream. These Reverse Stress Tests figure in the European Banking Authority guidelines and are emerging all over the banking environment. Those guidelines of EBA aim to sum up common practices emerging in european banking institutions.

Moreover, the implementation of Reverse Stress Tests in Groupe Caisse des Dépôts also allows to upgrade its banking practices to comply with the ACPR (Autorité de Contrôle Prudentiel et de Résolution) expectations. Indeed, ACPR is integrated in the Banque de France and has become the new supervisor of Groupe Caisse des Dépôts since the PACTE law.

The Financial Direction of Groupe Caisse des Dépôts decided to implement those Reverse Stress Tests as steering decision tools for balance sheets. This paper aims to develop several considerations induced by the establishment of Reverse Stress Tests and its coding. In this paper Reverse Stress Tests are applied to equity risk, interest rate risk and estate risk.

This paper splits into four parts. The first one deals with mathematical reflexions concerning the development of Reverse Stress Tests on Excel, the second one aims to implement it. The third part focuses on statistic weakness analysis of this Excel module and the last one aims to compare several models implemented on R to evaluate the robustness of the Excel module.

Remerciements

Je tiens tout d'abord à remercier l'ISFA qui m'a offert l'opportunité de présenter un mémoire d'actuariat devant l'Institut des Actuaire ainsi que la Caisse des Dépôts et Consignations et particulièrement l'Audit du Groupe qui m'ont parfaitement accueilli professionnellement et ont permis la réalisation de ce mémoire.

Je tiens à saluer la qualité de l'accueil et à remercier l'ensemble du département 1 et particulièrement Nicolas Baby et Ximin Wan. Nicolas Baby, mon tuteur entreprise, m'a introduit dans mon poste et accompagné tout au long de l'élaboration de ce mémoire. Quant à Ximin Wan, ma superviseuse et manager, elle m'a épaulé et fait confiance dans l'ensemble de mes débuts professionnels. Merci également à Damien Nury, Marie Gabé, Fabien Chanthapanya et Marceau Pietroboni avec qui je réalisais en parallèle de ce mémoire des missions d'audit en très bonne entente et qui m'ont permis de me dégager du temps pour sa réalisation et sa rédaction.

Et un grand merci à Benoit Fache et Yannis Biolay, mes interlocuteurs de la Direction des Finances, qui ont encadré et apporté leurs avis dans la mise en place des Reverse Stress Tests ainsi que dans la rédaction de ce mémoire.

Je suis reconnaissant envers Clément Lorin, de la Direction des Finances, et Arnaud Freyder, de la Direction de l'Audit, qui ont permis aux différentes équipes de collaborer pour la mise en place de Reverse Stress Tests et d'en faire l'objet de ce mémoire. Je remercie également Nicolas Leboisne, directeur de l'ISFA et tuteur académique, avec qui j'ai pu échanger fréquemment au sujet de mon alternance et de mon mémoire.

Pour finir, je remercie ma famille qui m'a soutenu, encouragé tout au long de ce mémoire et qui m'a aidé dans sa rédaction grâce à ses relectures.

Introduction

L'Autorité Bancaire Européenne (ABE) a édité des lignes directrices ayant pour but de présenter des standards concernant la réalisation de tests de résistance, appelés Stress Tests. La Caisse des Dépôts et Consignations n'est pas explicitement soumise au respect de ces lignes directrices mais s'inscrit dans cette démarche.

Les tests de résistance inversés, appelés Reverse Stress Tests, sont une catégorie des Stress Tests et font partie des bonnes pratiques de surveillance d'une institution financière. D'après les lignes directrices de l'ABE¹, les Reverse Stress Tests sont des tests qui utilisent en entrée l'identification d'un résultat prédéfini et explorent les scénarios qui pourraient faire que ce dernier se réalise.

La loi PACTE (Plan d'Action pour la Croissance et la Transformation des Entreprises) promulguée le 22 mai 2019, place la Caisse des Dépôts et Consignations sous la supervision² de l'Autorité de Contrôle Prudentiel et de Résolution (ACPR), adossée à la Banque de France (cf. [2]). A ce titre, la Caisse des Dépôts et Consignations cherche à inscrire encore davantage ses dispositifs de gestion du risque dans les meilleurs standards de place. La Direction Financière du Groupe Caisse des Dépôts (DFIN) travaille ainsi à la mise en place de Reverse Stress Tests. Les Reverse Stress Tests auront pour but d'identifier le scénario le plus probable conduisant à un impact financier, fixé au préalable, significatif pour le Groupe. Les Reverse Stress Tests ont pour objectif de devenir des outils d'aide à la décision à destination de la gouvernance. Ils permettront par exemple d'identifier des scénarios probables sur des seuils d'alerte de solvabilité ou de liquidité du Groupe. Ces scénarios probables identifiés permettront notamment à la gouvernance d'apprécier la mise en place de mesures de couverture sur ces événements afin d'en mitiger les impacts.

L'objet de ce mémoire est donc d'initier des réflexions sur l'élaboration d'une méthodologie de Reverse Stress Tests ainsi que de produire des premiers modules permettant de réaliser des Reverse Stress Tests. Le travail sur la formalisation des résultats en sortie du module est un point d'attention constant tout au long du mémoire afin de faciliter la présentation des résultats du Reverse Stress Test à la gouvernance. Cette première réflexion sur le Reverse Stress Test est appliquée à trois facteurs de risque : un pour le risque action, un pour le risque de taux et un pour le risque immobilier. Ces trois facteurs de risque représentent les principaux risques financiers de l'Etablissement Public Caisse des Dépôts. Les sensibilités à ces trois facteurs de risque ont été déterminées par la Direction des Risques de la Caisse des Dépôts et sont prises comme des *inputs*

1. « Reverse Stress Tests means an institution stress test that starts from the identification of the pre-defined outcome [...] and then explores scenarios and circumstances that might cause this to occur. », EBA/GL/2018/04.

2. « L'Autorité de contrôle prudentiel et de résolution contrôle [...] les activités bancaires et financières de la Caisse des Dépôts et Consignations », Loi PACTE, article 112.

dans ce mémoire. Ces sensibilités sont estimées sur les pertes de valeur du portefeuille dont les valeurs présentes dans ce mémoire ont été transformées par souci de confidentialité.

La première étape de ce mémoire (cf. Partie 3) a donc été de réfléchir à l'élaboration statistique, méthodologique d'un Reverse Stress Test. Il a été choisi, dans un premier temps, de privilégier une implémentation sur Excel afin d'avoir un outil forçant une méthodologie simplifiée (moins d'outils statistiques sont présents sur Excel que sur d'autres langages informatiques) ainsi qu'une facile prise en main du module par des collaborateurs, notamment au niveau du Groupe, et dont certains peuvent être moins expérimentés à l'utilisation de modèles mathématiques ou de programmes informatiques.

La deuxième étape (cf. Partie 4) porte sur l'implémentation du modèle statistique retenu sur Excel et sur la réflexion concernant le *story telling* entourant les résultats du modèle afin de faciliter d'une part la compréhension du modèle statistique et d'autre part de transmettre puis d'expliquer le plus facilement possible, les résultats.

L'élaboration de ce module Excel a nécessité l'utilisation d'hypothèses mathématiques, possiblement assez fortes. C'est pourquoi la troisième étape porte sur le *challenge* de ses hypothèses (cf. Partie 5) et sur la réflexion concernant l'implémentation d'autres modèles statistiques de Reverse Stress Tests plus robustes sur un logiciel statistique plus puissant : R via R Studio (cf. Partie 6).

Enfin, la dernière étape du mémoire se concentre sur la comparaison des différents modèles alternatifs à celui implémenté sur Excel afin de quantifier le poids des hypothèses mathématiques (cf. Partie 7). Cette comparaison aux autres modèles hors Excel permet alors soit de confirmer la robustesse statistique du module Excel soit d'en souligner les faiblesses. Au regard de ces résultats, des correctifs au module Excel peuvent alors être appliqués ou la migration du modèle au sein d'un module hors Excel être demandée. Les éléments apportés permettront d'apporter une vision plus complète à DFIN sur les différentes possibilités de mise en place afin d'orienter le futur du modèle de Reverse Stress Test en connaissance de cause.

Cadre du mémoire

I Contexte

1 Caisse des Dépôts et Consignations

La Caisse des Dépôts et Consignations est une institution publique française créée en 1816 par Louis XVIII. L'objectif était de racheter la dette publique française des guerres napoléoniennes grâce aux consignations et aux dépôts des notaires.

Depuis la Caisse des Dépôts est toujours restée publique, ses capacités se sont développées et ses missions sont devenues plus diverses : gestion du livret A, prêts aux collectivités territoriales, gestion de régimes de retraite, investisseur institutionnel... La Caisse des Dépôts est souvent vue comme étant le bras armé de l'Etat français dans l'économie. Elle est responsable devant le Parlement français, elle est placée sous le contrôle d'une commission de surveillance comprenant des élus (notamment trois députés et deux sénateurs). Elle est dirigée par un directeur général nommé par décret du Président de la République. Le directeur général actuel est Eric Lombard depuis sa nomination le 8 décembre 2017.

2 Structure de la Caisse des Dépôts

La Caisse des Dépôts a la particularité d'avoir deux bilans, celui de la Section Générale et celui du Fonds d'Epargne. En effet, cela permet de s'assurer que l'épargne réglementée (dépôts du livret A notamment) serve bien à financer le logement social et que les dépôts juridiques servent à financer l'économie française. De plus, cela garantit l'étanchéité entre les dépôts des français et les fonds publics.

Ainsi, le Fonds d'Epargne transforme l'épargne réglementée en prêts à long terme. Ces prêts financent des programmes d'intérêt général comme la politique de logement social, la politique de la ville ou encore l'investissement dans le secteur local. C'est la Direction Financière du Fonds d'Epargne (DFFE) qui s'assure de son pilotage financier.

La Section Générale, quant à elle, gère les consignations, les dépôts juridiques, les comptes en déshérence (assurances vie, comptes bancaires). Elle est pilotée par la **Direction des Finances**

du Groupe (DFIN) qui s'assure de la solvabilité et de la liquidité du Groupe, garantit son financement, construit le plan stratégique du Groupe etc.

Les Reverse Stress Tests présentés dans cette étude sont mis en place dans le cadre de la Section Générale et du Groupe Caisse des Dépôts (*i. e.* hors Fonds d'Épargne). C'est l'équipe en charge du pilotage du bilan et de la gestion financière chez DFIN qui a pour mission de les mettre en place. Bien que personnellement rattaché à la Direction d'Audit interne DGAu, il y a eu collaboration avec cette équipe afin de mettre en place des Reverse Stress Tests.

3 Métiers et filiales de la Caisse des Dépôts

Le Groupe Caisse des Dépôts inclut diverses filiales détenues de manière majoritaire ou non par l'Établissement Public. L'objectif est ici de couvrir l'ensemble du Groupe. La Caisse des Dépôts se décompose en métiers, en divisions de la structure en fonction des rôles de chacun.

Les 6 métiers de la CDC :

- Gestion D'Actifs - GDA est le métier gérant les actifs financiers de la CDC. GDA regroupe quatre filiales :
 1. **CDC Croissance**
 2. **CDC Investissement Immobilier**
 3. **CDC Placement**
 4. **Société Forestière** - Détention et gestion des actifs forestiers de la CDC et d'autres établissements financiers

- Direction des Politiques Sociales - DPS est le métier de la CDC qui a pour but d'accompagner les parcours de vie : gestion de régimes de retraite, du compte formation, du parcours handicap...

- La Banque des Territoires - BdT est le métier de la CDC qui se charge notamment de prêter aux collectivités locales, aux organismes de logements sociaux. Son but est de réduire les fractures territoriales en France.

La Banque des Territoires est divisée en :

 - 16 Directions Régionales
 - 2 filiales :
 1. **CDC Habitat** - construction de logements
 2. la Société Centrale pour l'Équipement du Territoire (**SCET**) - accompagnement des entreprises locales par conseils et expertise

- Gestion des Participations Stratégiques - GPS accompagne les différentes filiales par une participation dans l'actionnariat des différentes entités. Cela permet à la CDC d'avoir un rôle stratégique dans différents groupes pour représenter l'institution publique.

Parmi les filiales, voici quelques exemples :

 1. **Compagnie des Alpes** - exploitation des domaines skiables
 2. **EGIS** - entreprise d'ingénierie concernant le transport, l'environnement, l'énergie, etc.
 3. **ICADE** - groupe immobilier
 4. **SFIL** - banque d'investissement local

- Banque Publique d'Investissement - **Bpifrance** filiale détenue à presque 50% par la Caisse des Dépôts, prête aux entreprises françaises pour les accompagner dans leur développement.
- Le Groupe La Poste dont la Caisse des Dépôts est devenue actionnaire majoritaire en 2020 et par la même occasion liée à ses filiales **La Banque Postale** et **CNP Assurances**.

4 Sensibilités de la Caisse des Dépôts aux facteurs de risque

Pour l'élaboration de ces Reverse Stress Tests, des travaux d'échanges ont été réalisés par DFIN, en consultant notamment la Direction des Risques (DRG) qui est en charge de la cartographie des risques du Groupe. Ces travaux ont eu pour objectif d'identifier les principaux risques de l'Établissement Public Caisse des Dépôts et d'estimer les sensibilités consolidées du Groupe Caisse des Dépôts à ces risques. Un second travail a été initié avec les différentes filiales et entités du Groupe afin d'étudier l'ajout de facteurs de risque essentiellement représentatif de filiales du Groupe. Le risque de crédit est, notamment, un risque plutôt minime au niveau de la Section Générale mais très important pour La Banque Postale, Bpifrance ou encore SFIL.

Ainsi, grâce à ces échanges réalisés, il est possible de regrouper pour les différents facteurs de risque les sensibilités des différents métiers et filiales de la Caisse des Dépôts et Consignations (côté Section Générale) :

1. **Risque Action**

Sensibilité à la baisse des actions pour :

- GDA
- Bpifrance
- GPS

2. **Risque de Taux**

Sensibilité à la hausse des taux pour :

- GDA
- BdT
- GPS

3. **Risque Immobilier**

Sensibilité à la baisse du prix de l'immobilier pour :

- GDA
- BdT
- GPS

Ces trois risques sont les risques financiers les plus importants au niveau du bilan de la Section Générale. Ils ont donc été retenus pour élaborer le Reverse Stress Test. Mais de nombreux autres risques comme le risque de crédit ou le risque de change pourraient à terme remplacer ou s'ajouter à ces trois facteurs de risque en fonction des besoins de la gouvernance ou d'une vision consolidée plus précise du Groupe.

II Introduction aux Reverse Stress Tests

1 Définition des Stress Tests

Les Stress Tests (tests de résistance) sont une série de tests qui permettent de s'assurer de la résistance d'une entreprise à différents scénarios : chocs financiers, chocs économiques, chocs pandémiques etc. Ces Stress Tests peuvent permettre, dans un cadre prudentiel, de tester la robustesse d'un modèle par le calcul d'indicateurs tels que des provisions, des marges ou des coussins de capitaux. Les chocs sont supposés représenter des événements plausibles, probables mais rares, comme décrits dans les lignes directrices de l'Autorité Bancaire Européenne (EBA : European Banking Authority) sur les Stress Tests (cf. [1]). Ces tests de résistance vont permettre de déterminer les conséquences potentielles de tels scénarios sur un établissement financier. La métrique des conséquences peut être très variée comme un impact en valeur sur son portefeuille d'actifs, sur ses engagements, sur son bilan comptable, sur sa solvabilité etc. en fonction de ce sur quoi la gouvernance souhaite être éclairée.

Les tests de résistance intègrent donc des données exogènes en amont, comme le type de scénarios à simuler ou la métrique de l'impact à mesurer. Une fois les tests réalisés, la gouvernance de l'établissement va pouvoir ainsi se concentrer sur la couverture des scénarios ayant les conséquences les plus néfastes.

Ces Stress Tests peuvent se résumer en trois types de tests :

- **Les Stress Tests Globaux**, comportant le périmètre de l'ensemble du Groupe
- **Les Stress Tests Spécifiques** se divisant eux en deux sous cas :
 1. Ceux spécifiques à un périmètre particulier : une filiale ou une portion d'un portefeuille en particulier, par exemple.
 2. Ceux spécifiques à un type de risque : risque opérationnel, risque de taux, risque de marché etc.

Ces tests permettent donc bien de déterminer des conséquences de scénarios mais encore faut-il bien définir ces scénarios. En effet, s'ils sont trop spécifiques, pas assez exhaustifs, certains chocs probables, pouvant aussi avoir de lourdes conséquences sur l'institution, pourraient être omis. S'ils sont trop vagues, les conséquences seront difficiles à estimer.

D'après les lignes directrices de l'ABE (cf. [1]), les Stress Tests doivent notamment comporter un travail conséquent sur :

1. Les facteurs de risque
 - Identification
 - Hypothèses du modèle sur les facteurs de risque
 - Sensibilité des résultats du modèle à ces hypothèses
 - Rôle des analyses à dire d'expert
2. Les performances statistiques du modèle

2 Définition des Reverse Stress Tests

Les **Reverse Stress Tests** (tests de résistance inversés) permettent d'inverser l'aval et l'amont des Stress Tests habituels. En effet, ici le but va être, à partir d'une donnée exogène cible (exemple : perte maximale de valeur supportée de X milliards d'euros du portefeuille), de déterminer les scénarios les plus probables amenant à ce genre de conséquences.



FIGURE 1 – Comparaison entre un Stress Test et un Reverse Stress Test

Comme précédemment, des Reverse Stress Tests plus ou moins spécifiques, à un périmètre ou à un type risque, sont réalisables. Les lignes directrices de l'Autorité Bancaire Européenne (ABE) (cf. [1]) préconisent d'intégrer les Reverse Stress Tests à l'ensemble des Stress Tests réalisés par l'institution financière¹.

La construction d'un Reverse Stress Test peut être représentée par les différentes étapes suivantes :

1. Détermination d'un périmètre de test
2. Définition d'une conséquence dite « objectif » (en général une conséquence non souhaitée)
3. Identification des principaux facteurs de risque pouvant provoquer cet objectif
4. Calcul de la sensibilité de cet objectif aux différents facteurs de risque
5. Identification des scénarios les plus probables de cet objectif grâce à des estimations statistiques des différents facteurs de risque

DFIN a ensuite décidé de suivre les lignes directrices établies par l'ABE. Ces Reverse Stress Tests permettront donc d'être :

- un outil de gestion des risques pour identifier les vulnérabilités de l'établissement
- un indicateur de couverture pour corriger les défaillances et aligner son appétit au risque sur les risques réels mis en évidence

1. « Institutions should also include reverse stress testing and reverse stress testing scenarios in their stress testing programmes », EBA/GL/2018/04.

Etude des données

I Choix du jeu de données

La difficulté des Reverse Stress Tests est de déterminer les scénarios les plus probables dans un environnement stressé. En effet, l'objectif est d'identifier les scénarios à couvrir afin de limiter le risque et de ne pas dépasser l'impact maximum supporté par le Groupe.

Il est donc intéressant de réfléchir à déterminer ce qu'est un environnement stressé. Idéalement, il serait déterminé en faisant fi de l'historique, afin de ne pas avoir un regard biaisé qui consisterait à répliquer simplement les scénarios passés. Cependant, cela permet tout de même d'avoir une bonne compréhension de l'objet étudié ainsi qu'une bonne idée des tendances, des événements passés pour les extrapoler. De plus, il est compliqué de réaliser un outil mathématique, financier ne se basant sur aucune donnée.

En statistiques, il est courant d'utiliser différents outils tels que les distributions probabilistes, les quantiles de distribution, les espérances, les écarts-types etc. Les espérances et les écarts-types sont plutôt utilisés pour quantifier un impact alors que les densités et les quantiles sont plutôt utilisés pour hiérarchiser des probabilités d'occurrence d'événements.

Cependant ces outils statistiques peuvent s'appliquer à tout jeu de données quantitatif. Il faut donc déterminer lesquels semblent les plus adéquats. Deux questions sont soulevées rapidement une fois le jeu de données retenu pour représenter les facteurs de risque. Quelle profondeur d'historique et quelle granularité des données choisir ?

1 Profondeur d'historique

Plus la profondeur d'un historique est grande, plus les données seront nombreuses et les tests statistiques robustes. De plus, dans le cas de calcul de cas extrêmes (scénarios stressés), il est important de représenter les données des périodes dites de crise du facteur de risque. Ici, les facteurs de risque sont financiers et ce sont notamment les deux crises de 2008 (crise des « *subprimes* ») et de 2020 (début de la pandémie du Covid) qui sont intéressantes et nécessaires à capter.

Cependant, plus un historique est grand, plus les données récentes sont diluées, c'est à dire celles qui sont les plus représentatives de l'environnement actuel, dans des données plus anciennes. C'est pourquoi, il serait inutile et même malvenu de prendre un jeu de données remontant jusqu'aux années 70 afin d'avoir des données concernant la crise pétrolière de 1973 par exemple. En effet, les données sont tellement anciennes que certaines données des facteurs de risque peuvent être inexistantes et les valeurs économiques tellement lointaines (courbes des taux, inflation, taux de change, monnaies différentes) que cela conduirait sans doute à comparer des éléments trop différents pour que cela soit pertinent. De plus, certains indices sont affectés par des effets de recomposition. C'est à dire que des sous-jacents sur lesquels sont calculés l'indice sont actualisés, remplacés, ajoutés, retirés etc. afin de mieux correspondre à la définition de l'indice en fonction des évolutions sectorielles ou économiques.

Le choix retenu a été de prendre un **historique de 20 ans** sur les données les plus récentes disponibles en fonction des facteurs de risque. Ce choix permet aussi de retenir dans l'historique le krach boursier au début des années 2000 (bulle internet). Cette profondeur est en cohérence avec celles retenues au sein du modèle prudentiel de la Section Générale. L'historique sélectionné étant, par souci de cohérence, de même profondeur pour chaque facteur de risque.

2 Granularité des données

La granularité désigne, en analyse de données, la périodicité de la valeur de chaque donnée. En effet, le choix peut être d'extraire des données annuelles, mensuelles ou encore journalières.

Plus la granularité est fine, plus le jeu de données sera conséquent et plus les tests statistiques robustes. Cependant, l'étude pourra être faussée par de nombreux « bruits » dans les données. Ce terme de « bruit » provient, historiquement, du terme physique en traitement du signal audio. Il correspond à l'ensemble des bruits parasites captés lors d'un enregistrement (et qui provient souvent de l'instrument lui-même) qui brouille le signal étudié. En finance, cela correspond, par analogie, aux variations frénétiques des cours qui sont aléatoires mais d'espérance nulle. Elles peuvent être expliquées par la fréquence discrète (discrète dans le sens opposé à continue) des achats et ventes sur le marché.

Plus la granularité est large, plus ces bruits sont masqués. Mais l'analyse peut revenir alors à étudier seulement des tendances globales et perdre de l'information sur l'évolution des cours à court terme. Or les études statistiques cherchent le plus souvent à capter cette information court-termiste, les tendances étant plus souvent étudiées qualitativement à dire d'expert.

Le calcul des marginales ou des rendements des cours requiert une granularité relativement fine. Le choix d'équilibre entre ces deux jambes a donc été de retenir une granularité hebdomadaire ou mensuelle. Ce choix pourra être affiné en fonction des facteurs de risque retenus mais devra être le même pour tous.

II Choix des facteurs de risque

Le but de cette section est de présenter les choix des différents facteurs de risque. Ces différents risques peuvent chacun être décomposés en multiples facteurs de risque afin de les représenter exhaustivement. Cependant, multiplier le nombre de facteurs de risque multiplie aussi la dimension mathématique du Reverse Stress Test complexifiant ainsi son implémentation et sa compréhension. La décision a donc été de retenir, à ce stade, un seul facteur par risque.

1 Risque Action

Le périmètre du portefeuille de la Caisse des Dépôts en actions se concentre majoritairement en France.

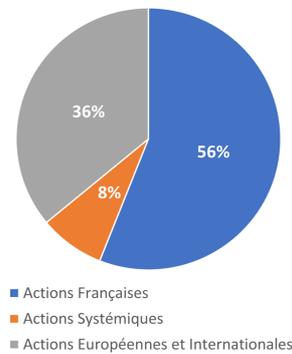


FIGURE 2 – Répartition du type d'actions cotées de la Section Générale au 30 juin 2020

Le choix s'est assez facilement porté sur l'indice boursier des actions cotées françaises : le **CAC 40**. Il est calculé en pondérant, selon leurs valeurs, les 40 actions cotées dont les échanges sont abondants et représentatifs de l'économie française. Mais l'ajout de facteurs du risque action pour représenter des actions européennes ou encore le périmètre des sociétés non cotées serait tout à fait envisageable.

Le cours du CAC 40 est extrait via Bloomberg (cf. [23]) et sera noté X^1 pour l'ensemble du mémoire.

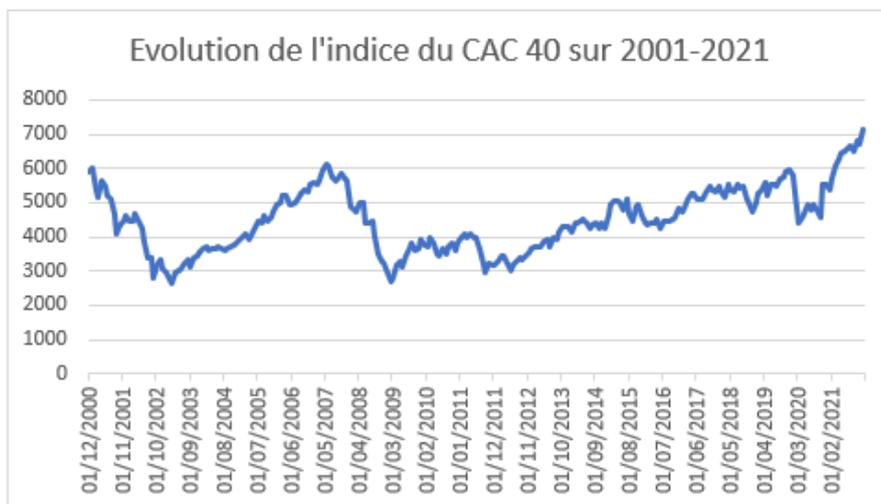


FIGURE 3 – Cours du facteur de risque action sur 21 ans

L'évolution du CAC 40 souligne nettement l'impact négatif sur le marché action français lors des crises financières du début des années 2000, de 2008 et de 2020.

2 Risque Taux

La Caisse des Dépôts est soumise au risque de taux, notamment par le biais du calcul des valeurs actuelles nettes de portefeuille où les taux représentent les taux d'actualisation utilisés. C'est pourquoi les sensibilités négatives sont plutôt associées à une hausse des taux.

De même que pour les actions, au vu du périmètre d'investissement de la Caisse des Dépôts, un taux français à retenir a été privilégié plutôt qu'un taux européen par exemple. Le choix, pour estimer les taux d'intérêt français, s'est porté sur les Obligations Assimilables du Trésor (OAT). Ce sont les taux annuels des titres de dette de l'Etat français. Ces obligations représentent ainsi un taux d'intérêt par la sûreté de remboursement de l'Etat.

Les taux sont donc utilisés pour actualiser les *cash flows* futurs qui peuvent avoir des maturités courtes (< 1 an) mais aussi très longues (> 30 ans). Dans le cadre de la prise en compte du risque global de taux, DFIN disposait de calcul sur la sensibilité de la valeur actuelle nette en fonction des différentes maturités de taux et c'est la maturité de 10 ans qui s'est trouvée être la plus contributrice aux variations de la valeur actuelle. Ce sont donc les **OAT 10 ans**, notées par la suite, par simplicité, OAT 10Y, qui ont été retenues.

Mais l'ajout de facteurs supplémentaires du risque de taux pour représenter des maturités ou un périmètre d'investissement élargi serait tout à fait envisageable.

L'historique des taux est extrait via Bloomberg (cf. [23]) et sera noté X^2 pour l'ensemble du mémoire.

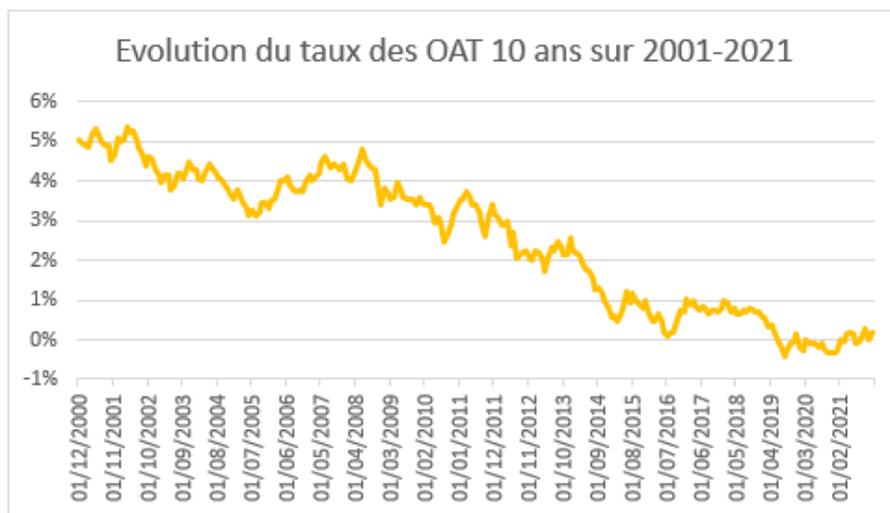


FIGURE 4 – Cours du facteur de risque taux sur 21 ans

Malgré une remontée des taux en 2021, la période se marque globalement par la baisse continue des taux, jusqu'aux taux négatifs en 2019 (pour ceux de maturité 10 ans).

3 Risque Immobilier

La Caisse des Dépôts possède aussi des actifs immobilisés français, notamment de l'immobilier. C'est pourquoi le risque immobilier, soit la variation des prix de l'immobilier, a un impact conséquent sur le portefeuille de la Caisse des Dépôts.

L'**Investment Property Databank** (noté IPD) est un baromètre qui mesure la performance immobilière en France. Il englobe des indices mesurant le rendement locatif courant et le rendement en capital. Il fournit des indices de performance de qualité aussi bien pour l'immobilier de commerces que pour l'immobilier de bureaux, l'immobilier résidentiel etc. Cependant, ce baromètre très représentatif est publié annuellement en mars. A date d'élaboration du Reverse Stress Test la valeur de l'IPD la plus récente est de 2020 (extraite en mars 2021).

DFIN mensualise cet indice chaque année grâce à l'indice de l'Institut de l'Epargne Immobilière et Foncière (IEIF) dont le relevé mensuel est possible. L'indice est composé de l'ensemble des Sociétés d'Investissements Immobiliers Cotées (SIIC).

L'ajout d'autres facteurs de risque pour représenter le reste des actifs immobilisés de la CDC, comme les forêts, serait tout aussi envisageable.

L'indice IEIF est extrait via Bloomberg (cf. [23]). L'indice IPD annuel est extrait manuellement et mensualisé par DFIN via Excel et sera noté X^3 pour l'ensemble du mémoire.

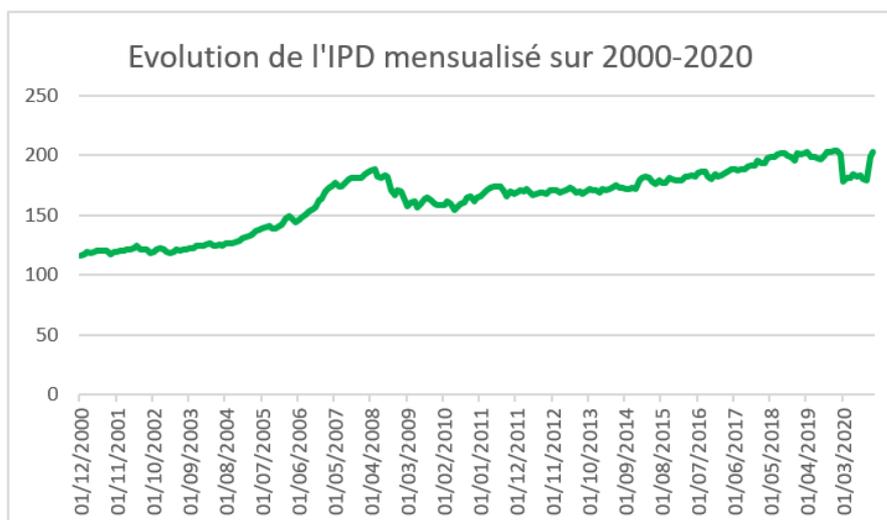


FIGURE 5 – Cours du facteur de risque immobilier sur 20 ans

Le cours affiché permet bien de visualiser l'impact à la baisse des crises financières de 2008 et de 2020 sur le marché de l'immobilier.

Ce facteur de risque mensualisé incite donc à retenir une **granularité mensuelle** des données pour les facteurs de risque taux et action. De plus, pour assurer une cohérence entre les facteurs de risque, l'échéance annuelle de l'IPD contraint à se limiter à des données jusqu'à fin 2020.

Méthode théorique pour déterminer une relation d'ordre de probabilité

Le but *in fine* du Reverse Stress Test est de hiérarchiser des évènements provoquant un évènement « objectif » selon leur probabilité d'occurrence. Il est donc nécessaire de déterminer une méthode de classification, une méthode de score, qui permette cette hiérarchisation. Le but ne sera pas nécessairement d'avoir une méthode dont la valeur a une signification mais essentiellement de pouvoir créer une relation d'ordre entre les évènements.

Pour réfléchir à cette problématique, la réduction de la dimension du problème (ici de trois par le nombre de facteurs de risque) est une bonne première approche afin de simplifier le problème.

I Réflexions unidimensionnelles

Commencer par un seul facteur permet de simplifier le problème et notamment d'exclure, pour le moment, les problématiques concernant les structures de corrélation. De plus, les questions concernant le traitement des différentes échelles de valeur entre un facteur en point et un autre en pourcentage par exemple sont aussi évitées.

Dans cette partie, le choix de l'unique facteur de risque se portera sur le cours hebdomadaire du cours CAC 40 censé représenter de façon simplifiée le risque action. Bien qu'il a été décidé de retenir *in fine* une granularité mensuelle pour le modèle, la granularité hebdomadaire se trouve plus illustrative, plus instructive grâce à son plus grand jeu de données pour cette partie.

Dans cette partie, l'objectif est donc de déterminer une méthode pouvant permettre de calculer des probabilités futures sur la valeur du CAC 40 (X_t^1 sa valeur notée en t). La méthode retenue pourra ensuite être assez facilement généralisable à d'autres facteurs de risque dans un cadre multidimensionnel.

1 Etude en valeur absolue

Le terme d'absolu, ici employé, a pour but de l'opposer au terme relatif, évoqué plus tard. Une étude statistique en absolu ne va donc étudier que les valeurs brutes prises par le facteur de risque (ici le CAC 40) indépendamment les unes des autres. Voici, ci-dessous, les différents quartiles de distribution du CAC 40.

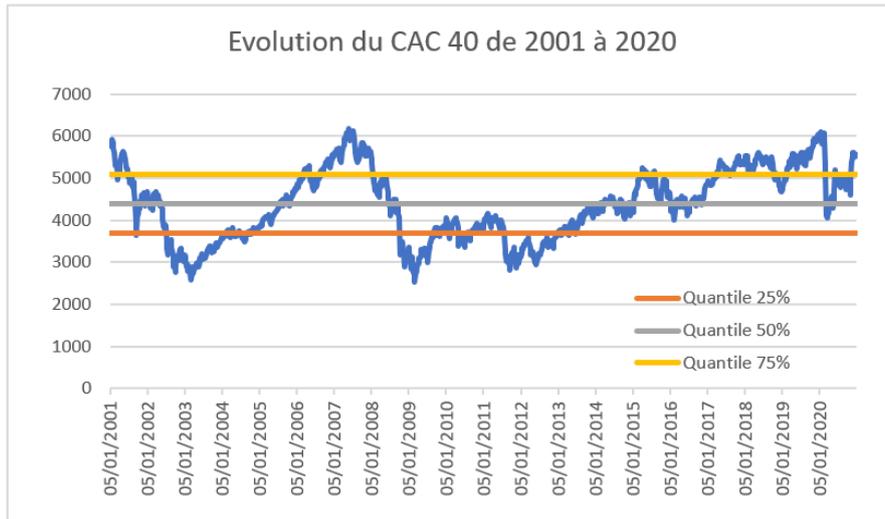


FIGURE 6 – Représentation graphique des quartiles du CAC 40

Déterminer la probabilité du CAC 40 d'atteindre un certain niveau à horizon j noté x_j à partir d'une valeur actuelle x_0 pourrait donc consister à compter le nombre historique de ce couple de valeurs survenu dans l'ensemble des paires de l'historique de même horizon j .

$$P(X_j^1 = x_j | X_0^1 = x_0) = \frac{|\{i, (X_i^1 = x_0, X_{i+j}^1 = x_j)\}|}{|\{i, (X_i^1, X_{i+j}^1)\}|}$$

Le problème est que la probabilité future ne serait alors qu'une réplique parfaite de la fréquence passée. La fonction serait alors très limitée par la taille du jeu de données (impossibilité de calculer une probabilité non nulle d'un événement non encore survenu) avec donc aucune flexibilité.

- Avantages :
 1. Très parlant
 2. Facilité d'implémentation
- Inconvénients :
 1. Reproduction parfaite de l'historique
 2. Très limité par sa profondeur

Cette méthode a peu de chances d'être la plus optimale. Cependant, avoir une vue des événements passés peut être très intéressant à titre d'information pour juger des événements en sortie du Reverse Stress Test pour *backtester* le modèle et juger si les événements prédits sont déjà survenus.

2 Etude en variation absolue

La Caisse des Dépôts est sensible à la baisse du cours du CAC 40, il est donc intéressant de regarder les principales chutes. Le terme anglais « *drawdown* » ci-dessous utilisé signifie la perte en valeur la plus importante survenue pour un horizon fixé.



FIGURE 7 – Représentation graphique des principaux *drawdowns* du CAC 40 sur 20 ans

Les deux dernières crises financières sont ainsi bien révélatrices des plus fortes baisses du CAC 40 : la crise de 2008 sur un horizon plus long que celle de la crise de 2020 qui est, elle, plus forte mais plus courte. Cette étude illustre l'importance, logique, de l'horizon choisi afin d'identifier les chocs sévères pour calibrer les quantiles adéquats.

Déterminer la probabilité du CAC 40 d'atteindre un certain niveau à horizon j noté x_j à partir d'une valeur actuelle x_0 pourrait donc revenir à calculer la probabilité de survenance d'une perte ou de gain du CAC 40 de valeur $x_j - x_0$ sur un horizon j .

$$P(X_j^1 = x_j | X_0^1 = x_0) = P(X_j^1 - X_0^1 = x_j - x_0)$$

- Avantages :
 1. Propriété d'additivité des variations
 2. Possibilité d'atteindre des événements jamais arrivés par une succession de variations absolues déjà réalisées
- Inconvénients :
 1. Hypothèse sous-jacente d'équivalence des variations absolues : le CAC a autant de probabilités de passer de 5000 points à 3500 points dans le futur que de 1500 points à 0 point
 2. Limité par la profondeur de l'historique

Une loi de probabilité peut aussi être calibrée, pour un horizon j fixé, sur l'ensemble des $\{i, X_{i+j}^1 - X_i^1\}$. Ces variations sont appelées des **marginales absolues**. Cela permettrait d'estimer une probabilité (toujours faible mais non nulle) à des variations jamais arrivées. Dans le cas de lois normales, cela permet d'obtenir des formules multidimensionnelles fermées, implémentables sur Excel.

3 Etude en variation relative

Les variations relatives sont souvent utilisées pour évoquer des variations de valeur : perdre $X\%$ plutôt que Y millions d'euros. Cela permet notamment de faire fi de l'hypothèse induite lors de l'étude en variation absolue, ainsi que de ne pas tenir compte de l'importance de la valeur de la monnaie : une perte de 1000 € en 2000 n'a pas la même importance qu'une perte de 1000 € en 2020.

Les gains ou pertes relatives hebdomadaires du CAC 40 peuvent être déterminés en calculant les $\left\{ \frac{X_{j+7}^1 - X_j^1}{X_j^1} \right\}$ sur toutes les semaines de l'historique.

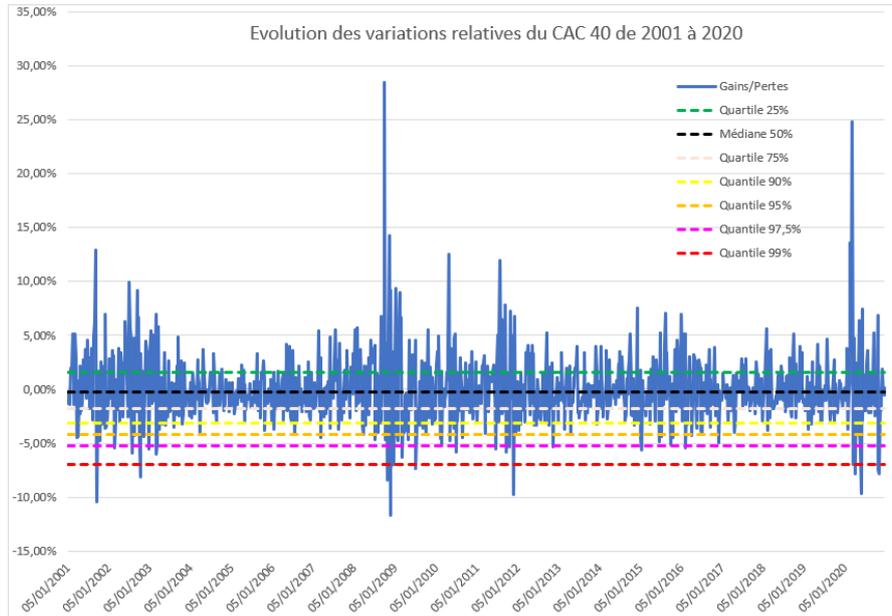


FIGURE 8 – Représentation graphique des quantiles des variations relatives du CAC 40

Les pertes importantes du CAC pendant les deux dernières crises financières sont bien mises en valeur par cette représentation du jeu de données.

Déterminer la probabilité du CAC 40 d'atteindre un certain niveau X à horizon j noté x_j à partir d'une valeur actuelle x_0 pourrait donc revenir à calculer la probabilité de survenance du gain ou de perte relative du CAC 40 à l'aide des $\frac{X_{j+7}^1 - X_j^1}{X_j^1}$ ou en passant par les log-rendements à l'aide des $\ln\left(\frac{X_{j+7}^1}{X_j^1}\right)$.

Soient P et Q deux mesures de probabilité envisageables :

$$P(X_j^1 = x_j | X_0^1 = x_0) = P\left(\frac{X_j^1 - X_0^1}{X_0^1} = \frac{x_j - x_0}{x_0}\right)$$

$$Q(X_j^1 = x_j | X_0^1 = x_0) = Q\left(\ln\left(\frac{X_j^1}{X_0^1}\right) = \ln\left(\frac{x_j}{x_0}\right)\right)$$

Les log-rendements conservent le sens de variation relative tout en disposant de la propriété mathématique d'additivité du logarithme qui peut être pratique pour se faire succéder des variations relatives.

Propriété d'additivité des logarithmes

Soit $i, j \in \mathbb{N}$ avec $j > i$ et X un jeu de données non nulles

$$\sum_{k=i}^j \ln \left(\frac{X_{k+1}}{X_k} \right) = \ln \left(\prod_{k=i}^j \frac{X_{k+1}}{X_k} \right) = \ln \left(\frac{X_{j+1}}{X_i} \right)$$

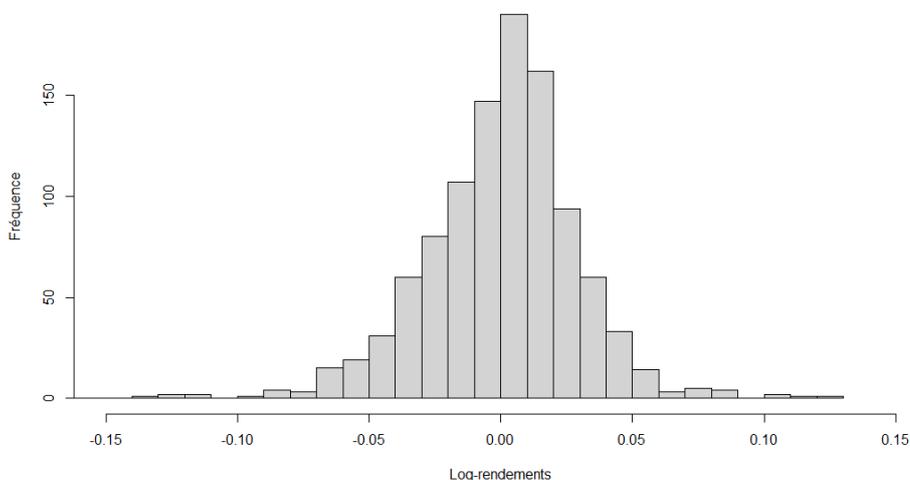


FIGURE 9 – Représentation graphique en histogramme des log-rendements hebdomadaires du CAC 40

L'histogramme des log-rendements hebdomadaires du CAC 40 souligne une forme de densité en cloche qui pourrait s'apparenter à une loi normale.

- Avantages :
 1. Propriété d'additivité des variations (dans le cas des log-rendements seulement)
 2. Plus pertinent à extrapoler historiquement
 3. Possibilité d'atteindre des évènements jamais arrivés par une succession de pertes relatives déjà réalisées

- Inconvénients :
 1. Hypothèse sous-jacente d'équivalence des pertes relatives : le CAC a autant de chances de perdre 30% de valeur à 5000 points que à 3500 points
 2. Limité par la profondeur de l'historique

Une loi de probabilité peut aussi être calibrée sur l'ensemble des $\frac{X_{j+7}^1 - X_j^1}{X_j^1}$ ou des $\ln\left(\frac{X_{j+7}^1}{X_j^1}\right)$. Ces variations sont appelées des **marginales relatives**. Cela permettra de donner une probabilité (toujours faible mais non nulle) à des variations jamais arrivées. Dans le cas de lois normales, cela permet d'obtenir des formules multidimensionnelles fermées, implémentables sur Excel.

4 Choix de la modélisation des facteurs de risque

L'étude absolue, en terme de valeurs, est une approche trop rigide pour permettre d'extrapoler les données dans des prévisions futures. Une étude sur les variations semble plus appropriée, plus flexible, plus généralisable.

Reste à sélectionner le type d'étude entre des variations relatives ou absolues. Le choix s'est porté selon le type de risque.

- Pour les cours représentant une valeur physique (cours d'actions, indice économique), l'hypothèse la moins forte est d'utiliser des variations relatives (en pourcentage). En effet, une variation en pourcentage est très représentative pour des actifs (actions, immobilier) et ne nécessite pas de prise en compte de l'inflation pour son extrapolation à aujourd'hui. Le choix se portera sur l'étude des log-rendements afin de conserver la propriété d'additivité des marginales (cf. Partie 3-I_3).
- Cependant pour des taux (en pourcentage) il est plus pertinent d'utiliser des variations absolues (en points de pourcentage), une variation relative d'une valeur représentant déjà un pourcentage étant moins compréhensible.

Pour rappel, l'IPD a orienté le choix sur l'utilisation mensuelle des cours (cf. Partie 2-II_3).

Avec X^1 , X^2 et X^3 représentant respectivement les cours du CAC 40, des OAT 10 ans et de l'IPD, les trois jeux de données sont donc transformés en 3 jeux de données de **marginales mensuelles** entre décembre 2000 ($m = 0$) et décembre 2020 ($m=240$).

$$\text{Marginales du CAC 40 : } \left\{ \ln \left(\frac{X_{m+1}^1}{X_m^1} \right) \right\}_{m \in [0, 239]}$$

$$\text{Marginales des OAT 10Y : } \left\{ (X_{m+1}^2 - X_m^2) \right\}_{m \in [0, 239]}$$

$$\text{Marginales de l'IPD : } \left\{ \ln \left(\frac{X_{m+1}^3}{X_m^3} \right) \right\}_{m \in [0, 239]}$$

Cependant, l'utilisation des valeurs des marginales (ici absolues ou relatives selon les facteurs de risque) comme un jeu de données pour calibrer une loi de probabilité implique que les données sont supposées indépendantes entre-elles justifiant ainsi la non conservation de l'ordre d'apparition des marginales successives.

Si cette hypothèse n'est pas justifiée, il pourrait être nécessaire d'altérer les données afin d'annihiler l'interdépendance des marginales ou alors de se tourner vers d'autres modélisations des facteurs de risque via des séries temporelles, par exemple. Les méthodes sur l'étude de l'autocorrélation des données sont notamment détaillées par Arthur Charpentier (cf. [22]).

Fonction d'autocorrélation

La fonction d'autocorrélation s'applique à un jeu de données et permet d'identifier, ou rejeter, une autocorrélation entre les données de ce jeu.

La fonction d'autocorrélation renvoie des coefficients de corrélation, compris donc entre -1 et 1, pour un pas de temps k fixé. Appliqué à un jeu de données de taille N X_1, \dots, X_N , le k^{eme} degré d'autocorrélation vaut :

$$r_k = \frac{\sum_{i=1}^{N-k} (X_i - \bar{X})(X_{i+k} - \bar{X})}{\sum_{i=1}^N (X_i - \bar{X})^2}$$

avec

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

Afin de rejeter l'autocorrélation du jeu de données, il est important de regarder l'ensemble des valeurs pour chaque pas de temps k jusqu'à ce que ceux-ci décroissent naturellement. A noter que r_0 est nécessairement égal à 1. Les packages statistiques informatiques permettent parfois d'intégrer un seuil de significativité au delà duquel la valeur est jugée significativement non-nulle. Ce seuil est souvent fixé, pour un seuil de significativité à 5%, à $\pm 1,96/\sqrt{N}$.

La représentation graphique (réalisée sous R cf. Annexe 1) de la fonction d'autocorrélation des log-rendements des marginales mensuelles du CAC 40 confirme leur auto-indépendance. En effet, hormis le degré 0 aucune autocorrélation ne dépasse le seuil de significativité. Le calibrage de lois statistiques sur les marginales est donc bien validé pour le CAC 40.

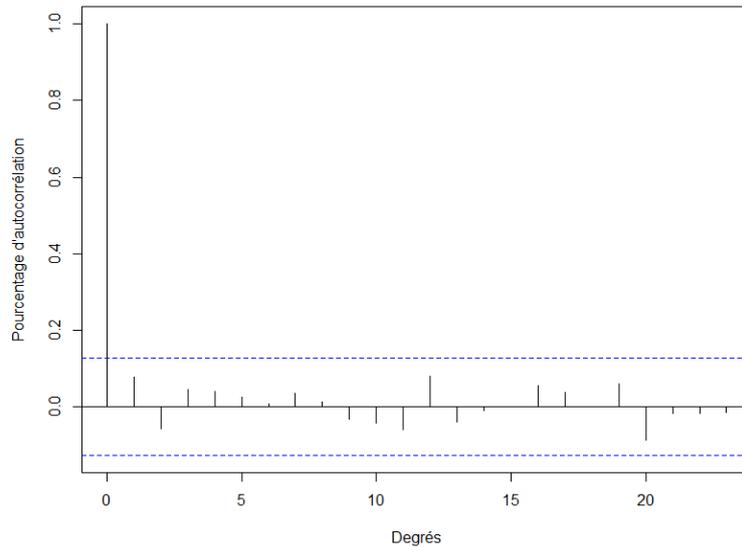


FIGURE 10 – Fonction d'autocorrélation du CAC 40

Il en va de même pour les OAT 10 ans et l'IPD dont les fonctions d'autocorrélation justifient le calibrage de lois théoriques sur le jeu de données de marginales.

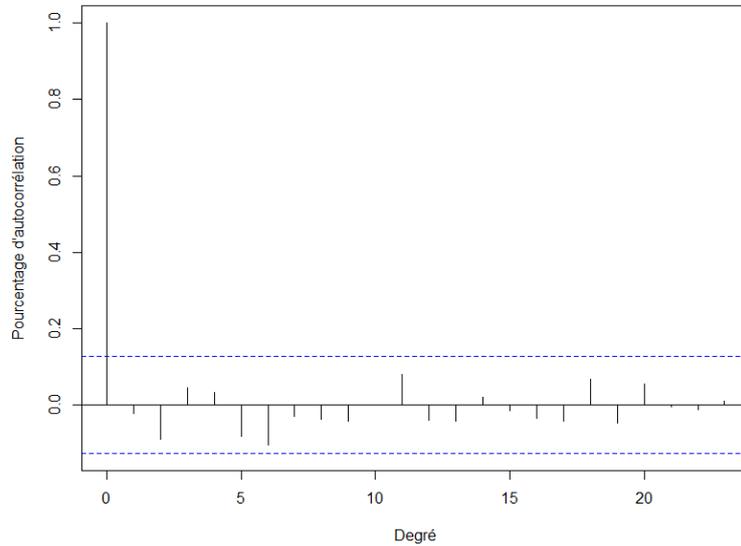


FIGURE 11 – Fonction d'autocorrélation des OAT 10Y

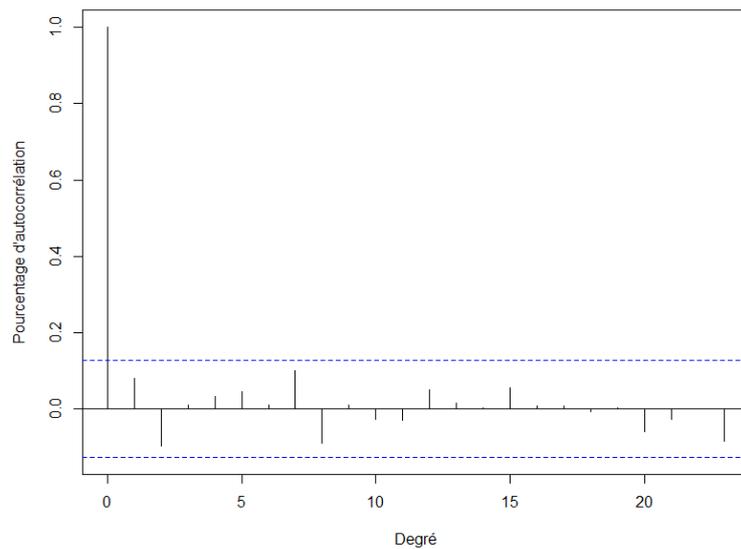


FIGURE 12 – Fonction d'autocorrélation de l'IPD

La première étape sera donc bien de calibrer des lois théoriques sur les marginales des facteurs de risque.

En cas de jeux de données autocorrélés, il est possible d'approfondir en utilisant la fonction d'autocorrélation partielle afin d'étudier plus précisément les degrés de corrélation identifiés par la fonction d'autocorrélation (cf. Annexe 6).

5 Prise en compte des tendances

Il peut être intéressant de réfléchir à neutraliser les tendances historiques (à la hausse pour le cours de l'immobilier, à la baisse pour les taux) contenues nécessairement dans le jeu de données des marginales afin de ne pas reproduire les tendances passées. Le modèle mathématique se concentrerait alors sur l'étude de la volatilité des facteurs de risque uniquement.

Pour reprendre l'exemple du CAC 40, neutraliser les tendances reviendrait à « centrer » les variations du CAC 40 autour d'une valeur constante, non évolutive dans le temps.

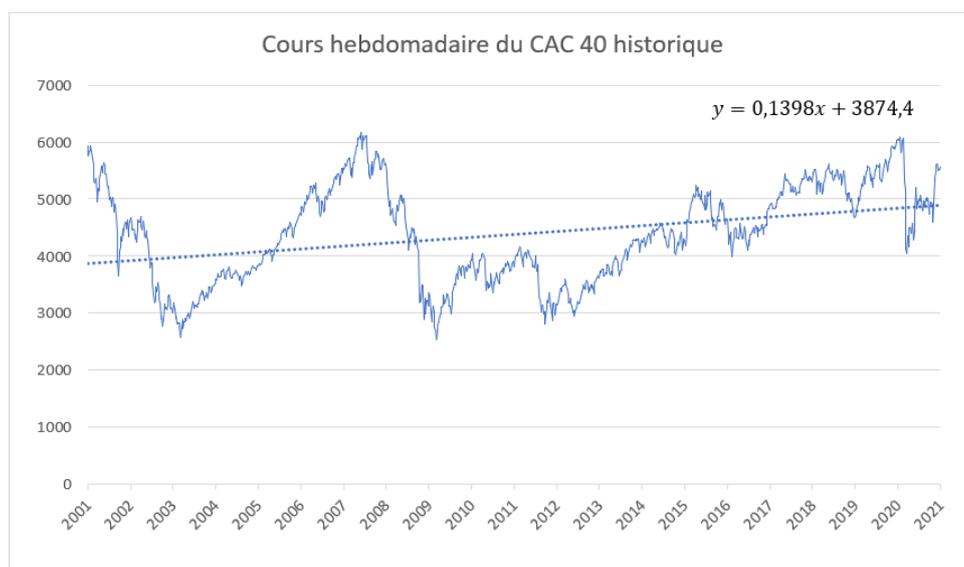


FIGURE 13 – Régression linéaire sur le cours hebdomadaire historique du CAC 40

La régression linéaire du CAC met en lumière une certaine tendance à l'augmentation du CAC 40. Elle est calculée avec un pas journalier via Excel.

- $x = 0$ correspond à la valeur de de la régression linéaire du CAC 40 au 29 décembre 2000
- $x = 7308$ correspond à la valeur de de la régression linéaire du CAC 40 au 01 janvier 2021

Ce coefficient directeur peut être rendu nul en appliquant un facteur d'actualisation aux données.

Notons X'_t les données du cours actualisé et a le coefficient directeur de la droite de régression linéaire ici égal à 0,1398.

$$X'_t = X_t + a \times t \text{ avec } t \text{ en jours}$$

Ce changement permet bien de modifier les valeurs du cours les plus lointaines d'un point de vue temporel et de modifier faiblement les valeurs récentes.

En représentant, graphiquement le cours des X' la tendance est bien neutralisée :

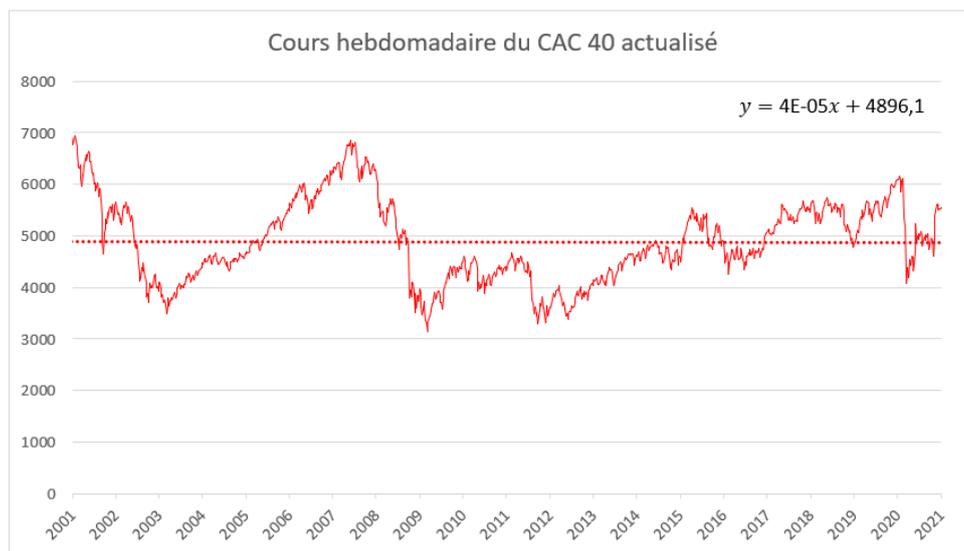


FIGURE 14 – Régression linéaire sur le cours hebdomadaire actualisé du CAC 40

Ainsi, le coefficient directeur de la droite est bien proche de 0. Les marginales seraient ainsi centrées (espérance proche de 0).

Cependant, cette prise en compte des tendances modifie la structure des marginales, notamment ses écarts-types, d'autant plus si les tendances sont importantes. C'est pourquoi il a été décidé de ne pas réaliser cette prise en compte.

Il a été choisi d'opter plutôt pour une altération de l'espérance historique (qui représente la tendance) des marginales. Cela peut se faire soit en centrant les marginales soit en forçant les paramètres des lois statistiques calibrées sur les marginales. L'objectif étant de ne pas extrapoler les tendances passées en tendances futures. Une introduction des tendances à dire d'expert (par exemple des tendances macroéconomiques) sur les tendances actuelles des différents cours pourrait être pertinente une fois le module final de Reverse Stress Test déterminé.

II Réflexions bidimensionnelles pour Excel

Il semble maintenant opportun d'étudier le passage à l'introduction d'un nouveau facteur de risque. Pour rappel, il a été décidé d'estimer les probabilités en calibrant des lois sur les marginales des facteurs de risque. La contrainte de réalisation via Excel aiguille le choix de loi. En effet, il est nécessaire que les paramètres soient facilement estimables et qu'une fonction multidimensionnelle existe, cette fonction multidimensionnelle sera appelée **fonction de score**.

C'est pourquoi **il a été retenu de calibrer des lois normales sur les marginales des facteurs de risque**. L'espérance estimée des marginales ne sera, quant à elle, pas retenue dans la modélisation pour ne pas intégrer dans la projection future l'effet des tendances passées.

Le passage d'un facteur de risque à deux induit la nécessité de traiter les corrélations des facteurs de risque entre eux. Il est aussi important de réfléchir à la pertinence de comparer des facteurs possédant des structures très différentes.

Pour cette partie à deux dimensions, le facteur de risque sur les actions françaises représenté par le CAC 40 est conservé, auquel est ajoutée l'étude des Obligations Assimilables du Trésor à maturité 10 ans (notées OAT 10Y). Le CAC 40 est un facteur de risque à points alors que l'OAT 10Y est un taux (donc en pourcentage).

1 Structure de corrélation

C'est la **fonction de densité normale bidimensionnelle qui a été retenue**. En effet, cela représente le cas particulier de la copule gaussienne où les lois de probabilité corrélées sont aussi des lois normales.

Comme expliqué précédemment, les paramètres d'espérance des lois normales μ_1 et μ_2 sont fixés nuls pour déterminer une fonction de score neutre des tendances.

Ainsi, pour rappel avec X^1 représentant le CAC 40 et X^2 représentant les OAT 10 Y ; des lois normales sont calibrées sur les marginales, en posant les hypothèses suivantes :

$$\ln\left(\frac{X_{m+1}^1}{X_m^1}\right) \sim \mathcal{N}(0, \sigma_1^2)$$
$$(X_{m+1}^2 - X_m^2) \sim \mathcal{N}(0, \sigma_2^2)$$

avec m représentant un mois.

La fonction de score (loi normale bidimensionnelle des marginales) a pour expression :

$$f(x_1, x_2) = \frac{1}{2\pi \cdot \sigma_1 \cdot \sigma_2 \cdot \sqrt{(1-r^2)}} \times \exp\left[-\frac{1}{2(1-r^2)} \times \left(\left(\frac{x_1}{\sigma_1}\right)^2 + \left(\frac{x_2}{\sigma_2}\right)^2 - 2r \cdot \frac{x_1}{\sigma_1} \cdot \frac{x_2}{\sigma_2}\right)\right]$$

avec r le coefficient de corrélation linéaire de Pearson des deux facteurs de risque.

2 Définition du coefficient de corrélation linéaire r

Etant donné que la loi de probabilité est calculée à l'aide de marginales d'un historique de 20 ans, il semble cohérent de faire, dans un premier temps, de même pour le calcul de la corrélation. La fonction normale bidimensionnelle est définie par l'utilisation du coefficient de corrélation de Pearson.

Coefficient de corrélation de Pearson

Aussi appelé coefficient de corrélation linéaire, il est le plus usuel. Il permet de mesurer si deux jeux de données ont tendance à être inférieurs ou supérieurs à leur moyenne respective plus ou moins au même moment.

Soient X et Y deux jeux de données de tailles n , $X = (x_1, \dots, x_n)$ et $Y = (y_1, \dots, y_n)$. Le coefficient de corrélation de Pearson est défini et estimé comme suit :

$$r_P(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{\sum_{i=1}^n (x_i - \mu_X)(y_i - \mu_Y)}{\sqrt{\sum_{i=1}^n (x_i - \mu_X)^2 \times \sum_{i=1}^n (y_i - \mu_Y)^2}}$$

avec μ_X et μ_Y pouvant être estimés respectivement par les estimateurs $\widehat{\mu}_X$ et $\widehat{\mu}_Y$:

$$\widehat{\mu}_X = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{et} \quad \widehat{\mu}_Y = \frac{1}{n} \sum_{i=1}^n y_i$$

- Avantages :

1. Très utilisé
2. Capte l'impact des valeurs extrêmes
3. Stable par transformation affine
4. Coefficient utilisé, entre autres, pour les structures de corrélations des copules gaussiennes et de Student

- Inconvénients :

1. Moins pertinent pour des corrélations non affines
2. Capte l'impact des valeurs aberrantes
3. Instable par transformation monotone non affine

D'autres coefficients comme le Rho de Spearman ou le Tau de Kendall existent (cf. Annexe 5) et pourraient se trouver pertinents par la suite dans le cas où il serait choisi de retenir une fonction de score autre que la loi normale multidimensionnelle.

3 Mise à la même échelle

Deux autres problématiques sont induites par le fait d'avoir deux facteurs de risque de structure différente (log-rendement de points et différence de taux) et de taille différente (écart-type des marginales des OAT environ dix fois supérieures à celui des marginales du CAC).

Le premier point est de vérifier que le coefficient de corrélation prend bien en compte cette structure de tailles différentes. La formule du coefficient de corrélation de Pearson, par la division par les écarts-types, revient bien à réduire et rendre comparables les données. Ce point est donc bien pris en compte.

Le deuxième point est de vérifier que la structure de corrélation le prend bien aussi en compte. De même, la formule de la loi normale multidimensionnelle utilisée divise bien chaque observation x_1 ou x_2 par son écart-type. Sans la prise en compte des tendances (hypothèse que $\mu_1 = 0$ et $\mu_2 = 0$), cela revient à normaliser et réduire les jeux de données de marginales, puis, à utiliser une loi normale multidimensionnelle sur deux lois normales $\mathcal{N}(0,1)$ corrélées. Ce point est donc aussi bien pris en compte.

4 Choix de l'horizon de probabilité

Les marginales sont calculées sur des données mensuelles. La densité de probabilité permet donc de déterminer des probabilités d'évènements à horizon 1 mois. Cependant, au vu de la faible volatilité des cours, les évènements probables à horizon 1 mois risquent de ne pas être très différents de ceux actuels. De plus, l'objectif des Reverse Stress Tests est d'analyser le risque à un horizon plus lointain.

Le choix de l'horizon n'est pas déterminant étant donné que la fonction à construire a pour caractéristique nécessaire, uniquement, d'être ordinale. Mais se fixer un horizon permet tout de même de mieux quantifier les probabilités des évènements. De plus, un horizon permet de capter l'information sur le cumul des marginales.

Il a été choisi, au vu des différentes temporalités de bilans et de choix de gouvernance, de prendre un horizon d'un an soit **douze mois**. Ce choix pourra être revu si l'utilité s'en faisait ressentir.

Cette extrapolation d'une projection 1 mois à 1 an est facilement réalisable grâce à la propriété d'additivité des marginales :

$$\ln \left(\frac{X_{m+12}^1}{X_m^1} \right) = \sum_{i=m}^{m+11} \ln \left(\frac{X_{i+1}^1}{X_i^1} \right)$$
$$(X_{m+12}^2 - X_m^2) = \sum_{i=m}^{m+11} (X_{i+1}^2 - X_i^2)$$

De plus, par propriété d'additivité des lois normales indépendantes (qui a été vérifiée par la fonction d'autocorrélation des marginales de chaque facteur de risque) la fonction de score peut avoir une écriture simplifiée.

Propriété d'additivité de lois normales indépendantes

$$\mathcal{N}(\mu_1, \sigma_1^2) + \mathcal{N}(\mu_2, \sigma_2^2) = \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Cette formule se simplifie avec l'égalité des μ_i et des σ_i du fait que les lois normales additionnées soient des lois normales calibrées sur un même historique de marginales d'un même facteur de risque et en fixant les espérances égales à 0 pour neutraliser l'effet des tendances. Ainsi, $\mu_1 = \mu_2 = 0$ et $\sigma_1 = \sigma_2$. Puis la formule se généralise ensuite facilement avec une somme de douze termes.

L'extrapolation annuelle de la calibration mensuelle des marginales peut être faite comme suit :

$$\ln\left(\frac{X_{m+12}^1}{X_m^1}\right) \sim \mathcal{N}(\mu'_1, \sigma_1'^2)$$

$$(X_{m+1}^2 - X_m^2) \sim \mathcal{N}(\mu'_2, \sigma_2'^2)$$

avec

$$\mu'_1 = 12\mu_1 \text{ et } \mu'_2 = 12\mu_2$$

$$\sigma'_1 = \sqrt{12\sigma_1^2} \text{ et } \sigma'_2 = \sqrt{12\sigma_2^2}$$

A laquelle peut être appliquée la fonction de score f comme vu précédemment sans utiliser μ'_1 et μ'_2 .

5 Illustrations de la fonction de score

Les valeurs prises par la fonction de score sont représentables graphiquement en discrétisant les valeurs envisageables par le CAC 40 et les OAT 10 ans. Comme vu précédemment, cela nécessite au préalable les estimations de plusieurs paramètres : valeurs récentes des facteurs de risque, écarts-types des marginales et coefficient de corrélation de Pearson.

	CAC 40	OAT 10 ans
Valeurs récentes		
31/12/2020	5551,41	-0,338
Ecart-type des marginales		
Mensuel	0,0230	0,1933
Annuel	0,0795	0,6697
Corrélation de Pearson des marginales		
21,32%		

FIGURE 15 – Tableau des paramètres du CAC 40 et des OAT 10Y

Il est ensuite possible, par des pas de discrétisation des valeurs futures envisageables respectives de chaque facteur de risque, d'observer les valeurs prises par la fonction de score sur cette discrétisation et de visualiser les événements les plus probables.

		OAT 10Y	-2,5	-2	-1,5	-1	-0,5	0	0,5	1	1,5	2
		N+1 - N	-2,16	-1,66	-1,16	-0,66	-0,16	0,34	0,84	1,34	1,84	2,34
Valeurs CAC 40	Log(N+1/N)	Valeurs Fonction Score										
2400	-0,36	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
2800	-0,30	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
3200	-0,24	0,000	0,000	0,001	0,002	0,002	0,001	0,000	0,000	0,000	0,000	0,000
3600	-0,19	0,000	0,001	0,004	0,009	0,010	0,006	0,002	0,000	0,000	0,000	0,000
4000	-0,14	0,000	0,003	0,013	0,027	0,033	0,022	0,008	0,002	0,000	0,000	0,000
4400	-0,10	0,001	0,006	0,024	0,056	0,073	0,053	0,022	0,005	0,001	0,000	0,000
4800	-0,06	0,001	0,007	0,033	0,084	0,119	0,094	0,041	0,010	0,001	0,000	0,000
5200	-0,03	0,001	0,007	0,036	0,099	0,151	0,128	0,061	0,016	0,002	0,000	0,000
5600	0,00	0,001	0,006	0,033	0,096	0,157	0,143	0,073	0,021	0,003	0,000	0,000
6000	0,03	0,000	0,005	0,026	0,081	0,141	0,136	0,073	0,022	0,004	0,000	0,000
6400	0,06	0,000	0,003	0,018	0,060	0,110	0,113	0,065	0,021	0,004	0,000	0,000
6800	0,09	0,000	0,002	0,012	0,040	0,078	0,085	0,051	0,017	0,003	0,000	0,000
7200	0,11	0,000	0,001	0,007	0,025	0,051	0,058	0,037	0,013	0,003	0,000	0,000
7600	0,14	0,000	0,001	0,004	0,014	0,031	0,037	0,025	0,009	0,002	0,000	0,000
8000	0,16	0,000	0,000	0,002	0,008	0,018	0,022	0,016	0,006	0,001	0,000	0,000
8400	0,18	0,000	0,000	0,001	0,004	0,010	0,013	0,009	0,004	0,001	0,000	0,000
8800	0,20	0,000	0,000	0,000	0,002	0,005	0,007	0,005	0,002	0,001	0,000	0,000
9200	0,22	0,000	0,000	0,000	0,001	0,003	0,004	0,003	0,001	0,000	0,000	0,000
9600	0,24	0,000	0,000	0,000	0,000	0,001	0,002	0,002	0,001	0,000	0,000	0,000
10000	0,26	0,000	0,000	0,000	0,000	0,001	0,001	0,001	0,000	0,000	0,000	0,000

Le plus probable ↑

Le moins probable ↓

FIGURE 16 – Discrétisation de la fonction de score en 2 dimensions entre le CAC 40 et les OAT 10Y

Ce tableau permet de se représenter visuellement la structure de probabilité et de corrélation entre les deux facteurs de risque.

Les colonnes et les lignes affichées permettent de montrer les différentes étapes de construction de la fonction de score :

1. Les valeurs envisageables du CAC 40 (pas de 400 points) et des OAT 10 ans (pas de 0,5%) sont discrétisées.
2. Les valeurs des marginales correspondantes de ces valeurs envisageables sont calculées par rapport aux dernières valeurs récentes.
3. La fonction de score f est calculée en utilisant le coefficient de corrélation entre les marginales, pour chaque élément de la grille.

La fonction de *Mise en Forme Conditionnelle* est ensuite utilisée via Excel afin de différencier visuellement les événements probables (en vert) des événements peu probables (en rouge). Les valeurs en bleu correspondent à l'évènement probable qui est le plus proche de l'évènement récent puisque l'hypothèse $\mu_1 = \mu_2 = 0$ a été forcée.

6 Fonction de score en deux dimensions avec un 3^{me} facteur de risque

Dans la perspective de l'élaboration complète d'un Reverse Stress Test à trois dimensions, la fonction de score pour les deux autres combinaisons des facteurs de risque peut être illustrée comme précédemment en deux dimensions.

Le 3^{me} facteur à introduire est donc, comme défini précédemment, l'Investment Property Databank (IPD) représentant le risque immobilier. Pour rappel, ce facteur est un indice, l'estimation retenue est donc semblable à celle du CAC 40. Avec X^3 son cours, ses marginales sont représentées par des log-rendements sur lesquels une loi normale centrée est calibrée.

Complétons dans un premier temps la mise à jour des données nécessaires pour calculer la fonction de score :

	CAC 40	OAT 10 ans	IPD
Valeurs récentes			
31/12/2020	5551,41	-0,338	202,53
Ecart-type des marginales			
Mensuel	0,0230	0,1933	0,0075
Annuel	0,0795	0,6697	0,0259
Matrice de corrélation			
Pearson	CAC 40	OAT 10 ans	IPD
CAC 40	1	21,32%	64,54%
OAT 10 ans	21,32%	1	-2,26%
IPD	64,54%	-2,26%	1

FIGURE 17 – Tableau des paramètres du CAC 40, des OAT 10Y et de l'IPD

De la même manière, il est possible d'illustrer par un tableau Excel les valeurs prises par les fonctions de score bidimensionnelles basées sur CAC & IPD puis OAT & IPD à l'aide de grilles discrétisées.

		IPD										
		160	170	180	190	200	210	220	230	240	250	260
Log(N+1/N)		-0,10	-0,08	-0,05	-0,03	-0,01	0,02	0,04	0,06	0,07	0,09	0,11
CAC 40	Log(N+1/N)	Valeurs Fonction Score										
2400	-0,36	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
2800	-0,30	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
3200	-0,24	0,000	0,001	0,002	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000
3600	-0,19	0,000	0,002	0,011	0,011	0,003	0,000	0,000	0,000	0,000	0,000	0,000
4000	-0,14	0,000	0,003	0,023	0,042	0,020	0,003	0,000	0,000	0,000	0,000	0,000
4400	-0,10	0,000	0,002	0,029	0,088	0,068	0,016	0,001	0,000	0,000	0,000	0,000
4800	-0,06	0,000	0,001	0,024	0,116	0,141	0,052	0,007	0,000	0,000	0,000	0,000
5200	-0,03	0,000	0,000	0,014	0,107	0,195	0,107	0,021	0,002	0,000	0,000	0,000
5600	0,00	0,000	0,000	0,007	0,073	0,198	0,156	0,043	0,005	0,000	0,000	0,000
6000	0,03	0,000	0,000	0,002	0,040	0,156	0,173	0,066	0,010	0,001	0,000	0,000
6400	0,06	0,000	0,000	0,001	0,018	0,100	0,153	0,079	0,016	0,001	0,000	0,000
6800	0,09	0,000	0,000	0,000	0,007	0,054	0,112	0,077	0,020	0,002	0,000	0,000
7200	0,11	0,000	0,000	0,000	0,003	0,026	0,070	0,063	0,021	0,003	0,000	0,000
7600	0,14	0,000	0,000	0,000	0,001	0,011	0,039	0,045	0,019	0,004	0,000	0,000
8000	0,16	0,000	0,000	0,000	0,000	0,004	0,019	0,028	0,015	0,004	0,000	0,000
8400	0,18	0,000	0,000	0,000	0,000	0,001	0,009	0,016	0,011	0,003	0,000	0,000
8800	0,20	0,000	0,000	0,000	0,000	0,000	0,004	0,008	0,007	0,002	0,000	0,000
9200	0,22	0,000	0,000	0,000	0,000	0,000	0,001	0,004	0,004	0,002	0,000	0,000
9600	0,24	0,000	0,000	0,000	0,000	0,000	0,001	0,002	0,002	0,001	0,000	0,000
10000	0,26	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,001	0,001	0,000	0,000

FIGURE 18 – Discrétisation de la fonction de score en 2 dimensions entre le CAC 40 et l'IPD

		IPD												
		150	160	170	180	190	200	210	220	230	240	250	260	270
LOG(N+1/N)		-0,13	-0,10	-0,08	-0,05	-0,03	-0,01	0,02	0,04	0,06	0,07	0,09	0,11	0,12
OAT 10Y	N+1 - N	Valeurs Fonction Score												
-4	-3,66	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
-3,5	-3,16	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
-3	-2,66	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
-2,5	-2,16	0,000	0,000	0,000	0,000	0,000	0,001	0,001	0,000	0,000	0,000	0,000	0,000	0,000
-2	-1,66	0,000	0,000	0,000	0,001	0,004	0,007	0,006	0,003	0,001	0,000	0,000	0,000	0,000
-1,5	-1,16	0,000	0,000	0,000	0,005	0,019	0,034	0,030	0,014	0,004	0,001	0,000	0,000	0,000
-1	-0,66	0,000	0,000	0,001	0,013	0,054	0,095	0,082	0,038	0,010	0,002	0,000	0,000	0,000
-0,5	-0,16	0,000	0,000	0,002	0,022	0,087	0,151	0,129	0,059	0,016	0,003	0,000	0,000	0,000
0	0,34	0,000	0,000	0,002	0,020	0,080	0,137	0,116	0,053	0,014	0,002	0,000	0,000	0,000
0,5	0,84	0,000	0,000	0,001	0,011	0,042	0,072	0,059	0,027	0,007	0,001	0,000	0,000	0,000
1	1,34	0,000	0,000	0,000	0,003	0,013	0,021	0,017	0,008	0,002	0,000	0,000	0,000	0,000
1,5	1,84	0,000	0,000	0,000	0,001	0,002	0,004	0,003	0,001	0,000	0,000	0,000	0,000	0,000
2	2,34	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
2,5	2,84	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
3	3,34	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000

FIGURE 19 – Discrétisation de la fonction de score en 2 dimensions entre l'OAT 10Y et l'IPD

L'impact de la corrélation (environ 65%) entre les marginales du CAC 40 et de l'IPD, dans la détermination des valeurs de la fonction de score, est bien souligné par la mise en forme conditionnelle d'Excel.

III Fonction de score tridimensionnelle

Maintenant que les réflexions sur l'estimation de fonction de score à deux dimensions ont été menées, un premier Reverse Stress Test complet à trois dimensions peut être réalisé. Pour rappel, la structure de corrélation retenue est une copule gaussienne avec le coefficient de corrélation linéaire de Pearson. Ceci n'a pas nécessité de travail préalable sur les données pour les mettre sur une échelle commune. La fonction de score renvoie les valeurs de probabilité à horizon 1 an.

Rappelons que les 3 facteurs de risque CAC 40, OAT 10Y et IPD (resp. X^1 , X^2 et X^3) sont représentés par leurs marginales mensuelles sur lesquelles sont calibrées des lois normales.

Les hypothèses suivantes sont posées :

$$\ln \left(\frac{X_{m+1}^1}{X_m^1} \right) \sim \mathcal{N}(\mu_1, \sigma_1^2)$$

$$(X_{m+1}^2 - X_m^2) \sim \mathcal{N}(\mu_2, \sigma_2^2)$$

$$\ln \left(\frac{X_{m+1}^3}{X_m^3} \right) \sim \mathcal{N}(\mu_3, \sigma_3^2)$$

où les espérances et les écarts-types sont estimés statistiquement sur n (ici $n = 240$) marginales mensuelles d'historique.

Ainsi en les approximant par leurs estimateurs usuels :

$$\widehat{\mu}_1 = \frac{1}{n} \sum_{i=0}^{n-1} \ln \left(\frac{X_{i+1}^1}{X_i^1} \right)$$

$$\widehat{\mu}_2 = \frac{1}{n} \sum_{i=0}^{n-1} (X_{i+1}^2 - X_i^2)$$

$$\widehat{\mu}_3 = \frac{1}{n} \sum_{i=0}^{n-1} \ln \left(\frac{X_{i+1}^3}{X_i^3} \right)$$

et

$$\widehat{\sigma}_1 = \sqrt{\frac{1}{n-1} \sum_{i=0}^{n-1} \left(\ln \left(\frac{X_{i+1}^1}{X_i^1} \right) - \widehat{\mu}_1 \right)^2}$$

$$\widehat{\sigma}_2 = \sqrt{\frac{1}{n-1} \sum_{i=0}^{n-1} (X_{i+1}^2 - X_i^2 - \widehat{\mu}_2)^2}$$

$$\widehat{\sigma}_3 = \sqrt{\frac{1}{n-1} \sum_{i=0}^{n-1} \left(\ln \left(\frac{X_{i+1}^3}{X_i^3} \right) - \widehat{\mu}_3 \right)^2}$$

Il a été décidé de ne pas prendre en compte les tendances passées. Ainsi, $\widehat{\mu}_1 = \widehat{\mu}_2 = \widehat{\mu}_3 = 0$.

La fonction de score est estimée sur un horizon de 1 an, les données en paramètre de la fonction sont donc estimées sur une extrapolation annuelle des marginales mensuelles :

$$\ln\left(\frac{X_{m+12}^1}{X_m^1}\right) \sim \mathcal{N}(0, \sigma_1'^2)$$

$$(X_{m+12}^2 - X_m^2) \sim \mathcal{N}(0, \sigma_2'^2)$$

$$\ln\left(\frac{X_{m+12}^3}{X_m^3}\right) \sim \mathcal{N}(0, \sigma_3'^2)$$

avec leurs estimateurs respectifs :

$$\widehat{\sigma}_1' = \sqrt{12\widehat{\sigma}_1^2}, \widehat{\sigma}_2' = \sqrt{12\widehat{\sigma}_2^2} \text{ et } \widehat{\sigma}_3' = \sqrt{12\widehat{\sigma}_3^2}$$

Ces trois hypothèses sont satisfaites dans la mesure où les marginales suivent bien des lois normales et sont supposées respecter le critère de martingalité.

La **fonction de score**, de structure de corrélation copule gaussienne, est donc logiquement à trois dimensions :

$$f(x_1, x_2, x_3) = \frac{1}{(2\pi)^{3/2} \times \sqrt{|\Sigma|}} \times \exp\left(-\frac{1}{2} \times X^T \Sigma^{-1} X\right)$$

avec

$$X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

et

$$x_1 = \ln\left(\frac{X_{252}^1}{X_{240}^1}\right), x_2 = (X_{252}^2 - X_{240}^2), x_3 = \ln\left(\frac{X_{252}^3}{X_{240}^3}\right)$$

où X_{252}^i représente n'importe quelle valeur possible du facteur de risque i pour estimer sa probabilité de survenir dans 1 an et X_{240}^i la dernière valeur récente du facteur de risque i.

Ainsi qu'avec la matrice de covariance suivante :

$$\Sigma = \begin{pmatrix} \sigma_1'^2 & r_{1,2} \cdot \sigma_1' \cdot \sigma_2' & r_{1,3} \cdot \sigma_1' \cdot \sigma_3' \\ r_{1,2} \cdot \sigma_1' \cdot \sigma_2' & \sigma_2'^2 & r_{2,3} \cdot \sigma_2' \cdot \sigma_3' \\ r_{1,3} \cdot \sigma_1' \cdot \sigma_3' & r_{2,3} \cdot \sigma_2' \cdot \sigma_3' & \sigma_3'^2 \end{pmatrix}$$

et les $r_{i,j}$ représentant les corrélations linéaires de Pearson calculées sur les marginales mensuelles des facteurs de risque i et j.

Soient $\forall i, j \in \llbracket 1, 3 \rrbracket$ $\widehat{r}_{i,j}$ les estimateurs des $r_{i,j}$:

$$\widehat{r}_{1,2} = \frac{\sum_{i=0}^{n-1} \ln\left(\frac{X_{i+1}^1}{X_i^1}\right) \times (X_{i+1}^2 - X_i^2)}{\sqrt{\sum_{i=0}^{n-1} \ln\left(\frac{X_{i+1}^1}{X_i^1}\right)^2 \times \sum_{i=0}^{n-1} (X_{i+1}^2 - X_i^2)^2}}$$

$$\widehat{r}_{1,3} = \frac{\sum_{i=0}^{n-1} \ln\left(\frac{X_{i+1}^1}{X_i^1}\right) \times \ln\left(\frac{X_{i+1}^3}{X_i^3}\right)}{\sqrt{\sum_{i=0}^{n-1} \ln\left(\frac{X_{i+1}^1}{X_i^1}\right)^2 \times \sum_{i=0}^{n-1} \ln\left(\frac{X_{i+1}^3}{X_i^3}\right)^2}}$$

$$\widehat{r}_{2,3} = \frac{\sum_{i=0}^{n-1} (X_{i+1}^2 - X_i^2) \times \ln\left(\frac{X_{i+1}^3}{X_i^3}\right)}{\sqrt{\sum_{i=0}^{n-1} (X_{i+1}^2 - X_i^2)^2 \times \sum_{i=0}^{n-1} \ln\left(\frac{X_{i+1}^3}{X_i^3}\right)^2}}$$

Rappel ci-dessous des paramètres en entrée de la fonction de score :

	CAC 40	OAT 10 ans	IPD
Valeurs récentes			
31/12/2020	$X_{240}^1 = 5551,41$	$X_{240}^2 = -0,338$	$X_{240}^3 = 202,53$
Ecart-type des marginales			
Mensuel	$\widehat{\sigma}_{1,2} = 0,0230$	$\widehat{\sigma}_{1,3} = 0,1933$	$\widehat{\sigma}_{2,3} = 0,0075$
Annuel	$\widehat{\sigma}'_{1,2} = 0,0795$	$\widehat{\sigma}'_{1,3} = 0,6697$	$\widehat{\sigma}'_{2,3} = 0,0259$
Coefficient de Corrélation Linéaire (Pearson)			
Facteur n°1	CAC 40	CAC 40	OAT 10Y
Facteur n°2	OAT 10Y	IPD	IPD
Valeur	$\widehat{r}_{1,2} = 21,32\%$	$\widehat{r}_{1,3} = 64,54\%$	$\widehat{r}_{2,3} = -2,26\%$

FIGURE 20 – Tableau des paramètres des facteurs de risque en entrée du Reverse Stress Test

Maintenant que l'ensemble des paramètres de la fonction de score a été défini, l'étape suivante consiste à la lier aux sensibilités de la Caisse des Dépôts sur chaque facteur de risque afin d'estimer l'évènement le plus probable d'une perte fixée.

Reverse Stress Test via Excel

La fonction de score obtenue permet de hiérarchiser les évènements futurs selon leur probabilité estimée d'occurrence. Une fois les scénarios économiques et financiers hiérarchisés, il est nécessaire de déterminer comment ces scénarios impactent le bilan, la solvabilité, la liquidité (selon ce que l'institution financière souhaite modéliser), et *in fine* la métrique cible sur laquelle s'appuiera le Reverse Stress Test. Pour cela, il faut aussi déterminer le lien entre cette métrique et les valeurs des différents facteurs de risque.

I Objectif

L'« objectif » correspond ici à la métrique cible sur laquelle le Reverse Stress Test va se baser. C'est en général une valeur dont elle veut se couvrir.

1 Cadre général

Plusieurs métriques peuvent figurer derrière cette valeur :

- Métrique de marché : Valeur du portefeuille d'actifs
- Métrique de risque : *Risk Weighted Assets*, Besoin en Fonds Propres
- Métrique de richesse : Fonds Propres
- Métrique réglementaire bancaire : Le ratio CET1 (Common Equity Tier 1)

$$\text{CET 1} = \frac{\text{Fonds Propres de catégorie 1}}{\text{Risk Weighted Assets}}$$

Les Fonds Propres de catégorie 1 sont définis par la réglementation européenne dans l'article 50 du CRR (cf. [3]) et correspondent à la partie la moins risquée des capitaux propres d'une institution financière.

Les *Risk Weighted Assets* se calculent, quant à eux, en pondérant les actifs par leur risque respectif.

Ces métriques ne sont pas exhaustives et peuvent s'appliquer à bien d'autres ratios de solvabilité, de liquidité etc. D'autres métriques plus pertinentes existent aussi pour les entreprises d'assurance comme cela a déjà été étudié dans le mémoire d'actuariat sur les Reverse Stress Tests de Dylan Lathuille (2019, cf. [8]).

2 Application à la Caisse des Dépôts

DFIN souhaite utiliser *in fine* le Reverse Stress Test sur leur métrique de solvabilité : l'Ecart à la cible.

$$\text{Ecart à la cible} = \text{Capital Economique} - \text{Besoin en Fonds Propres}$$

Le Capital Economique est composé des capitaux propres de l'institution et de la richesse des filiales financières régulées. Cette métrique peut être assimilée à la richesse de la Caisse des Dépôts.

Les Besoins en Fonds Propres représentent la métrique de risque de la Caisse des Dépôts. Elle résulte de la somme des risques associés aux risques matériels de la Caisse des Dépôts (dont font partie les risques action, taux et immobilier) estimés par des modèles spécifiques ou par des modèles plus communs de type Value-at-Risk.

L'Ecart à la cible peut donc être vu comme la différence entre la richesse et les risques de l'institution. Cette métrique est peu usuelle et propre à la Caisse des Dépôts.

3 Application à ce mémoire

Dans ce mémoire, la variation de **valeur du portefeuille** d'actifs de la Caisse des Dépôts fera office de valeur « objectif ». En effet, le lien entre facteurs de risque et valeur du portefeuille demande une étude moins approfondie pour DFIN pour les déterminer que l'Ecart à la cible.

L'objectif d'une perte de la valeur du portefeuille de la Caisse des Dépôts est fixé arbitrairement à **3 milliards d'euros**.

Objectif	
Impact cible (en euros)	-3 000 000 000

FIGURE 21 – Impact cible du Reverse Stress Test sur la variation de valeur du portefeuille d'actifs de la Caisse des Dépôts

II Sensibilités des facteurs de risque

Il reste donc maintenant à faire le lien entre les sensibilités des 3 facteurs de risque (CAC 40, OAT 10Y et IPD) et la valeur du portefeuille. Pour cela, DFIN a fait appel à des experts risques de la Direction des Risques (DRG) du Groupe Caisse des Dépôts qui leur ont transmis des sensibilités du portefeuille à une variation des facteurs de risque (valeurs factices dans le mémoire). Ces sensibilités sont supposées linéaires des facteurs de risque. Mais des sensibilités non linéaires pourraient être implémentées facilement, d'un point de vue mathématique, dans le module.

Sensibilités			
Facteurs	CAC 40	OAT 10Y	IPD
Variation	-1 pt	-1 pb	-1%
Impact (en euros)	-6 000 000	2 000 000	-200 000 000

FIGURE 22 – Tableau des sensibilités factices du portefeuille de la Caisse des Dépôts aux facteurs de risque

Ainsi à titre factice et illustratif :

- Une diminution du CAC 40 de 1 point entraînerait donc une perte sur la valeur du portefeuille d'actifs estimée à 6 millions d'euros.
- Une diminution des taux de 1 point de base (0,01% en absolu) entraînerait un gain sur la valeur du portefeuille d'actifs estimé à 2 millions d'euros.
- Une diminution de 1% (en relatif : multiplication par 0,99 de la valeur) de la valeur de l'IPD entraînerait une perte estimée à 200 millions d'euros sur la valeur du portefeuille d'actifs.

III Solveur Excel

L'ensemble de ce modèle est implémenté sur Excel. Le solveur Excel va donc permettre de tester algorithmiquement l'ensemble des valeurs envisageables et de déterminer quel triplet approchant du mieux possible l'objectif est le plus probable de survenir.

Le problème d'optimisation se résume donc à maximiser la fonction de score f sous contraintes selon les valeurs envisageables des facteurs de risque à 1 an du CAC 40, des OAT 10Y et de l'IPD (resp. X_{252}^1 , X_{252}^2 et X_{252}^3) en calculant les marginales correspondantes par rapport aux dernières valeurs récentes des facteurs de risque au 31 décembre 2020 (resp. X_{240}^1 , X_{240}^2 et X_{240}^3) :

$$\begin{cases} \arg \max_{x_1, x_2, x_3} & f(x_1, x_2, x_3) \\ \text{s.c} & g(x_1, x_2, x_3) = -3\,000\,000\,000 \end{cases}$$

avec

$$\begin{aligned} x_1 &= \ln\left(\frac{X_{252}^1}{X_{240}^1}\right) \quad \text{et} \quad X_{240}^1 = 5551,41 \\ x_2 &= (X_{252}^2 - X_{240}^2) \quad \text{et} \quad X_{240}^2 = -0,338\% \\ x_3 &= \ln\left(\frac{X_{252}^3}{X_{240}^3}\right) \quad \text{et} \quad X_{240}^3 = 202,53 \end{aligned}$$

et g la fonction des sensibilités :

$$g(x_1, x_2, x_3) = (e^{x_1} - 1) \cdot X_{240}^1 \times 6\,000\,000 - x_2 \times 2\,000\,000 + \frac{e^{x_3} - 1}{0,01} \times 200\,000\,000$$

Trois algorithmes d'optimisation sont disponibles sur Excel :

1. **Simplex PL** : Méthode du simplexe pour résoudre les problèmes linéaires.
2. **GRG non linéaire** : Méthode de réduction du gradient pour les problèmes simples non linéaires.
3. **Évolutionnaire** : Méthode d'optimisation heuristique pour résoudre les problèmes complexes.

Au vu de la fonction de score à optimiser qui est probablement non convexe, la méthode évolutionnaire semble plus adéquate. Cependant, il est nécessaire de réfléchir à des démarches afin d'éviter au maximum de se trouver dans un maximum local et non global. En cas de mise en production du module Excel, la discrétisation de l'espace de recherche serait une approche envisageable permettant de diminuer le risque de se trouver dans un maximum local. Par exemple, ces étapes de discrétisation pourraient consister à faire un déroulement de solutions itératives où les contraintes se resserrent progressivement selon la position de la dernière solution du solveur pour converger vers une solution optimale finale.

IV Définition des métriques de score

Le modèle de Reverse Stress Test n'intègre pas les tendances des différents des facteurs de risque, le modèle se concentre sur leurs volatilités (prises en compte dans le calibrage des écarts-types des lois normales des marginales des facteurs de risque). L'évènement reproduisant les dernières valeurs récentes, mesurées au 31 décembre 2020, est donc l'évènement maximisant la fonction de score et dont l'impact en valeur de portefeuille est nul. Cet évènement serait donc l'*output* du Reverse Stress Test pour un impact cible nul.

Dernières valeurs récentes			
Date	CAC 40	OAT 10Y	IPD
31/12/2020	5551,41	-0,338	202,53
Impact (en euros)	0		
Densité	0,0871		
Score sur 100 000	100 000		
Probabilité	Probable		

FIGURE 23 – Valeurs Récentes des facteurs de risque pour calculer les marginales

La probabilité du scénario est déterminée par sa **Densité** représentant la valeur de la fonction de score à trois dimensions déterminée antérieurement.

Dans le tableau ci-dessus, deux métriques sont utilisées afin de quantifier le résultat. En effet, la valeur de densité de la loi normale tridimensionnelle n'a pas une valeur interprétable. Rappelons que ce Reverse Stress Test est à destination de la gouvernance pour être un outil d'aide à la décision.

- **Score sur 100 000** : cette métrique consiste à ramener la densité en base 100 000 par rapport à celle du scénario le plus probable de densité 0,0871.
- **Probabilité** : cette seconde métrique est une notion qualitative en fonction des valeurs du score. Elle a été déterminée en fonction du score d'évènements qui peuvent être actuellement jugés comme « probables », « improbables » etc. à l'aide de seuils semblant cohérents. Ces seuils sont donc choisis de façon arbitraire mais sont présentés pour mieux illustrer la fonction de score.

Score Base 100 000		
Score compris entre		Probabilité
50 000	100 000	Probable
5 000	50 000	Peu Probable
500	5 000	Improbable
0	500	Très Improbable

FIGURE 24 – Métrique qualitative du score

V Solution du problème

Après avoir lancé plusieurs fois le solveur selon différentes valeurs d'initialisation, le scénario obtenu maximisant au mieux la fonction de score et atteignant un impact cible de - 3 milliards d'euros est le suivant :

Solveur			
Solutions	CAC 40	OAT 10 Y	IPD
	5116,12	-0,387	198,51
Impact (en euros)	- 2 611 735 865	9 789 181	- 397 553 497
	87,07%	-0,33%	13,25%
	- 2 999 500 181		
Millions d'écart avec l'objectif	-		
Densité	0,0787		
Score sur 100 000	90 348		
Probabilité	Probable		

FIGURE 25 – Tableau statistique de la solution du solveur

Ainsi, le solveur retient l'évènement d'un CAC 40 à 5116,12 points, des obligations à maturité de l'Etat à un taux de -0,387% et les IPD à 198,51 points comme l'évènement le plus probable provoquant une perte de valeurs de 3 milliards d'euros du portefeuille de la Caisse des Dépôts.

La densité maximale des scénarios d'impact de - 3 milliards d'euros est donc de 0,0787. Cette densité correspond à un score de 90 348 en base 100 000 qui lui affecte le commentaire qualitatif de « Probable ».

Le module Excel permet aussi de comparer les contributions dans l'atteinte de l'objectif entre les trois facteurs de risque. Dans cet évènement, solution de l'optimisation, le CAC 40 est la raison majoritaire de la perte en valeur du portefeuille de la CDC, la contribution des facteurs de risque dans l'impact cible est très grandement inégale. En effet, le CAC 40 contribue majoritairement à plus de 80% de l'impact alors que les OAT 10Y semblent non significatifs et contribuent à moins de 1%. Ces contributions inégales peuvent être causées par un déséquilibre des sensibilités de la CDC aux facteurs de risque déterminées précédemment (cf. Partie 4-II).

VI Sensibilités du solveur

Dans ce module de Reverse Stress Test sur Excel, trois types de valeurs statistiques sont déterminantes pour la sélection de la solution optimale par le solveur : les sensibilités de la CDC aux trois facteurs de risque, les corrélations deux-à-deux entre ces facteurs et les écarts-types des trois lois marginales.

Les valeurs de sensibilité de la CDC aux facteurs de risque sont des données déterminées par la Direction des Risques et donc supposées exogènes. Mais les autres valeurs proviennent directement de l'estimation statistique réalisée, il peut donc être intéressant de mesurer la sensibilité des trois *outputs* du Reverse Stress Test, solutions de l'optimisation, à ces estimations statistiques.

Pour cela, une première sensibilité de 25% (à la hausse ou à la baisse) est appliquée simultanément sur les trois écarts-types estimés des marginales des facteurs de risque et une seconde sensibilité de 25% (à la hausse ou à la baisse) simultanément sur les trois coefficients de corrélation linéaires des facteurs de risque deux-à-deux. Ces deux sensibilités sont testées et séparément et cumulativement.

Sensibilités		Solution			Evènement		
		CAC 40	OAT 10Y	IPD	Score sur 100 000	Probabilité	
		5116,12	-0,387	198,51	90 348	Probable	
Ecart-type	Corrélation	Variations sur la solution			Variations sur le score		
A	+25%	-	0,06	- 0,001	0,02	3 374	Probable
B	-25%	-	0,04	-	0,03	6 829	Probable
C	-	+25%	8,59	- 0,012	0,52	451	Probable
D	-	-25%	8,98	0,011	0,59	459	Probable
E	+25%	+25%	8,71	- 0,011	0,53	3 662	Probable
F	-25%	+25%	9,02	- 0,011	0,54	6 115	Probable
G	+25%	-25%	8,47	0,012	0,57	3 058	Probable
H	-25%	-25%	8,51	0,014	0,57	7 611	Probable

FIGURE 26 – Sensibilités du scénario optimal et des métriques associées à un choc sur l'écart-type et/ou la corrélation

Ces différentes sensibilités combinées entre elles soulignent un impact mesuré sur les solutions du solveur (colonnes 3 à 5). En effet, le choc sur l'écart-type des marginales (colonne 1) pris indépendamment semble non significatif (comparaison des lignes A et B avec la solution, des lignes C-E-F entre elles et des lignes D-G-H entre elles), ne jouant sur la solution du CAC 40 et l'IPD que de l'ordre du dixième et ne jouant sur la solution des OAT 10Y que de l'ordre du millième. Cependant, le choc de 25% sur les corrélations (colonne 2) semble quant à lui avoir un impact plus conséquent (comparaison des lignes C et D avec la solution, des lignes A-E-G entre elles et des lignes B-F-H entre elles) faisant varier l'*output* du CAC 40 d'une dizaine de points, l'*output* des OAT 10Y d'un point de base et l'*output* de l'IPD d'un demi point. Ces sensibilités peuvent sembler marginales mais le choc de 25% a été choisi très arbitrairement, il peut donc être opportun d'étudier ce choix de corrélation plus attentivement.

A contrario, c'est le choc de 25% sur les écarts-types des marginales (colonne 1) qui semblent souligner l'impact sur les métriques de score (colonnes 6 à 8), faisant varier le score sur 100 000 d'une dizaine de milliers de points (comparaison des lignes A et B avec la solution, des lignes C-E-F entre elles et des lignes D-G-H entre elles). **Ces sensibilités soulignent donc une faiblesse de la métrique de score** dans la mesure où son sens varierait d'une année sur l'autre en fonction des écarts-types des marginales des facteurs de risque alors même que les *outputs* seraient très sensiblement identiques. De plus, le choc de 25% a été choisi également arbitrairement et restait une variation brute assez faible par rapport aux valeurs des écarts-types retenus (cf. Partie 3-III).

C'est pourquoi, il est retenu d'approfondir les pistes de sensibilités aux hypothèses statistiques endogènes au modèle de Reverse Stress Test mis en place au sein du module Excel (cf. Partie 5).

VII Contribution des facteurs de risque à l'impact cible

Il peut également être intéressant d'étudier si les facteurs de risque ont vraiment contribué de façon inégalitaire à l'impact cible à cause des sensibilités de la Caisse des Dépôts à ces facteurs. Pour cela, le module Excel est utilisé pour plusieurs impacts cibles et identifie si les contributions des facteurs de risque sont structurellement inégales ou si elles dépendent de l'impact cible prédéfini.

Impact cible	Solutions			Scores		Contributions à l'impact		
	CAC 40	OAT 10Y	IPD	Sur 100 000	Probabilité	CAC 40	OAT 10Y	IPD
- 1 Md €	5406,16	-0,354	201,21	98 934	Probable	87,28%	-0,31%	13,03%
- 2 Md €	5261,05	-0,370	199,88	95 704	Probable	87,22%	-0,32%	13,11%
- 3 Md €	5116,12	-0,387	198,51	90 348	Probable	87,07%	-0,33%	13,25%
- 4 Md €	4972,10	-0,404	197,13	83 128	Probable	86,99%	-0,33%	13,34%
- 5 Md €	4828,65	-0,419	195,70	74 357	Probable	86,82%	-0,33%	13,50%
- 6 Md €	4686,60	-0,430	194,20	64 532	Probable	86,57%	-0,31%	13,73%
- 7 Md €	4542,45	-0,454	192,79	54 195	Probable	86,57%	-0,33%	13,76%
- 8 Md €	4402,72	-0,470	191,13	43 941	Peu Probable	86,24%	-0,33%	14,09%
- 9 Md €	4259,31	-0,486	189,69	34 275	Peu Probable	86,23%	-0,33%	14,10%

FIGURE 27 – Contributions à différents impacts cibles pour chaque facteur de risque

Tout d'abord, le module Excel souligne une certaine cohérence entre les différents impacts cibles. En effet, plus l'impact cible est conséquent, plus les *outputs*, solutions du solveur, sont à la baisse (et donc s'éloignent de leurs dernières valeurs récentes) et plus le score de probabilité de l'évènement diminue. Il y a donc une corrélation linéaire positive entre l'impact et les cours des facteurs de risque.

Quant à la contribution à l'impact cible de chaque facteur de risque, une certaine constance est soulignée par le tableau. Ce sont donc bien les sensibilités de la CDC aux facteurs de risque qui sont assez inégales entre elles. Le CAC 40 contribue sur l'ensemble des solutions à plus de 80% de l'impact cible. Ce résultat est cohérent dans la mesure où le risque action est le risque principal de la Caisse des Dépôts. L'étude approfondie des sensibilités de la CDC aux facteurs de risque ne fait pas l'objet de ce mémoire, les données ont été transmises par la Direction des Risques. Cependant, ce tableau souligne l'importance de ces hypothèses exogènes au modèle, les sensibilités de la CDC aux facteurs de risque demanderont donc une attention toute particulière dans la vie du modèle.

Alternatives aux hypothèses simplificatrices du module Excel

Le module de Reverse Stress Test permet ainsi d'identifier l'évènement financier jugé le plus probable amenant une perte financière fixée préalablement. Cependant, afin de pouvoir le réaliser (notamment sur Excel), beaucoup d'hypothèses mathématiques fortes ont dû être posées.

Dans l'objectif de réaliser un module le plus proche possible de la réalité, la suite logique est de lever certaines hypothèses en commençant par celles qui sont les plus fortes.

Pour cela, voici les trois pistes envisagées dans la suite de ce mémoire pour *challenge* le module Excel :

- **Calibrer des lois marginales par d'autres lois que des lois normales en fonction des facteurs de risque** (cf. Partie 5 - I)
- **Réfléchir à l'utilisation d'une copule non gaussienne** (cf. Partie 5 - II)
- **Quantifier l'impact du choix du coefficient de corrélation utilisé** (cf. Partie 5 - III)

Pour l'ensemble de ces points, des tests statistiques ou des études statistiques seront réalisés afin de relativiser la matérialité de l'hypothèse prise puis des pistes alternatives seront étudiées et comparées. Ces pistes alternatives peuvent exiger d'implémenter le Reverse Stress Test dans un outil plus puissant que Excel (sur R par exemple). Les modèles implémentés seront ensuite comparés avec celui sur Excel.

I Alternatives aux lois marginales gaussiennes

1 Tests de normalité

Dans l'optique de réaliser un premier module de Reverse Stress Test sur Excel, des lois normales ont été calibrées sur les données des marginales afin d'obtenir une formule fermée pour la fonction de score avec une structure de corrélation gaussienne. Cependant, les lois normales ont deux caractéristiques assez fortes : elles sont symétriques et ont des queues de distribution fines. Ces deux caractéristiques sont généralement rarement vérifiées. C'est pourquoi cette hypothèse peut être considérée comme forte. Cette hypothèse a notamment interpellé la Direction des Risques lors de la mission (concluante) de validation du module Excel.

Différents tests statistiques existent pour accepter ou rejeter la normalité d'un jeu de données. En voici par exemple deux : le test de Shapiro-Wilk et le test de Jarque-Bera.

Test de Jarque-Bera

Ce test a été publié en 1980 par Carlos M. Jarque et Anil K. Bera (cf. [14]). L'hypothèse nulle considère que l'échantillon de taille n x_1, \dots, x_n est normalement distribué.

La statistique de test vaut :

$$JB = \frac{n}{6} \left(S^2 + \frac{(K-3)^2}{4} \right)$$

avec S le coefficient d'asymétrie :

$$S = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}) \right)^{3/2}}$$

et K le coefficient d'aplatissement :

$$K = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2}$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

L'hypothèse nulle H_0 revient donc à $S = 0$ et $K = 3$.

Sous H_0 , la statistique est supposée suivre une loi du χ^2 à deux degrés de liberté.

L'hypothèse nulle de normalité de l'échantillon est rejetée (resp. acceptée) si, pour un niveau de confiance α , la p.value (calculée grâce à la statistique JB) est inférieure à α (resp. supérieure).

Test de Shapiro-Wilk

Ce test a été publié en 1965 par Samuel Sanford Shapiro et Martin Bradbury Wilk (cf. [18]). L'hypothèse nulle considère que l'échantillon de taille n x_1, \dots, x_n est normalement distribué. La statistique de test vaut :

$$W = \frac{\left(\sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} x_{(n-i+1)} - x_{(i)} \right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

avec

$x_{(i)}$ la i^{me} plus petite valeur de l'échantillon

et

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

L'hypothèse nulle de normalité de l'échantillon est rejetée (resp. acceptée) si, pour un niveau de confiance α , la statistique W est inférieure à α (resp. supérieure).

		Marginales		
		p.value	CAC 40	OAT 10Y
Tests statistiques	Shapiro-Wilk	5,73E-05	1,56%	3,67E-04
	Jarque-Bera	7,68E-08	73,98%	<2,20E-16

FIGURE 28 – p.value des tests statistiques de normalité sur les marginales des trois facteurs de risque

Pour un niveau de confiance à 10%, d'après le tableau ci-dessus, la normalité des marginales du CAC 40 et de l'IPD sont rejetées pour les deux tests statistiques. Le test de Shapiro-Wilk rejette aussi la normalité des marginales des OAT 10Y. Cependant, le test de Jarque Bera estime que les marginales des OAT sont symétriques et ont un aplatissement équivalent à celle d'une loi normale.

L'hypothèse de normalité des marginales du CAC 40 et de l'IPD peut donc bien être considérée comme forte. Mais cette hypothèse semble moins forte pour les OAT 10Y.

Il est donc intéressant d'étudier de nouvelles lois marginales pertinentes afin de remplacer la loi normale des marginales, *a minima*, du CAC 40 et de l'IPD.

2 Alternatives à la loi normale

2-A Lois théoriques alternatives

DFIN actualise annuellement un cahier des marginales sur l'ensemble des actions au bilan de la Section Générale. Pour chaque cours, plusieurs lois théoriques sont testées et comparées. Il a donc semblé pertinent de réutiliser certaines de ces lois qui sont, en plus de la **loi normale**, la **loi normale asymétrique** et la **loi normale-inverse gaussienne**. D'autres lois auraient pu être envisagées comme la loi logistique et la loi de Johnson's SU (cf. Annexe 7), les critères nécessaires étant que les lois aient des supports définis sur \mathbb{R} et qu'elles soient globalement en forme de cloche.

Loi normale asymétrique

Cette loi permet de généraliser la loi normale à des coefficients d'asymétrie non nuls. Cette loi est composée de 3 paramètres : un paramètre de position (noté ξ), un paramètre d'échelle (noté ω) et un paramètre de forme (noté α).

La fonction de densité est définie telle que :

$\forall x \in \mathbb{R}$

$$f_{\xi, \omega, \alpha}(x) = \frac{1}{\omega\pi} e^{-\frac{(x-\xi)^2}{2\omega^2}} \int_{-\infty}^{\alpha(\frac{x-\xi}{\omega})} e^{-\frac{t^2}{2}} dt$$

avec

$$\xi, \alpha \in \mathbb{R}$$

$$\omega \in \mathbb{R}_+^*$$

Il n'existe pas de formules fermées pour estimer les 3 paramètres mais ceux-ci peuvent être estimés numériquement.

Loi normale-inverse gaussienne (notée NIG)

La loi fait son apparition dans un article de Ole Barndorff-Nielsen en 1977 (cf. [12]). Cette distribution a été définie à partir d'une loi mélange. Elle est composée de 4 paramètres : un paramètre d'épaisseur de la queue de distribution (noté α), un paramètre d'asymétrie (noté β), un paramètre d'échelle (noté δ) et un paramètre de position (noté μ).

La fonction de densité est définie telle que :

$\forall x \in \mathbb{R}$

$$f_{\alpha, \beta, \delta, \mu}(x) = e^{\delta\sqrt{\alpha^2 - \beta^2}} \frac{\alpha\delta}{\pi\sqrt{\delta^2 + (x - \mu)^2}} K_1(\alpha\sqrt{\delta^2 + (x - \mu)^2}) e^{\beta(x - \mu)}$$

avec

K_1 la fonction de Bessel modifiée de troisième nature d'ordre 1

et

$$\alpha, \beta, \delta, \mu \in \mathbb{R}$$

Les 4 paramètres peuvent être aussi estimés numériquement.

2-B Statistiques de comparaison

Maintenant que les trois jeux de lois marginales à comparer ont été déterminés (lois marginales normales, normales asymétriques et NIG), il reste à déterminer un processus de sélection pour choisir la loi adéquate. DFIN décide dans son cahier des marginales de partir de la loi la plus simple, donc la plus rapide en exécution informatique, à la loi la plus complexe. Pour ce mémoire cela revient donc à commencer par la loi normale pour terminer sur la NIG.

La même méthode est retenue mais les critères à tester doivent être réfléchis pour être adéquats avec l'objectif des Reverse Stress Tests, très différent du modèle de calcul de la VaR Actions visé *in fine* par le cahier des marginales.

Beaucoup de critères peuvent être retenus pour comparer une loi théorique avec un jeu de données. Voici une liste non-exhaustive des critères envisageables concernant les :

- Tests statistiques sur la distribution des lois
- Moments (espérance, variance)
- Quantiles
- Coefficients de forme (asymétrie, aplatissement)
- Mesures de risque (VaR, CVaR)

Il a été décidé de ne pas complexifier les comparaisons des trois lois marginales avec trop de critères. Trois critères ont donc été finalement retenus.

Le premier critère repose sur le test statistique de Kolmogorov-Smirnov de comparaison des fonctions de répartition.

Test statistique de Kolmogorov-Smirnov

Ce test statistique non-paramétrique, introduit en 1939 par Vladimir Ivanovitch Smirnov grâce aux travaux antérieurs de Andreï Nicolaïévitch Kolmogorov (1933, cf. [15]), permet de quantifier la comparaison de distribution entre deux échantillons, par la comparaison de leurs fonctions de répartition empiriques respectives.

Soient deux échantillons X et Y, de tailles pouvant être différentes n et m, pour valeurs x_1, \dots, x_n et y_1, \dots, y_m . L'hypothèse nulle est que les deux échantillons sont identiquement distribués.

La statistique de test vaut :

$$D_{n,m} = \sup_t |F_X(t) - F_Y(t)|$$

avec F_X et F_Y les fonctions de répartition empiriques des deux jeux de données X et Y

$$F_X(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{x_i \leq t} \text{ et } F_Y(t) = \frac{1}{m} \sum_{i=1}^m \mathbb{1}_{y_i \leq t}$$

L'hypothèse nulle (équivalence des distributions) est rejetée, pour un niveau de confiance α , si :

$$D_{n,m} > \sqrt{-\ln\left(\frac{\alpha}{2}\right) \frac{1 + \frac{m}{n}}{2m}}$$

Ce test est pertinent comme critère de sélection de marginale car il permet de comparer la loi théorique calibrée sur le jeu de données, en réalisant un tirage aléatoire, avec le jeu de données initial.

Numériquement l'idée est donc premièrement d'estimer les paramètres des différentes lois pour un jeu de données puis de réaliser un tirage de taille conséquente pour que sa fonction de répartition empirique ne varie pas. Enfin la dernière étape est de réaliser ce test qui permet de calculer la p.value à partir de cette statistique. Si cette p.value est inférieure (resp. supérieure) à 10% alors les deux jeux de données sont supposés de distribution significativement différentes (resp. équivalentes).

Le second critère repose sur le coefficient d'asymétrie des distributions.

Coefficient d'asymétrie

Le coefficient d'asymétrie S permet de mesurer, assez logiquement, l'asymétrie d'une distribution. Pour un échantillon de taille n x_1, \dots, x_n :

$$S = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})\right)^{3/2}} \quad \text{avec} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Un jeu de données est dit d'asymétrie positive (resp. négative) si sa médiane est inférieure à sa moyenne (resp. supérieure). Un jeu de données d'asymétrie nulle, *i. e.* symétrique, a une moyenne égale à la médiane.

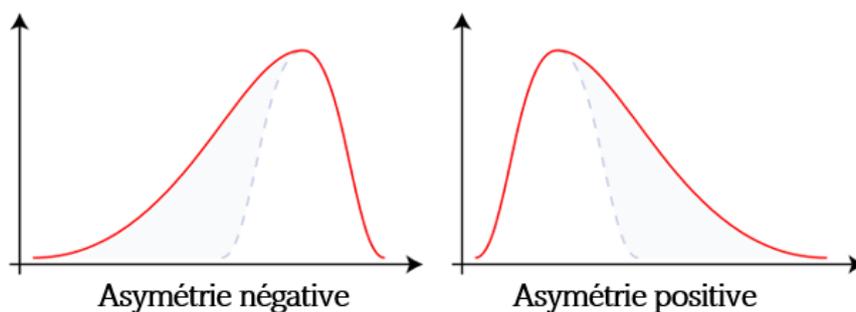


FIGURE 29 – Illustration graphique du coefficient d'asymétrie

Ce critère est pertinent à retenir car il permet de conserver une asymétrie (si existante) entre la baisse et la hausse des cours des facteurs de risque.

Numériquement l'idée est donc premièrement d'estimer les paramètres des différentes lois pour un jeu de données puis de réaliser un tirage de taille conséquente pour que sa fonction de répartition empirique ne varie pas. Enfin la dernière étape est de calculer le coefficient d'asymétrie sur les deux jeux de données et de les comparer. Le critère a été fixé à deux dixièmes d'écart maximum (resp. minimum) entre les valeurs des coefficients pour accepter (resp. rejeter) la similitude d'asymétrie.

Le troisième critère repose sur la mesure de risque de la *Conditional Tail Value-at-Risk*.

Conditional Value-at-Risk

La CVaR aussi appelée *Tail Value-at-Risk* ou *Average Value-at-Risk* représente la perte attendue moyenne à partir d'un certain quantile de distribution. Cette condition permet de moyenner les pertes au sein des événements considérés comme les plus extrêmes.

Posons deux définitions : une CVaR Gauche pour observer les valeurs extrêmes de la queue de distribution gauche (valeurs extrêmes négatives des marginales *i. e.* à la baisse extrême des cours) et une CVaR Droite pour observer les valeurs extrêmes de la queue de distribution droite (valeurs extrêmes positives des marginales *i. e.* à la hausse extrême des cours).

Pour le quantile extrême α fixé à observer, la CVaR Gauche d'une variable aléatoire X (à support sur \mathbb{R}) vaut :

$$\text{CVaR Gauche}_\alpha(X) = \frac{1}{\alpha} \int_0^\alpha \text{VaR}_\gamma(X) d\gamma$$

et la CVaR Droite de cette même variable vaut :

$$\text{CVaR Droite}_\alpha(X) = \frac{1}{\alpha} \int_{1-\alpha}^1 \text{VaR}_\gamma(X) d\gamma$$

avec $\text{VaR}_\gamma(X)$ la Value-at-Risk de quantile γ de la v.a X :

$$\text{VaR}_\gamma(X) = \inf\{x \in \mathbb{R}, P(X \leq x) \geq \gamma\}$$

Ce critère est pertinent à retenir pour aiguiller le bon choix de loi marginale car il permet de conserver l'aspect extrême des tests de résistance. La Value-at-Risk (notée VaR) aurait également pu être retenue mais la CVaR a la caractéristique supplémentaire de respecter strictement les mesures de risque (le critère de sous-additivité n'étant pas validé par la VaR).

Numériquement l'idée est donc premièrement d'estimer les paramètres des différentes lois pour un jeu de données puis de réaliser un tirage de taille conséquente pour que sa fonction de répartition empirique ne varie pas. Enfin la dernière étape est de calculer la CVaR Gauche pour les facteurs de risque dont le risque est à la baisse (CAC 40, IPD) et la CVaR Droite pour les facteurs de risque dont le risque est à la hausse (OAT 10Y) afin de les comparer avec celles calculées sur le jeu de données historique. Il est préférable que le modèle soit prudent, la loi théorique est donc retenue uniquement si sa CVaR a un impact financier négatif sur la Caisse des Dépôts plus important que l'historique.

Le quantile est fixé à 10% pour observer des événements relativement rares mais pas trop extrêmes : les tests de résistance doivent représenter des événements rares, graves mais tout de même plausibles.

2-C Proposition de procédure de sélection des marginales

Les lois théoriques suivantes sont donc testées l'une après l'autre :

1. Loi Normale
2. Loi Normale asymétrique
3. Loi NIG

Les paramètres de chaque loi sont estimés pour chaque facteur de risque, puis un tirage de 100 000 marginales dites théoriques est réalisé afin de tester les critères retenus entre l'historique (noté h) et chaque loi théorique (notée th).

Les conditions d'acceptation des tests retenus sont les suivants :

- Critère de Kolmogorov-Smirnov :

$$p.value^{K-S} > 10\%$$

- Critère d'asymétrie :

$$|S_{th} - S_h| < 0,2$$

- Critère de prudence :

$$CVaR_{10\%}^{Gauche}(X_{th}) < CVaR_{10\%}^{Gauche}(X_h)$$

ou

$$CVaR_{10\%}^{Droite}(X_{th}) > CVaR_{10\%}^{Droite}(X_h)$$

Si un des trois tests n'est pas vérifié, alors la loi est rejetée et c'est la suivante qui est testée, selon l'ordre précédemment explicité.

Etant données les différences de complexité des trois lois, le critère de Kolmogorov et le critère de prudence ont tendance à être de plus en plus vérifiés. La loi normale asymétrique se base sur le coefficient d'asymétrie pour estimer les paramètres, ce critère est donc toujours vérifié pour cette loi.

Sont présentées ci-dessous, les valeurs des trois critères des trois lois théoriques calibrées sur l'historique respectif de chaque facteur de risque.

CAC 40			Critères		
			p.value Kolmogorov	Coefficient d'asymétrie	CVaR 90% à gauche
Loi marginale	Historique	Valeur		-0,557	-0,103
	Normale	Valeur	21,18%	0,007	-0,093
		Test	OK	KO	KO
	Normale asymétrique	Valeur	76,98%	-0,57	-0,103
		Test	OK	OK	OK
	NIG	Valeur	97,36%	-0,863	-0,107
Test		OK	KO	OK	

FIGURE 30 – Valeurs des critères de marginale du CAC 40

Concernant le CAC 40, la loi normale est bien rejetée comme le suggéraient les tests de normalité réalisés précédemment. La loi normale asymétrique, quant à elle, vérifie bien les trois critères, elle est donc retenue comme loi marginale. La loi NIG n'a pas besoin d'être testée, les valeurs des critères sont données à titre illustratif.

OAT 10Y			Critères		
			p.value Kolmogorov	Coefficient d'asymétrie	CVaR 90% à droite
Loi marginale	Historique	Valeur		0,003	0,341
	Normale	Valeur	25,95%	0,001	0,341
		Test	OK	OK	OK
	Normale asymétrique	Valeur	26,33%	-0,013	0,338
		Test	OK	OK	KO
	NIG	Valeur	51,41%	0,178	0,356
Test		OK	OK	OK	

FIGURE 31 – Valeurs des critères de marginale des OAT 10Y

Les critères confirment que la loi marginale normale ne semble pas être une hypothèse très forte sur les OAT 10Y. Elle est donc retenue comme loi marginale.

IPD			Critères		
			p.value Kolmogorov	Coefficient d'asymétrie	CVaR 90% à gauche
Loi marginale	Historique	Valeur		-1,033	-0,018
	Normale	Valeur	2,46%	-0,01	-0,03
		Test	KO	KO	OK
	Normale asymétrique	Valeur	<i>L'estimation des paramètres ne converge pas</i>		
		Test			
	NIG	Valeur	95,35%	-1,05	-0,026
Test		OK	OK	OK	

FIGURE 32 – Valeurs des critères de marginale de l'IPD

Cette procédure de sélection confirme que le calibrage d'une loi normale sur les marginales de l'IPD n'est pas adéquat. L'estimation numérique des paramètres de la loi normale asymétrique n'a pas convergé, l'arborescence se poursuit donc directement sur la loi NIG. Celle-ci valide les trois critères et est donc retenue comme loi la plus adéquate pour représenter le jeu de marginales de l'IPD.

2-D Lois marginales retenues

Les lois marginales théoriques personnalisées retenues sont donc les suivantes :

- CAC 40 : Loi normale asymétrique
- OAT 10Y : Loi normale
- IPD : Loi normale inverse-gaussienne

3 Estimateurs à noyau

Il existe une autre méthode d'estimation de densité : l'estimateur à noyau. L'estimateur à noyau est une méthode d'estimation statistique non-paramétrique, elle a donc l'avantage de ne pas avoir à analyser un catalogue de lois théoriques pour trouver la plus pertinente.

Cette méthode a été créée par Emanuel Parzen (1962, cf. [16]) et Murray Rosenblatt (1956, cf. [17]) au milieu du XX^{ème} siècle. Cette méthode est une généralisation de la méthode d'estimation par histogramme où la densité en un point est estimée par des observations à proximité de x . L'estimateur à noyau permet de remplacer l'histogramme par une fonction continue centrée en x .

Cette fonction est dépendante d'un paramètre de largeur h permettant de déterminer la fenêtre des observations considérées comme à proximité. Plus cette fenêtre est large, moins l'estimation sera précise et se rapprochera de la moyenne mais plus la fenêtre est fine, plus l'estimation convergera vers des probabilités discrètes de chaque donnée d'historique. Ce paramètre est à estimer selon la particularité du jeu de données.

La fonction continue, appelée noyau, est aussi un paramètre de l'estimateur à noyau, beaucoup de fonctions peuvent être utilisées. Le choix usuel est la loi normale centrée réduite, c'est le choix retenu.

Estimateur à noyau gaussien

Pour un jeu de données de taille n , x_1, \dots, x_n , et une fenêtre h l'estimateur à noyau gaussien se définit par :

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)$$

avec K le noyau gaussien :

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

Cet estimateur à noyau utilisé sur les jeux de marginales des différents facteurs de risque permet donc bien d'avoir une fonction de densité pouvant aussi remplacer les marginales normales du module Excel. Cette fonction de densité, par sa complexité et sa proximité du jeu de données, valide nécessairement les critères de Kolmogorov-Smirnov et d'asymétrie. Cette fonction a aussi tendance à être plus conservatrice, sur le critère de CVaR, que les trois lois théoriques précédentes. Son implémentation est cependant plus lourde et consommatrice de temps.

4 Bilan sur les marginales retenues

Il est maintenant clair que l'hypothèse de loi normale est une hypothèse forte et des hypothèses alternatives moins fortes ont été fixées. Mais il reste à savoir comment quantifier l'importance de cette hypothèse de marginale sur les résultats du Reverse Stress Test.

Trois catégories de Reverse Stress Tests (RST) peuvent donc être envisagées concernant les différents choix de marginales :

- RST utilisant des **lois marginales normales**
- RST utilisant des **lois marginales personnalisées** (cf. Partie 5-I_2D)
- RST utilisant des **lois marginales estimées par un estimateur à noyau gaussien**

II Alternatives à la copule gaussienne

Dans le module Excel, concilier lois marginales normales avec une copule gaussienne a permis d'obtenir la formule fermée de loi normale multidimensionnelle. Cependant, la copule gaussienne est loin d'être la seule existante. C'est pourquoi il est intéressant d'étudier les alternatives à la copule gaussienne afin de retenir la structure de corrélation qui représente le mieux la corrélation entre les facteurs de risque.

1 Définition d'une copule

Le nom de copule a été introduit par Abe Sklar en 1959 (cf. [19]). C'est un élément mathématique qui permet de créer une structure de corrélation multivariée à partir de jeux de données univariés.

Définition d'une fonction copule

Une copule multivariée de taille n est une fonction de $[0, 1]^n \rightarrow [0, 1]$ qui vérifie les propriétés suivantes :

1. $C(0, \dots, 0) = 0$
2. $C(1, \dots, 1) = 1$
3. $C(u, 1, \dots, 1) = C(1, u, \dots, 1) = \dots = C(1, 1, \dots, u) = u$
4. $\exists i \in \llbracket 1, n \rrbracket$ tel que $u_i = 0 \Rightarrow C(u_1, \dots, u_n) = 0$
5. C est non-n-décroissante

Cela signifie pour exemple que si $n = 2$

$\forall u_1, u_2, v_1, v_2$ tel que $0 \leq u_1 \leq u_2 \leq 1$ et $0 \leq v_1 \leq v_2 \leq 1$:

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$$

La copule s'applique parfaitement pour créer une relation entre une fonction de répartition multivariée et les fonctions de répartition univariées, par l'introduction de corrélations.

Théorème de Sklar

Soit X un vecteur aléatoire composé de n variables aléatoires X_1, \dots, X_n de fonctions de répartition F_1, \dots, F_n .

Sa fonction de répartition est donc définie telle que :

$$F_X(x_1, \dots, x_n) = \mathcal{P}(X_1 \leq x_1, \dots, X_n \leq x_n)$$

D'après le théorème de Sklar,

$$\exists C \text{ une copule telle que } F_X(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n))$$

La réciproque est aussi vérifiée :

$$\forall C \text{ une copule, } C(F_1, \dots, F_n) \text{ est une fonction de répartition multivariée}$$

2 Définitions des copules envisagées

Il a été décidé de comparer trois autres copules usuelles à la copule **gaussienne** : la copule de **Student**, la copule de **Gumbel** et la copule de **Clayton**. Les caractéristiques communes et spécifiques de ces copules sont détaillées par Arthur Charpentier (2013, cf. [21]) ainsi que dans plusieurs mémoires d'actuariat, dont l'utilisation et la comparaison de copules sont l'objet d'une étude plus approfondie, comme celui d'Arnaud Duval (2012, cf. [5]) de Guillaume Galataud (2010, cf. [6]), d'Esterina Masiello (2010, cf. [9]) et d'Emmanuelle Scheid (2011, cf. [10]).

2-A Copules elliptiques

Copule gaussienne

La copule gaussienne est la copule la plus courante, c'est une copule elliptique qui est construite à partir de la distribution normale multivariée.

La copule gaussienne peut se définir telle que :

$$\forall u_1, \dots, u_n \in [0, 1]$$

$$C_{\Sigma}^{\text{Gauss}}(u_1, \dots, u_n) = \Phi_{\Sigma}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n))$$

avec Φ^{-1} la fonction de répartition normale inverse centrée réduite et Φ_{Σ} la fonction de répartition normale multivariée de matrice de corrélation $\Sigma \in [-1, 1]^{n \times n}$.

L'inconvénient de la copule gaussienne est qu'elle a tendance à minimiser la corrélation extrême, sur les queues de distribution. Or en finance, lors d'évènements stressés extrêmes, les facteurs de risque ont tendance à être d'autant plus corrélés.

Copule de Student

La copule de Student est aussi souvent utilisée pour sa simplicité de compréhension. A l'instar de la copule gaussienne, c'est une copule elliptique qui est construite à partir d'une distribution multivariée : celle de Student.

La copule de Student peut se définir telle que :

$$\forall u_1, \dots, u_n \in [0, 1]$$

$$C_{\nu, \Sigma}^{\text{Student}}(u_1, \dots, u_n) = t_{\nu, \Sigma}(t_{\nu}^{-1}(u_1), \dots, t_{\nu}^{-1}(u_n))$$

avec t_{ν}^{-1} la fonction de répartition de Student inverse de degré ν et $t_{\nu, \Sigma}$ la fonction de répartition mutivariée de Student de degré de liberté ν et de matrice de corrélation $\Sigma \in [-1, 1]^{n \times n}$.

Plus ce degré augmente plus la copule de Student converge vers la copule gaussienne.

$$C_{\nu, \Sigma}^{\text{Student}} \xrightarrow{\nu \rightarrow +\infty} C_{\Sigma}^{\text{Gauss}}$$

La copule de Student permet d'intégrer un coefficient de dépendance de queue à travers le degré de liberté ν , et ainsi de différencier la corrélation des extrêmes de celle du reste du jeu de données.

2-B Copules archimédiennes

Copule de Gumbel

La copule de Gumbel fait partie de la famille des copules archimédiennes et peut se définir telle que :

$\forall u_1, \dots, u_n \in [0, 1]$

$$C_\theta(u_1, \dots, u_n) = \exp \left(- \left[\sum_{i=1}^n (-\ln u_i)^\theta \right]^{\frac{1}{\theta}} \right)$$

avec $\theta \geq 1$.

Dans le cas où $n = 2$, θ est un paramètre lié de façon bijective avec le tau de Kendall (mesure de corrélation de rang) noté τ défini en Annexe 5 :

$$\tau = 1 - \frac{1}{\theta} \iff \theta = \frac{1}{1 - \tau}$$

Le paramètre θ étant nécessairement supérieur à 1, la copule n'appréhende que les dépendances positives.

La copule de Gumbel a la particularité de corrélérer fortement positivement les facteurs de risque au niveau de la queue de corrélation droite, c'est à dire de corrélérer fortement la hausse conjointe de facteurs de risque. La corrélation à la baisse est, elle, peu corrélée.

Copule de Clayton

La copule de Clayton fait aussi partie de la famille des copules archimédiennes et peut se définir telle que :

$\forall u_1, \dots, u_n \in [0, 1]$

$$C_\alpha(u_1, \dots, u_n) = \left(\sum_{i=1}^n u_i^{-\alpha} - (n-1) \right)^{-\frac{1}{\alpha}}$$

avec $\alpha > 0$

Dans le cas où $n = 2$, α est lui aussi un paramètre lié de façon bijective avec le tau de Kendall :

$$\tau = \frac{\alpha}{\alpha + 2} \iff \alpha = \frac{2\tau}{1 - \tau}$$

Le paramètre α étant nécessairement positif, la copule n'appréhende que les dépendances positives.

La copule de Clayton est un peu la symétrie de la copule de Gumbel, au niveau de la queue de distribution où la corrélation est amplifiée. En effet, la copule de Clayton corrèle fortement positivement les facteurs de risque au niveau de la queue de corrélation gauche, c'est à dire qu'elle corrèle fortement la baisse conjointe de facteurs de risque. Inversement, la corrélation à la hausse est, elle, peu corrélée.

3 Choix des copules adéquates

L'objectif est de comparer la corrélation historique entre les jeux de données avec des corrélations théoriques au sein de copules afin de sélectionner la plus adéquate.

Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire dont les variables aléatoires ont des fonctions de répartition théoriques F_1, \dots, F_n mais dont le vecteur a , lui, une fonction de répartition empirique F_X^h .

D'après le théorème de Sklar, la copule la plus adéquate vérifie :

$$C(u_1, \dots, u_n) = F_X^h(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n))$$

$\forall i \in [1, n]$

$F_i^{-1}(u_i)$ représente le quantile u_i de la variable X_i

Il faut donc trouver la fonction copule qui représente au mieux la fonction de répartition multivariée sur les quantiles historiques entre chacun des trois facteurs de risque.

C'est pourquoi les diagrammes quantiles-quantiles sont souvent utilisés pour comparer l'historique observé avec des copules théoriques calibrées sur le jeu de données. Ces diagrammes permettent d'observer la corrélation empirique, historique entre deux jeux de données.

Ces diagrammes se construisent en représentant sur chacun des deux axes les quantiles (donc compris entre 0 et 1). Chaque événement est représenté graphiquement par le quantile associé à sa première coordonnée sur l'axe des abscisses et par le quantile associé à sa deuxième coordonnée sur l'axe des ordonnées. Ainsi, il est possible d'observer graphiquement si les deux jeux de données ont tendance à évoluer conjointement dans le même sens, dans le sens opposé ou encore indépendamment l'un de l'autre.

Interprétations d'un diagramme quantiles-quantiles :

- Si la représentation du jeu de données tend vers la diagonale d'en bas à gauche vers en haut à droite du diagramme alors les deux jeux de données sont fortement corrélés positivement entre-eux.
- Si la représentation du jeu de données tend vers la diagonale d'en haut à gauche vers en bas à droite du diagramme alors les deux jeux de données sont fortement corrélés négativement entre-eux.
- Si aucune tendance linéaire ne semble apparaître alors les deux jeux de données sont indépendants ou très faiblement corrélés entre-eux.
- Les corrélations peuvent être plus ou moins marquées aux extrémités du diagramme, cela correspond à une corrélation plus ou moins forte des queues de distribution des deux jeux de données par rapport aux valeurs moyennes.

En notant q^1 , q^2 et q^3 les jeux de données des quantiles mensuels historiques sur 20 ans respectivement des jeux de données des marginales des facteurs de risque X^1 , X^2 et X^3 :

$$q^1 = \left\{ \frac{1}{240} \text{Rg}^1 \left(\ln \left(\frac{X_{m+1}^1}{X_m^1} \right) \right) \right\}_{m \in [0, 239]}$$

$$q^2 = \left\{ \frac{1}{240} \text{Rg}^2 (X_{m+1}^2 - X_m^2) \right\}_{m \in [0, 239]}$$

$$q^3 = \left\{ \frac{1}{240} \text{Rg}^3 \left(\ln \left(\frac{X_{m+1}^3}{X_m^3} \right) \right) \right\}_{m \in [0, 239]}$$

avec Rg^i la fonction de rang associée au jeu des marginales du jeu de données X^i .

Voici, ci-dessous, les diagrammes quantiles-quantiles des trois jeux de données q^1 , q^2 et q^3 deux-à-deux :

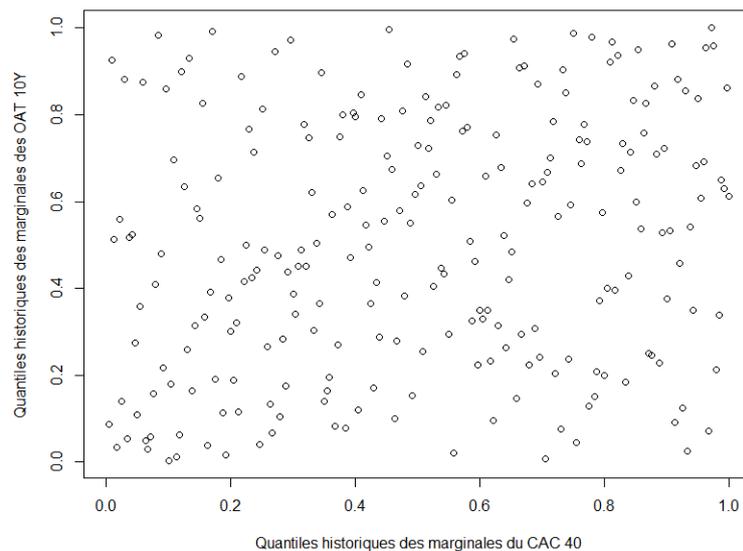


FIGURE 33 – Diagramme quantiles-quantiles entre les marginales du CAC 40 et des OAT 10Y

Pour rappel, la corrélation linéaire sur 20 ans entre les marginales du CAC 40 et des OAT 10Y est de 21%. Cette faible corrélation positive est très difficilement observable sur le diagramme. Cependant, la structure de corrélation semble symétrique et ne semble pas particulièrement s'accroître sur les queues de distribution.

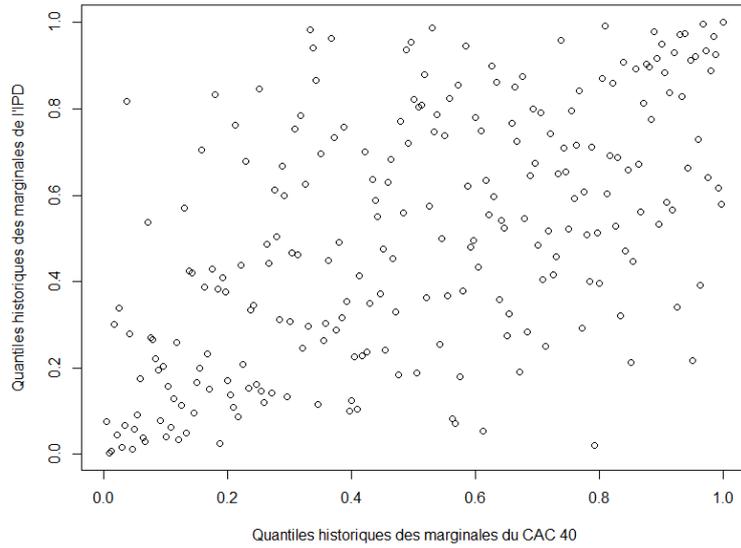


FIGURE 34 – Diagramme quantiles-quantiles entre les marginales du CAC 40 et de l'IPD

Ici, la corrélation linéaire sur 20 ans entre les marginales du CAC 40 et de l'IPD d'environ 65% est nettement plus observable. La structure de corrélation semble toujours symétrique mais cette fois-ci les queues de distribution (angle en bas à gauche et en haut à droite) semblent être d'autant plus corrélées positivement. Cela s'explique probablement par la réaction similaire à la chute puis à la hausse des deux cours lors des crises financières de 2008 et de 2020.

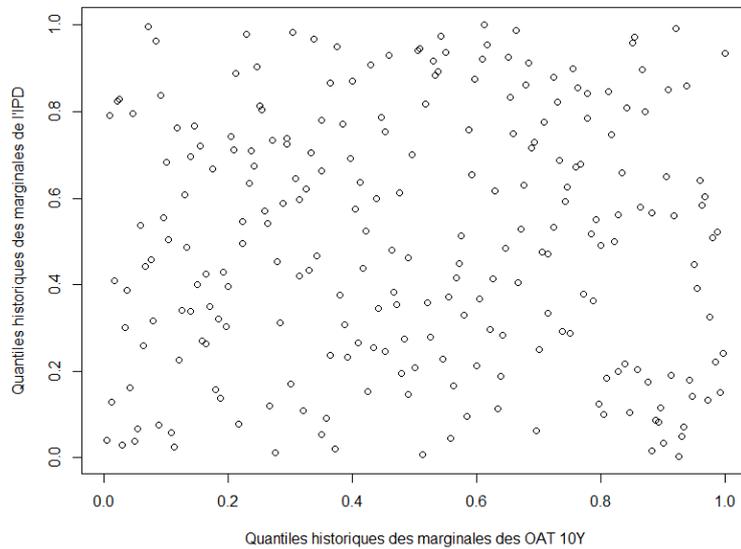


FIGURE 35 – Diagramme quantiles-quantiles entre les marginales des OAT 10Y et de l'IPD

La corrélation linéaire sur 20 ans entre les marginales des OAT 10Y et de l'IPD, d'à peine de -2% , est logiquement non visualisable et aucune déduction sur la structure de corrélation ne peut être à tirer de ce diagramme de deux jeux de données probablement indépendants.

L'absence d'asymétrie sur l'ensemble des structures bivariées de corrélation entre les queues gauches et droites invite à **rejeter l'idée des copules de Gumbel et de Clayton** qui ne semblent pas pertinentes. Ces graphiques, cependant, confortent dans la **pertinence de tester la copule de Student** par à rapport à la gaussienne au vu de la plus forte corrélation sur les queues de distribution notamment entre le CAC 40 et l'IPD.

Une approche plus quantitative avec la comparaison de ces diagrammes quantiles-quantiles historiques avec des théoriques basées sur les quatre copules calibrées sur le jeu de données sont disponibles en Annexe 2 et confirment les interprétations ci-dessus.

4 Bilan sur les copules retenues

De même que pour les lois marginales, le choix de la copule peut se révéler une hypothèse forte au vu de la corrélation plus forte aux extrêmes entre le CAC 40 et l'IPD. Reste donc à quantifier l'impact de cette hypothèse pour mesurer la robustesse du module Excel.

Deux catégories de Reverse Stress Tests (RST) sont donc retenues concernant les différents choix de copules :

- RST utilisant une **copule gaussienne**
- RST utilisant une **copule de Student**

III Alternatives à la mesure de corrélation brute

Il a été vu précédemment que modifier le coefficient de corrélation linéaire pouvait avoir des impacts sur les résultats du module Excel (cf. Partie 4-VI). Il semble donc pertinent de réfléchir aux valeurs alternatives envisageables pour ce coefficient. Ce coefficient pourrait aussi être à terme estimé à dire d'expert.

1 Mesures de corrélation alternatives

Il existe trois mesures de corrélation linéaire usuelles : le coefficient de corrélation linéaire de Pearson défini précédemment (cf. Partie 3 - II_2), le rho de Spearman et le tau de Kendall définis en annexe (cf. Annexe 5) qui sont deux mesures de rang.

Le tau de Kendall aurait été naturellement étudié si les copules de Gumbel et de Clayton avaient été retenues. Cependant, la copule gaussienne et la copule de Student sont définies avec le coefficient de corrélation linéaire de Pearson. C'est pourquoi les alternatives vont se restreindre à celui-ci.

2 Pistes de calcul du coefficient de Pearson

Comme vu précédemment, il est possible de modifier la valeur du coefficient de corrélation en fonction de l'historique de données retenu pour le calibrer. En effet, un historique de 20 ans est utilisé pour calibrer les marginales car cela permet d'avoir plus de données pour estimer les lois. Le coefficient de corrélation ne nécessite pas autant de données pour être estimé efficacement mais le choix de l'historique va orienter sur quelle portion historique il est préférable d'avoir une corrélation équivalente.

2-A *Modification de l'historique*

Il a donc été envisagé d'étudier l'évolution de la corrélation entre les différents facteurs de risque. Pour cela, il a été envisagé d'observer l'évolution de la corrélation mensuelle roulante sur 10 ans *i. e.* à partir de 2010 (jusqu'à 2020) à chaque mois est calculée la corrélation historique des 10 dernières années entre les valeurs des marginales des facteurs de risque pris deux-à-deux.

Pour chaque graphique est représenté, par une ligne horizontale en couleur, le coefficient de corrélation linéaire calculé sur les années 2001-2020 utilisé dans le module Excel.

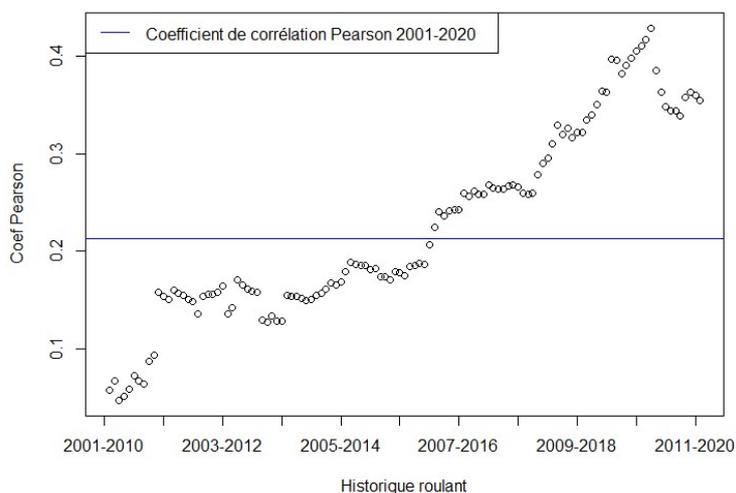


FIGURE 36 – Evolution du coefficient de corrélation linéaire roulant sur 10 ans entre les marginales du CAC 40 et des OAT 10Y

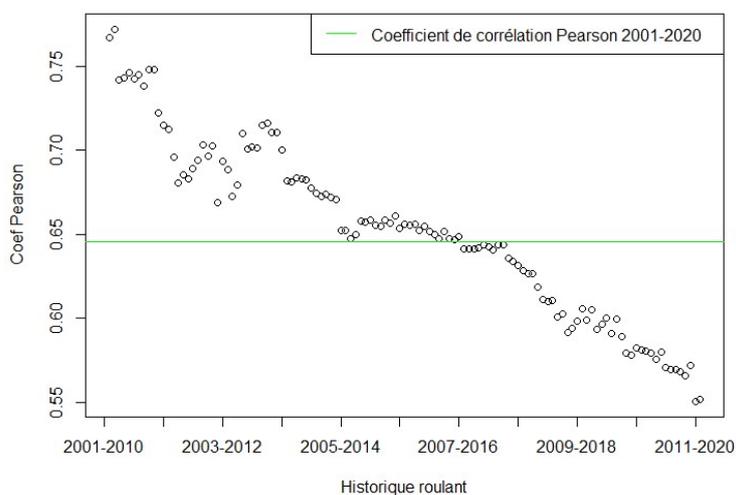


FIGURE 37 – Evolution du coefficient de corrélation linéaire roulant sur 10 ans entre les marginales du CAC 40 et de l'IPD

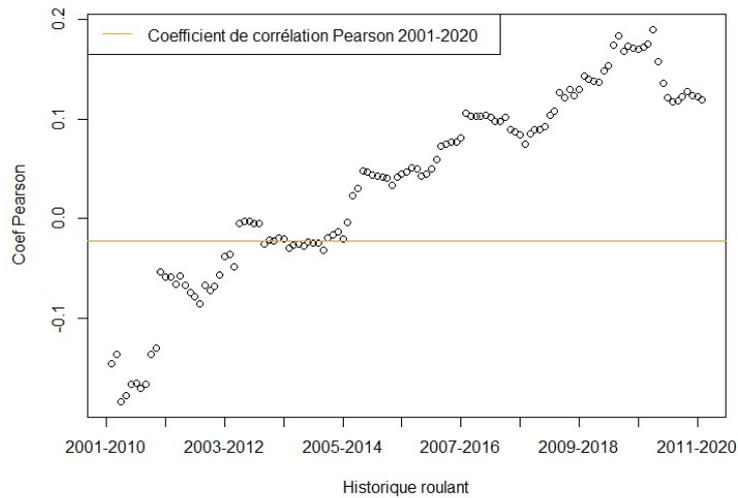


FIGURE 38 – Evolution du coefficient de corrélation linéaire roulant sur 10 ans entre les marginales des OAT 10Y et de l’IPD

Ces évolutions illustrent bien une tendance linéaire pour chacune des trois corrélations :

- Les marginales du CAC 40 et de l’OAT 10Y avaient tendance à évoluer de façon indépendante entre 2001 et 2010 pour être au final corrélées positivement ces dix dernières années.
- Les marginales du CAC 40 et de l’IPD étaient fortement corrélées mais le sont nettement de moins en moins au fil du temps.
- Les marginales de l’OAT 10Y et de l’IPD ont vu le sens de leur corrélation évoluer passant d’une faible corrélation négative à une faible corrélation positive.

L’étude de ces tendances de corrélations appuie l’idée qu’une estimation à dire d’expert pourrait être pertinente *in fine*.

Les corrélations calculées sur 20 ans (lignes horizontales des graphiques) sont une moyenne sur l’ensemble de cette évolution. Il peut sembler **pertinent d’envisager l’alternative de la corrélation sur les 10 dernières années (2011-2020)**. Cet historique demeure conséquent pour être relativement stable sur le temps court et il permet bien de conserver un crise financière (celle de mars 2020) affectant notamment les corrélations sur les queues de distribution extrêmes des marginales.

2-B Modification de la pondération

Une autre méthode consisterait à **conserver la profondeur d'historique de 20 ans tout en sur-pondérant les années récentes et en sous-pondérant les années plus anciennes.**

Voici donc la formule du coefficient de corrélation pondéré r' (un parmi d'autres envisageables) à utiliser sur les 20 années d'historique des marginales :

Soit X et Y deux jeux de données de taille n de valeurs respectives x_1, \dots, x_n et y_1, \dots, y_n .

$$r'(X, Y) = \frac{\sum_{i=1}^n p_i (x_i - \mu'_X)(y_i - \mu'_Y)}{\sqrt{\sum_{i=1}^n p_i (x_i - \mu'_X)^2 \times \sum_{i=1}^n p_i (y_i - \mu'_Y)^2}}$$

et d'espérances pondérées respectives :

$$\mu'_X = \sum_{i=1}^n p_i x_i$$

$$\mu'_Y = \sum_{i=1}^n p_i y_i$$

avec la pondération p_i devant vérifier :

$$\sum_{i=1}^n p_i = 1$$

Beaucoup de pondérations sont donc envisageables mais la pondération retenue dans le cadre ce mémoire est celle d'une suite arithmétique de raison 1 (mais repondérée pour valider la contrainte ci-dessus soit de raison $\frac{2}{n(n+1)}$) :

$$p_i = \frac{2i}{n(n+1)}$$

3 Bilan sur les coefficients de corrélation retenus

A l'instar du choix de la copule et des lois marginales, plusieurs choix ont été fixés afin de quantifier la performance du module Excel en le *challengeant* avec d'autres Reverse Stress Tests.

Trois catégories de Reverse Stress Tests (RST) sont donc retenues concernant les différents choix de calcul des coefficients de corrélation :

- RST calculant le **coefficient de corrélation de Pearson brut sur 20 ans**
- RST calculant le **coefficient de corrélation de Pearson brut sur 10 ans**
- RST calculant le **coefficient de corrélation de Pearson pondéré sur 20 ans**

Voici les valeurs de ces différentes catégories de coefficients pour chaque corrélation :

Corrélation de Pearson des marginales		Facteurs de risque 2 à 2		
		CAC 40 / OAT 10Y	CAC 40 / IPD	OAT 10Y / IPD
Historique	Calcul			
20 ans	Brut	21,32%	64,54%	-2,26%
10 ans	Brut	35,42%	55,16%	11,97%
20 ans	Pondéré	30,59%	55,63%	5,03%

FIGURE 39 – Valeurs des corrélations en fonction de l'historique et de la méthode de calcul retenus

Le coefficient brut sur 20 ans est le coefficient qui donne le plus d'importance aux années anciennes, *a contrario* de celui brut sur 10 ans qui correspond aux corrélations les plus récentes.

IV Synthèse des modèles de Reverse Stress Test

Les différentes alternatives mathématiques, en les croisant, permettent d'obtenir **18 modèles de Reverse Stress Test** à implémenter puis à comparer.

	Hypothèses mathématiques			Faisable sur Excel ?
	Lois des marginales	Copule	Corrélation de Pearson	
Modèle A	Normales	Gaussienne	Sur 20 ans sans pondération	Oui
Modèle B	Normales	Gaussienne	Sur 10 ans sans pondération	Oui
Modèle C	Normales	Gaussienne	Sur 20 ans avec pondération	Oui
Modèle D	Personnalisées	Gaussienne	Sur 20 ans sans pondération	Non
Modèle E	Personnalisées	Gaussienne	Sur 10 ans sans pondération	Non
Modèle F	Personnalisées	Gaussienne	Sur 20 ans avec pondération	Non
Modèle G	Estimateur à noyau gaussien	Gaussienne	Sur 20 ans sans pondération	Non
Modèle H	Estimateur à noyau gaussien	Gaussienne	Sur 10 ans sans pondération	Non
Modèle I	Estimateur à noyau gaussien	Gaussienne	Sur 20 ans avec pondération	Non
Modèle J	Normales	Student	Sur 20 ans sans pondération	Non
Modèle K	Normales	Student	Sur 10 ans sans pondération	Non
Modèle L	Normales	Student	Sur 20 ans avec pondération	Non
Modèle M	Personnalisées	Student	Sur 20 ans sans pondération	Non
Modèle N	Personnalisées	Student	Sur 10 ans sans pondération	Non
Modèle O	Personnalisées	Student	Sur 20 ans avec pondération	Non
Modèle P	Estimateur à noyau gaussien	Student	Sur 20 ans sans pondération	Non
Modèle Q	Estimateur à noyau gaussien	Student	Sur 10 ans sans pondération	Non
Modèle R	Estimateur à noyau gaussien	Student	Sur 20 ans avec pondération	Non

FIGURE 40 – Caractéristiques des différents Reverse Stress Tests

Le modèle A correspond donc au premier module de Reverse Stress Tests implémenté sur Excel. **Seuls les modèles A, B et C sont implémentables sur Excel.** Il faut donc créer un nouveau module de Reverse Stress Tests permettant d'implémenter l'ensemble des 18 modèles pour les comparer toutes choses étant égales par ailleurs.

Reverse Stress Test via R

Afin de comparer les différents modèles de Reverse Stress Tests il a été retenu d'implémenter une nouvelle méthode de Reverse Stress Tests sur R. Ce logiciel de programmation permet de faire l'ensemble de ce qui a été fait sur Excel. Cependant, les hypothèses mathématiques levées impliquent que le Reverse Stress Test ne se résume plus à l'optimisation sous contraintes d'une fonction multivariée des densités des marginales.

Une nouvelle méthode permettant d'obtenir à partir des mêmes *inputs* (marginales des facteurs de risque, valeurs récentes des facteurs de risque, sensibilités de la CDC à ces facteurs de risque, impact cible) le même type d'*output* (l'évènement le plus probable à date engendrant cet impact) a donc été élaborée.

I Méthodologie par tirage

La méthode la plus intuitive et la plus appropriée à l'utilisation de copules est de réaliser un Reverse Stress Test grâce à un tirage. Pour chaque modèle, et ses hypothèses correspondantes, les marginales sont corrélées dans le tirage afin d'obtenir une liste de triplets de marginales qui permettent ensuite d'en déduire les valeurs des facteurs de risque et les impacts sur la CDC en fonction des sensibilités introduites.

1 Optimisation en temps

La réalisation d'un Reverse Stress Test par tirage engendre une complexité en temps. En effet, plus le nombre d'évènements tirés est conséquent plus le tirage est long à réaliser mais plus le nombre d'évènements est faible moins le modèle est robuste et stable si le tirage est reproduit à nouveau.

Pour contourner cet effet, une possibilité est de ne pas faire un tirage à 1 mois comme les marginales mensuelles l'engendrent mais un **tirage à 1 an** par exemple en cumulant 12 marginales. Cela permet de faire apparaître des évènements stressés plus rapidement et donc de diminuer le nombre d'évènements nécessaires en fonction de l'impact cible (plus l'impact est élevé en valeur absolue plus les évènements sont rares et plus le tirage a besoin d'un nombre conséquent d'évènements tirés pour les voir survenir).

Le tirage comporte donc 100 000 simulations à horizon 1 an, ce nombre permet d'apporter une stabilité aux différents modèles pour les comparer au mieux avec un bruit aléatoire moindre.

2 Suppression des tendances

Avec le module Excel, les jeux de marginales n'ont pas à être altérés pour ne pas prendre en compte le non-centrage des marginales (tendances des facteurs de risque non nulles) puisque l'espérance est directement un paramètre de la loi marginale normale qu'il suffit de forcer à 0. *A contrario* sur R, comme ce paramètre d'espérance n'est pas présent sur l'ensemble des autres lois marginales, il est donc nécessaire de centrer les jeux de marginales avant de calibrer ces lois.

3 Prise en compte des *drifts*

De plus, le modèle par tirage implique un nouveau biais par rapport au module Excel. En effet, le tirage de marginales en log-rendement induit aussi un *drift* (appelé *drift* d'Itô).

Espérance des log-rendements

Pour X_t une série temporelle d'instant t ,

$$E\left(e^{\ln \frac{X_{t+1}}{X_t}}\right) \neq e^{E\left(\ln \frac{X_{t+1}}{X_t}\right)}$$

Il est donc nécessaire d'altérer les paramètres de calibrage des marginales log-rendements pour compenser ces *drifts*. Ces derniers diffèrent selon les lois théoriques retenues. Le *drift* de la loi normale marginale est relativement connu car il est présent dans la formule de *Black and Scholes* et il explique la moyenne théorique d'une distribution log-normale.

Ainsi, le *drift* d'Itô d'une variable aléatoire X distribuée selon une loi normale $\mathcal{N}(\mu, \sigma^2)$ vaut :

$$\text{Drift}_{\text{Itô}}^{\text{Norm}} = \mu + \frac{\sigma^2}{2}$$

Le *drift* d'Itô d'une variable aléatoire X distribuée selon une loi normale asymétrique $\text{Skew}(\epsilon, \omega, \alpha)$ vaut :

$$\text{Drift}_{\text{Itô}}^{\text{Skew}} = \ln\left(2e^{\epsilon+0.5\omega^2} \times \Phi\left(\frac{\omega\alpha}{\sqrt{1+\alpha^2}}\right)\right)$$

avec Φ la fonction de répartition d'une loi normale centrée réduite.

Le *drift* d'Itô d'une variable aléatoire X distribuée selon une loi normale-inverse gaussienne $\text{NIG}(\alpha, \beta, \gamma, \epsilon)$ vaut :

$$\text{Drift}_{\text{Itô}}^{\text{NIG}} = \mu + \delta \left(\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + 1)^2} \right)$$

Le *drift* d'Itô d'une variable aléatoire X distribuée selon un estimateur à noyau gaussien (x_1, \dots, x_n, h, n) vaut :

$$\text{Drift}_{\text{Itô}}^{\text{Kernel}} = \ln\left(\frac{1}{n} \sum_{i=1}^n e^{x_i+0.5h^2}\right)$$

Ces trois derniers *drifts* ne possèdent pas de formule fermée et ont été estimés par DFIN par simulation de Monte-Carlo.

4 Résumé des étapes du tirage du Reverse Stress Test via R

L'ensemble des valeurs des paramètres ayant permis le Reverse Stress Test sur R est présenté en Annexe 3. Les *packages* R ayant permis de les estimer sont disponibles en Annexe 1.

1. Importation des *inputs* :
 - Valeurs récentes des facteurs de risque
 - Historique des facteurs de risque entre 2001 et 2020
 - Sensibilités de la CDC à la baisse ou à la hausse des facteurs de risque

2. Transformation des cours en jeu de données de marginales
 - (a) Calcul des marginales selon le type de facteurs de risque
 - (b) Retrait des *drifts* pour les marginales en log-rendements, retrait des espérances pour les autres marginales

3. Calibrage des hypothèses mathématiques
 - (a) Calibrage des lois marginales (normales, personnalisées et par estimateurs gaussiens)
 - (b) Calibrage des coefficients de corrélation de Pearson (sur 20 ans sans pondération, sur 10 ans sans pondération et sur 20 ans avec pondération)
 - (c) Calibrage des copules (gaussienne et de Student) selon le coefficient de corrélation retenu

4. Tirage de 100 000 simulations à horizon 1 an
 - (a) Tirage de 12 quantiles mensuels grâce aux copules pour chaque modèle
 - (b) Transformation de ces quantiles en marginales grâce aux fonctions de répartition inverses
 - (c) Transformation de 12 marginales mensuelles en 1 valeur marginale annuelle grâce à leur propriété d'additivité (cf. Partie 3 - I_3)
 - (d) Transformation de ces marginales à 1 an en valeur de cours projeté
 - (e) Calcul des impacts sur la CDC pour chaque triplet d'évènement

II Sélection des *outputs* adéquats

Le tirage pour chaque modèle renvoie donc 100 000 évènements de quadruplets : les valeurs simulées du CAC 40, des OAT 10Y et de l'IPD dans 1 an ainsi que l'impact sur la CDC.

La première idée intuitive est de retourner en sortie du tirage, pour un impact cible, l'évènement le plus fréquemment tiré. Le problème est qu'aucun impact n'est certain d'être tiré exactement à l'euro près. Il est donc nécessaire de réaliser un arrondi sur les impacts pour regrouper les évènements ayant des impacts similaires. L'arrondi retenu par DFIN est de l'ordre du milliard d'euros.

Il est donc ainsi bien possible de retourner l'évènement le plus fréquemment tiré d'un modèle, pour un impact cible. Cependant, bien que le nombre d'évènements tirés soit conséquent, il demeure une certaine volatilité de l'évènement le plus fréquent pour un impact similaire.

Il est important d'avoir un modèle de Reverse Stress Test stable dans le temps. Donc pour tous les tirages de chaque modèle testé, la sortie du modèle qui a été retenue, pour un impact cible (noté I), est de **renvoyer en sortie les facteurs de risque moyens au sein des évènements ayant cet impact cible : un CAC 40 moyen (noté \bar{X}^1), un OAT 10Y moyen (noté \bar{X}^2) et un IPD moyen noté (\bar{X}^3)**. Ce sont ces facteurs de risque moyens qui seront appelés par la suite *outputs* du Reverse Stress Test.

Soit X_i^j la valeur à 1 an du $j^{\text{ème}}$ facteur de risque dans l'évènement n°i.

$\forall j \in \llbracket 1, 3 \rrbracket$ et $\forall N \in \mathbb{N} \ I = N \times 10^9$:

$$\bar{X}_I^j = E \left(X_i^j \mid \text{Impact} = I \right) = \frac{\sum_{i=1}^{100000} X_i^j \times \mathbb{1}_{\text{Impact}(i) = I}}{\sum_{i=1}^{100000} \mathbb{1}_{\text{Impact}(i) = I}}$$

avec $\text{Impact} (\llbracket 1, 100000 \rrbracket \rightarrow \mathbb{R})$ la fonction déterminant l'impact de l'évènement d'un tirage à partir de son numéro.

Par exemple sur un tirage de 100 000 scénarios projetés à 1 an, si l'impact cible est fixé à -3 Md € alors les scénarios possibles sont filtrés et ne sont retenus que ceux correspondants à l'impact (par exemple 5 000 scénarios). Et c'est sur ces scénarios que sont déterminés les *outputs* du Reverse Stress Test en calculant la moyenne de chaque facteur de risque.

Cette méthodologie par tirage permet de simplifier grandement la question autour du sens donné, pour la gouvernance, à l'évènement en sortie du Reverse Stress Test. En effet, l'idée d'optimisation sous contraintes d'une fonction de densité multivariée (dans le module Excel) est plus complexe à vulgariser qu'un évènement moyen dans un tirage.

Comparaison des modèles implémentés sur R

Pour l'ensemble des 18 modèles de Reverse Stress Tests (cf. Partie 5 - IV) implémentés sur R, la méthodologie permet bien d'obtenir, pour les mêmes *inputs* du module Excel, des *outputs* de même nature. Il reste maintenant à comparer ces différents *outputs* pour juger si certains peuvent être considérés significativement différents du modèle A, B ou C (implémentables via Excel).

Il est donc nécessaire de réfléchir à des méthodes de comparaison de ces modèles.

Tout d'abord, il peut être intéressant de comparer les modèles à différents impacts cibles, des impacts liés à des événements plus ou moins stressés. Les tests de résistance ayant pour objectif de tester la résilience de la CDC à des événements adverses, seuls des impacts cibles négatifs sont retenus.

Ainsi, **quatre impacts** en pertes de valeur du portefeuille de la CDC sont retenus par DFIN :

- - 10 Md €
- - 7 Md €
- - 5 Md €
- - 3 Md €

Pour chaque impact il est donc possible de comparer le CAC 40 moyen, l'OAT 10Y moyen et l'IPD moyen en sortie (*i. e.* 1 valeur moyenne par facteur de risque et par impact retenu soit **12 outputs**). Comme ces facteurs de risque sont des moyennes sur les années simulées, pour un impact cible, il peut aussi être intéressant d'observer les écarts-types de ces facteurs de risque (*i. e.* 1 écart-type par facteur de risque et par impact retenu soit **12 variables caractéristiques des modèles**). En effet, deux modèles peuvent très bien sortir un CAC moyen identique mais dont les écarts-types au sein du tirage (filtré sur un impact cible) sont sensiblement différents.

De plus, **1 variable quantitative** correspondant à la part des années simulées dans le tirage ayant eu des impacts négatifs sur la CDC plus conséquents que 15 Md € peut aussi être intéressante. Cette variable caractéristique du modèle peut permettre de déterminer la faisabilité du modèle sur un impact cible plus élevé que les quatre sélectionnés.

Chaque modèle peut donc être comparé à un autre au sein de **25 variables quantitatives** (*i.e.* 12 *outputs* et 13 variables caractéristiques du modèle) **qui vont permettre de comparer les 18 modèles** afin de déterminer les modèles proches et en déduire les hypothèses mathématiques les plus fortes parmi le choix des marginales, le choix de la copule et le choix de la méthode de calcul du coefficient de corrélation de Pearson. Pour représenter graphiquement ce jeu de données à 25 dimensions, il est nécessaire de synthétiser l'ensemble des variables au sein de nouvelles variables appelées composantes principales.

I Analyse en composantes principales

L'analyse en composantes principales (notée ACP) permet de transposer les variables d'un jeu de données en axes principaux orthogonaux les uns aux autres. Ces axes sont construits par combinaisons linéaires des variables initiales. Chaque axe représente ainsi un pourcentage d'explication de l'ensemble des variables. L'orthogonalité des axes permet d'obtenir une explication totale, de l'ensemble des axes, de 100%.

Ces axes permettent ainsi de représenter les individus du jeu de données en deux dimensions avec les deux principaux axes afin de simplifier au mieux le jeu de données multidimensionnel. De plus, les variables initiales peuvent aussi être intégrées dans la visualisation en deux dimensions à l'aide de leurs corrélations aux axes représentées par des angles de cosinus égaux à ces corrélations. Ainsi, la représentation d'une variable (sous forme d'une flèche) se rapprochant de l'horizontal (resp. la verticale) signifie que la variable est quasiment exclusivement contenue dans l'axe horizontal (resp. vertical) et donc quasiment indépendante de l'axe vertical (resp. horizontal).

1 Théorie

Soit une base de données composée de n individus et p variables représentée sous la forme d'une matrice :

$$\tilde{X} = (\tilde{X}_{i,j})_{\substack{i \in \llbracket 1, n \rrbracket \\ j \in \llbracket 1, p \rrbracket}}$$

La construction de l'analyse en composantes principales peut se résumer par les étapes suivantes :

1. Centrage et réduction des données :

$$X = (X_{i,j})_{\substack{i \in \llbracket 1, n \rrbracket \\ j \in \llbracket 1, p \rrbracket}} = \left(\frac{\tilde{X}_{i,j} - \mu_j}{\sigma_j} \right)_{\substack{i \in \llbracket 1, n \rrbracket \\ j \in \llbracket 1, p \rrbracket}}$$

avec $\forall j \in \llbracket 1, p \rrbracket$

$$\mu_j = \frac{1}{n} \sum_{i=1}^n \tilde{X}_{i,j}$$

$$\sigma_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\tilde{X}_{i,j} - \mu_j)^2}$$

2. Détermination de l'axe principal $u_1^1 X_1 + \dots + u_p^1 X_p$:

$$\arg \min_{u_1^1, \dots, u_p^1 \in \mathbb{R}} = \sum_{i=1}^n \sum_{j=1}^p (X_{i,j} - u_1^1 X_{i,1} - \dots - u_p^1 X_{i,p})^2$$

3. Détermination des $p - 1$ axes suivants de la même manière avec la condition supplémentaire que chaque nouvel axe k doit être orthogonal à ceux déjà déterminés :

$$\forall k \in \llbracket 2, p \rrbracket, A_k = u_1^k X_1 + \dots + u_p^k X_p$$

$$\forall l < k, A_k \perp A_l$$

Les axes sont aussi appelés des vecteurs propres.

4. Représentation des individus dans le nouvel espace E :

$$\forall i \in \llbracket 1, n \rrbracket, X_i = \left(\sum_{j=1}^p u_j^1 X_{i,j} \quad \dots \quad \sum_{j=1}^p u_j^p X_{i,j} \right)_E$$

5. Représentation des variables dans le nouvel espace E :

$$\forall j \in \llbracket 1, p \rrbracket, V_j = \begin{pmatrix} \cos(V_j, u_1^1 X_1 + \dots + u_p^1 X_p) \\ \vdots \\ \cos(V_j, u_1^p X_1 + \dots + u_p^p X_p) \end{pmatrix}_E = \begin{pmatrix} \frac{\sum_{k=1}^p u_k^1 \text{Cov}(V_j, X_k)}{\sqrt{\sum_{k=1}^p (u_k^1)^2}} \\ \dots \\ \frac{\sum_{k=1}^p u_k^p \text{Cov}(V_j, X_k)}{\sqrt{\sum_{k=1}^p (u_k^p)^2}} \end{pmatrix}_E$$

6. Calcul des valeurs propres des axes :

$$\forall k \in \llbracket 1, p \rrbracket, \lambda_k = \text{Var}(u_1^k X_1 + \dots + u_p^k X_p) = \sum_{l=1}^p (u_l^k)^2$$

7. Calcul des scores d'explicativités des axes :

$$\forall k \in \llbracket 1, p \rrbracket, S_k = \frac{\lambda_k}{\sum_{l=1}^p \lambda_l}$$

L'objectif était ici d'avoir une approche simplifiée de la construction théorique de l'analyse en composantes principales. Mais des définitions plus complètes, utilisant notamment les notions d'inertie, existent et sont détaillées dans les mémoires d'actuariat de Lucas Grandperrin (2019, cf. [7]) et de Thi Diep Trieu (2013, cf. [11]).

L'analyse en composantes principales va donc permettre de représenter l'ensemble des modèles en deux dimensions (en fonction des axes avec les scores les plus explicatifs) en les différenciant les uns des autres graphiquement par leurs coefficients sur chacun des deux axes. Les variables affichées, quant à elles, pourront permettre de comprendre la contribution des variables dans les axes principaux.

2 Analyse en composantes principales sur les 25 variables représentatives des modèles

2-A Détermination du nombre d'axes à observer

Tout d'abord, la représentation graphique peut se réaliser sur deux axes de l'ACP quels qu'ils soient. Il est donc important de connaître la contribution de chaque axe dans la variance totale du jeu de données afin de savoir si la représentation graphique sur les deux premiers axes est suffisante ou si d'autres graphiques avec introduction des troisième ou quatrième axes peuvent aussi être intéressants et contenir de l'information.

Pour déterminer le nombre d'axes à déterminer il est possible d'utiliser la **technique du coude**. Celle-ci a pour objectif de se concentrer sur l'éboulis (*i. e.* la décroissance) du score d'explicativité (ou de la valeur propre) des axes classés en ordre décroissant d'explicativité. La méthode consiste à ne retenir que les axes dont les scores d'explicativité sont significativement supérieurs à l'axe suivant.

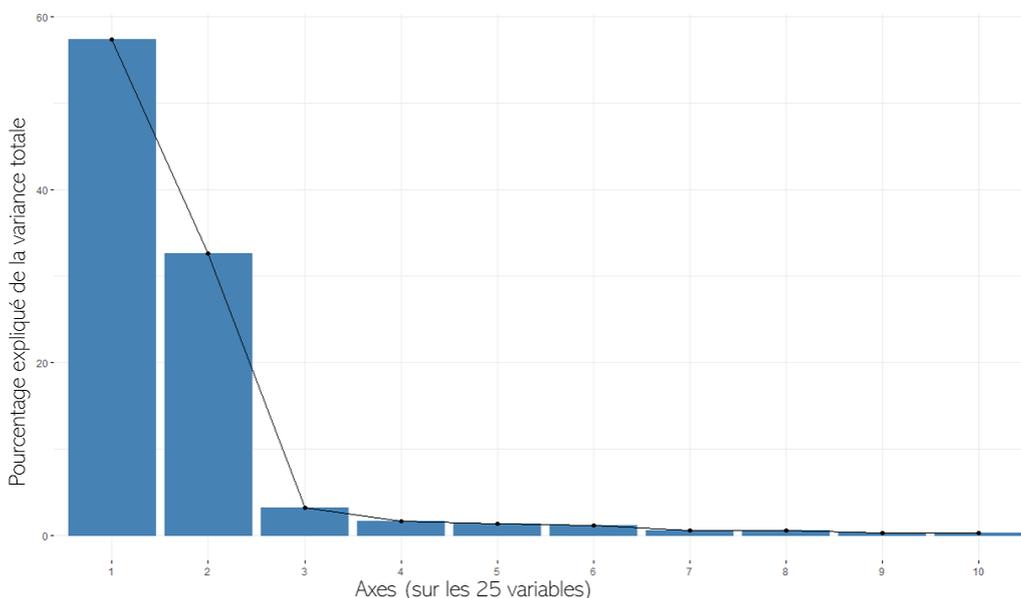


FIGURE 41 – Scores d'explicativité des axes de l'ACP sur les 25 variables

C'est le troisième axe le plus explicatif qui figure au niveau du coude des scores d'explicativité. La méthode du coude indique donc que les deux axes principaux sont suffisants pour représenter l'ensemble du jeu de données. De plus, la représentation en deux dimensions semble nettement suffisante avec l'axe principal expliquant 57,4% de la variance totale et l'axe secondaire expliquant 32,6%. La représentation graphique sur les deux premiers axes représente donc 90%. La représentation graphique des autres axes ne semble donc pas nécessaire.

2-B Représentation des variables

La représentation des variables selon les deux axes sur $[-1, 1]^2$ se réalise directement grâce à leurs coordonnées (cosinus avec les axes).

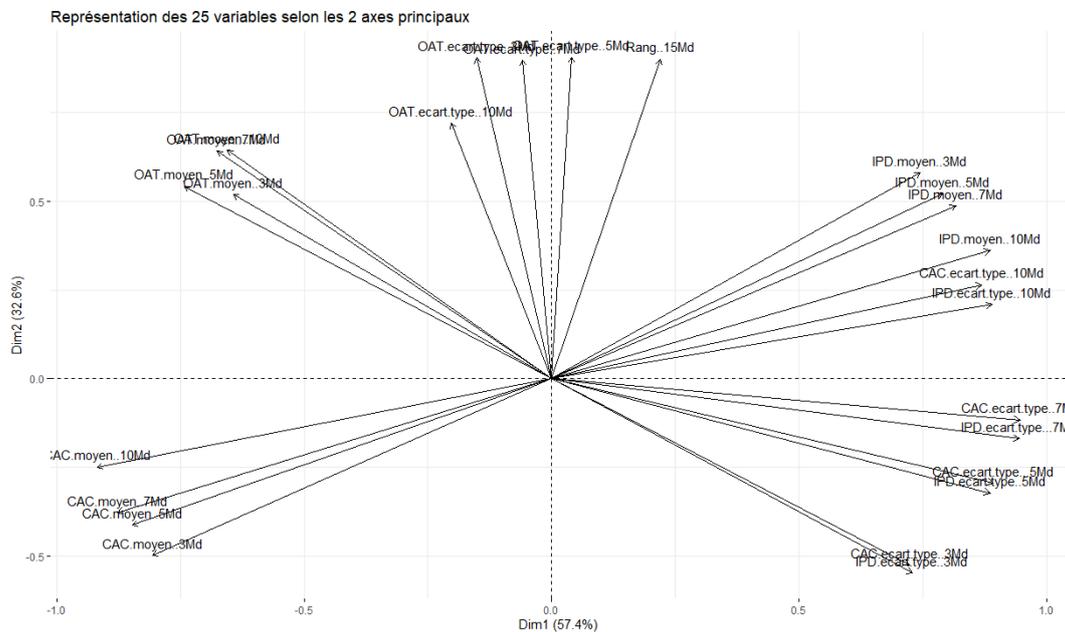


FIGURE 42 – Représentation de la contribution des 25 variables sur les deux axes principaux

L'information des variables des écarts-types des facteurs de risque CAC et IPD pour les impacts - 10 Md € - 7 Md € et - 5 Md € semble être principalement contenue dans l'axe principal (horizontal). *A contratio* l'information liée aux variables des écarts-types des OAT 10Y pour les quatre impacts et la variable de rang en dessous de - 15 Md € semble plutôt contenue dans l'axe secondaire (vertical). L'ensemble des variables *outputs* du modèle contribue de façon plus répartie entre les deux axes.

Chaque variable est représentée par une flèche, chaque flèche indique la direction et le sens selon laquelle la valeur de la variable croît. Les variables partent toutes de l'origine du graphique, un individu représenté à l'origine du graphique aurait donc des valeurs de variables correspondant respectivement aux moyennes des variables.

2-C Représentation des modèles

La représentation des individus selon les deux axes (combinaisons linéaires des variables initiales) sur \mathcal{R}^2 se réalise directement grâce à leurs coordonnées respectives sur ces deux combinaisons linéaires.

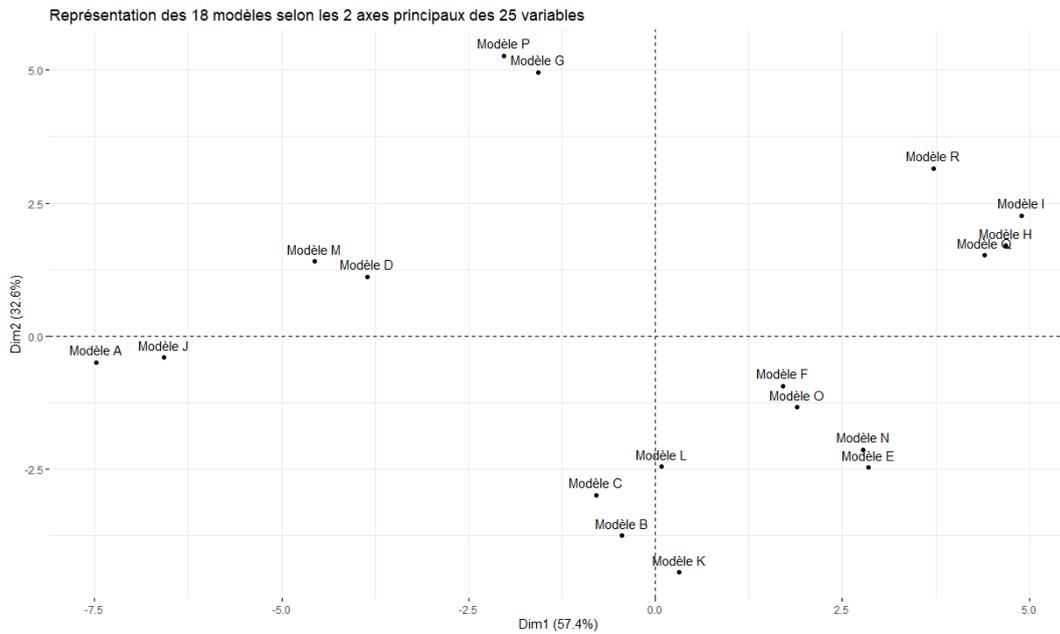


FIGURE 43 – Représentation des 18 modèles sur les deux axes principaux

La représentation souligne que l'information détenue par les variables des modèles A et J est essentiellement contenue dans l'axe principal (horizontal) et celle des modèles B, C, K et L dans l'axe secondaire.

2-D Représentation des 25 variables et des 18 modèles

Les variables peuvent être ajoutées à cette représentation, avec une transposition des coefficients des variables à l'échelle de ceux des individus.

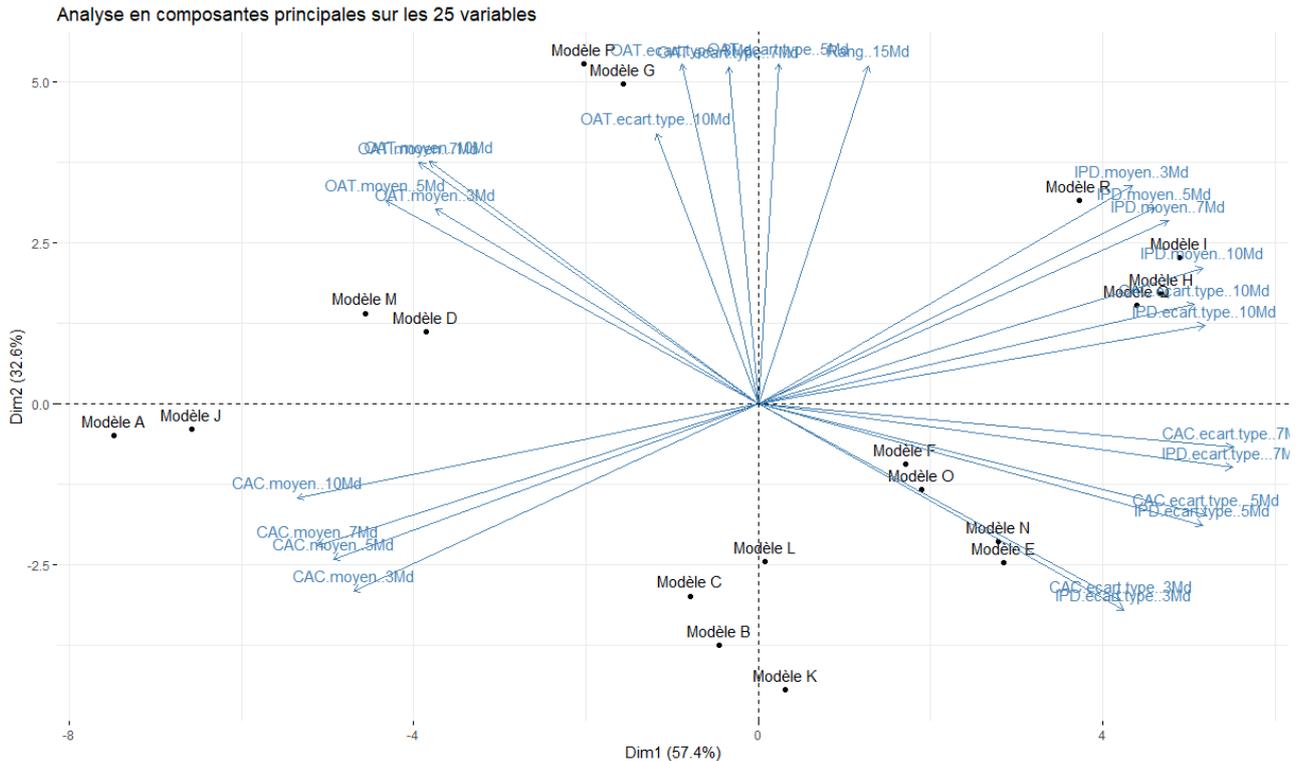


FIGURE 44 – Représentation de l'ACP sur 25 variables sur ses deux axes principaux

La représentation conjointe des modèles et de leurs variables permet d'établir un relation d'ordre au sein de chacune des 25 variables entre les 18 modèles. Pour cela, il suffit de projeter graphiquement chaque modèle représenté par un point du graphique sur les droites de direction des flèches de chaque variable.

- Un modèle projeté près de l'extrémité de la flèche d'une variable a une valeur haute par rapport à la moyenne de cette variable.
- Un modèle projeté près de l'origine a une valeur proche de la moyenne de cette variable.
- Un modèle projeté près de la symétrie de l'extrémité de la flèche d'une variable par rapport à l'origine a une valeur faible par rapport à la moyenne de cette variable.

Il est donc possible de tirer des premières conclusions sur l'impact du choix des hypothèses mathématiques sur les 25 variables (toutes choses étant égales par ailleurs) :

1. Lois marginales

(a) Lois normales (Modèles A, B, C, J, K et L)

Les *outputs* du CAC moyen sont plus hauts que la moyenne, les *outputs* de l'IPD sont plus faibles que la moyenne, les écarts-types des OAT 10Y sont plus faibles que la moyenne et la proportion d'évènements tirés d'impacts plus conséquents que - 15 Md € est en dessous de la moyenne.

(b) Lois personnalisées (Modèles D, E, F, M, N et O)

Les *outputs* du CAC et de l'IPD sont répartis autour de la moyenne des modèles. Idem pour les écarts-types des OAT 10Y et la variable de rang en dessous de - 15 Md €.

(c) Estimateurs gaussiens (Modèles G, H, I, P, Q et R)

Les *outputs* du CAC moyen sont plus faibles que la moyenne, les *outputs* de l'IPD sont plus hauts que la moyenne, les écarts-types des OAT 10Y sont plus hauts que la moyenne et la proportion d'évènements tirés d'impacts plus conséquents que - 15 Md € est au dessus de la moyenne.

2. Copule

(a) Copule gaussienne (Modèles A, B, C, D, E, F, G, H et I)

Aucun impact décelable à partir des analyses de données.

(b) Copule de Student (Modèles J, K, L M, N, O, P, Q et R)

Aucun impact décelable à partir des analyses de données.

3. Coefficient de corrélation

(a) Brut sur 20 ans (Modèles A, D, G, J, M et P)

Les *outputs* des OAT 10Y sont plus hauts que la moyenne et les variables d'écart-type d'impact - 3 Md €, - 5 Md € et - 7 Md € du CAC 40 et de l'IPD sont plus faibles que la moyenne.

(b) Brut sur 10 ans (Modèles B, E, H, K, N et Q)

Les *outputs* des OAT 10Y sont plus faibles que la moyenne et les variables d'écart-type d'impact - 3 Md €, - 5 Md € et - 7 Md € du CAC 40 et de l'IPD sont plus hautes que la moyenne.

(c) Pondéré sur 20 ans (Modèles C, F, I, L, O et R)

Idem que pour le coefficient de corrélation linéaire brut sur 10 ans.

L'analyse en composantes principales souligne donc une opposition d'une part entre les modèles de lois marginales normales et ceux de lois marginales par estimateur gaussien et d'autre part entre les modèles utilisant un coefficient de corrélation sans pondération sur 20 ans et ceux utilisant les autres méthodes de calcul du coefficient de corrélation.

L'analyse en composantes principales, par sa construction, étudie sur un même plan l'ensemble des variables. Or, dans ces 25 variables, seuls les *outputs* du modèle seront *in fine* transmis à la gouvernance. La réalisation de l'analyse en composantes principales sur les 12 *outputs* peut donc permettre de représenter et d'étudier les modèles uniquement sur l'utilisation qui sera faite des Reverse Stress Tests. Pour rappel, ces *outputs* correspondent aux 12 variables des trois facteurs de risque moyens pour chacun des quatre impacts retenus.

3 Analyse en composantes principales sur le jeu des 12 variables *outputs* des modèles

3-A Détermination du nombre d'axes à observer

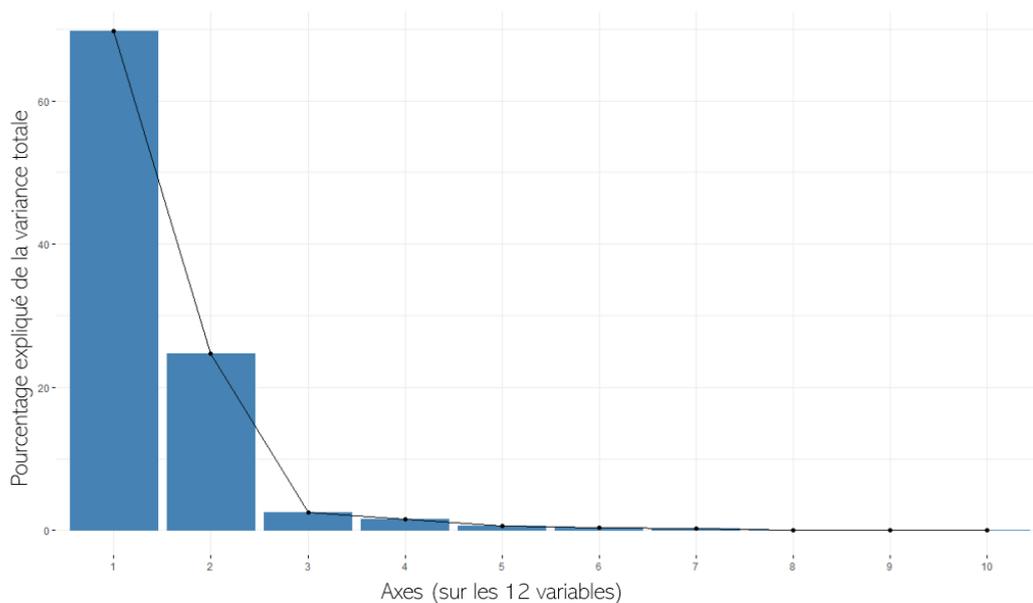


FIGURE 45 – Scores d'explicativité des axes de l'ACP sur les 12 variables

De même que précédemment, la méthode du coude souligne que les deux axes principaux sont suffisants. En se restreignant aux 12 *outputs*, l'explicativité du jeu de données se concentre sur l'axe principal expliquant quasiment 70% de la variance des 12 variables. La représentation sur les deux axes demeure de nouveau largement suffisante avec une représentation de quasiment 95% de la variance du jeu de 12 variables avec une contribution d'environ 25% de l'axe secondaire.

3-B Représentation des 12 variables et des 18 modèles

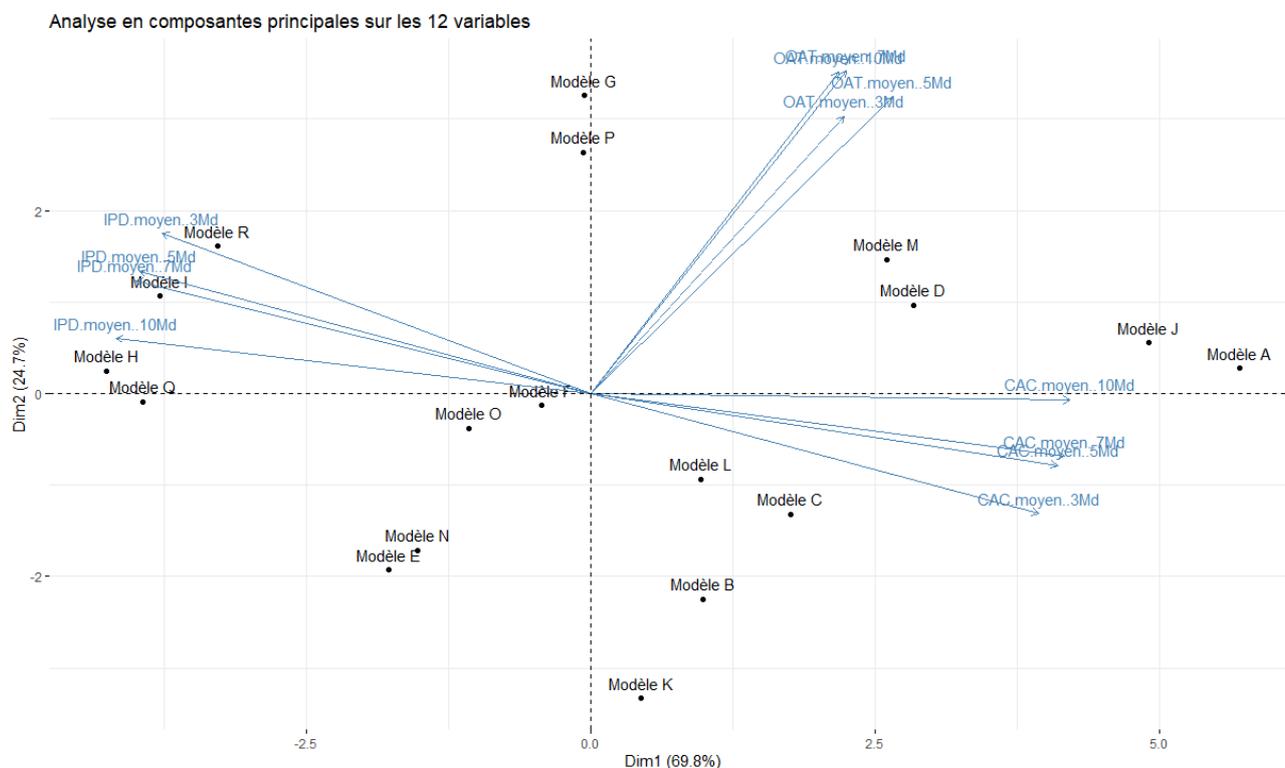


FIGURE 46 – Représentation de l'ACP sur 12 variables sur ses deux axes principaux

L'axe principal semble donc expliquer la balance entre le CAC 40 et l'IPD dans l'impact final comme observé dans le module Excel (cf. Partie 4-VII). Celui-ci soulignait, tout comme les valeurs des corrélations, que les OAT 10Y semblent indépendantes des deux autres facteurs de risque et ne contribuer que très marginalement à l'impact. L'impact des OAT 10Y dans les résultats du Reverse Stress Test est donc probablement très mesuré.

L'axe principal résume à la fois l'importance de l'hypothèse de marginale et du coefficient de corrélation. En effet, l'axe oppose les modèles A et J (lois marginales normales, coefficient de corrélation brut sur 20 ans) avec les modèles H, I, Q et R (lois marginales par estimateur gaussien, coefficient de corrélation brut sur 10 ans ou pondéré sur 20 ans). Aucune hiérarchisation d'importance entre les deux hypothèses ne semble se détacher. L'axe secondaire, quant à lui, oppose les modèles avec coefficients de corrélations bruts sur 20 ans représentés plutôt sur la partie supérieure du graphique et les autres modèles plutôt sur la partie inférieure.

Le choix du coefficient de corrélation linéaire contribue donc à la détermination des deux axes et semble donc être l'hypothèse mathématique la plus déterminante des résultats.

Cependant, l'analyse en composantes principales ne permet pas de quantifier la proximité ou non des différents modèles de Reverse Stress Test. C'est pourquoi cette analyse des données peut être complétée par une méthode de classification : le partitionnement en k-moyennes.

II Partitionnement en k-moyennes

Ce partitionnement est une méthode de classification qui permet de représenter dans un espace multidimensionnel des individus en fonction des valeurs de leurs variables respectives puis de les regrouper en k groupes en fonction de la proximité entre les individus.

1 Théorie

Cet algorithme possède deux paramètres principaux :

1. Le nombre de groupes k à former dans le jeu de données
2. La formule de la distance pour estimer la proximité des individus

Distance euclidienne

C'est la distance la plus usuelle.

Soit X et Y deux individus à n dimensions $X = (x_1, \dots, x_n)$ et $Y = (y_1, \dots, y_n)$ alors la distance euclidienne entre X et Y se définit par :

$$\text{dist}(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Pour que les variables soient équilibrées, au niveau de leur contribution dans le calcul de la distance, il est nécessaire de réduire le jeu de données.

Le partitionnement en k-moyennes a été proposé par plusieurs scientifiques sous différentes formes. Il est souvent basé sur l'algorithme de Lloyd en informatique. Mais c'est **l'algorithme de Hartigan et Wong** (1979) qui sera utilisé pour la comparaison des 18 modèles. En effet, ce dernier algorithme a pour but spécifique de créer une partition localement optimale en k portions. Le fonctionnement de l'algorithme est expliqué plus en détail par Laurence Morissette et Sylvain Chartier (2013, cf. [13]).

Algorithme de Hartigan et Wong (1/2)

Cette allocation optimale des groupes, réalisée en déplaçant l'ensemble des individus d'un groupe à un autre, permet de déterminer le centre de chaque groupe puis de retenir la classification qui minimise la distance des individus au centre de leur groupe et maximise par la même occasion la distance des individus des centres des autres groupes.

Le centre C_i du groupe n°i G_i vaut :

$$C_i = \min_X \sum_{Y \in G_i} \text{dist}(X, Y)$$

L'algorithme a pour objectif la minimisation de la distance intra-groupe (notée SSE) :

$$\text{SSE} = \sum_{i=1}^k \sum_{X \in G_i} (X - C_i)^2$$

Algorithme de Hartigan et Wong (2/2)

Les étapes d'affectation des individus aux groupes peuvent se synthétiser comme ceci :

1. Répartition aléatoire de l'ensemble des k centres de groupes parmi les individus.
2. Assignment des individus au groupe du centre le plus proche.
3. Calcul des centres des groupes.
4. Répétition de l'étape (2) et (3) jusqu'à convergence.

Sur le jeu de données, cet algorithme peut permettre de classifier les 18 modèles en groupes de similitude en fonction des valeurs des 25 variables représentatives ou en fonction des 12 *outputs*. Il sera alors possible de faire varier k pour estimer les principales sources de différenciation entre plusieurs groupes. Il est nécessaire de faire tourner l'algorithme à plusieurs reprises afin de bien déterminer la classification optimale en fonction des répartitions aléatoires de l'étape (1).

2 Application sur le jeu des 25 variables représentatives des modèles

2-A Détermination du paramètre k

Tout d'abord, il est nécessaire de déterminer le nombre de groupes de la partition. Pour cela, il suffit de comparer le SSE final en fonction du nombre de groupes choisi. Plus le nombre k se rapprochera du nombre d'individus (ici 18) plus le SSE tendra naturellement vers 0. Cependant, l'évolution du SSE en fonction de k n'est pas linéaire, un graphique peut donc permettre de déterminer le choix d'un k dont le SSE a bien diminué par rapport à $k = 1$ mais qui est aussi un nombre de groupes pas trop important.

Une partition peut être jugée adéquate si l'ajout d'un groupe a sensiblement fait diminuer le SSE de la partition précédente mais dont l'ajout d'un groupe supplémentaire n'apporterait pas de nette amélioration du SSE. Cette méthode est l'analogie de la méthode du coude vue dans l'ACP (cf. Partie 7-I_2A).

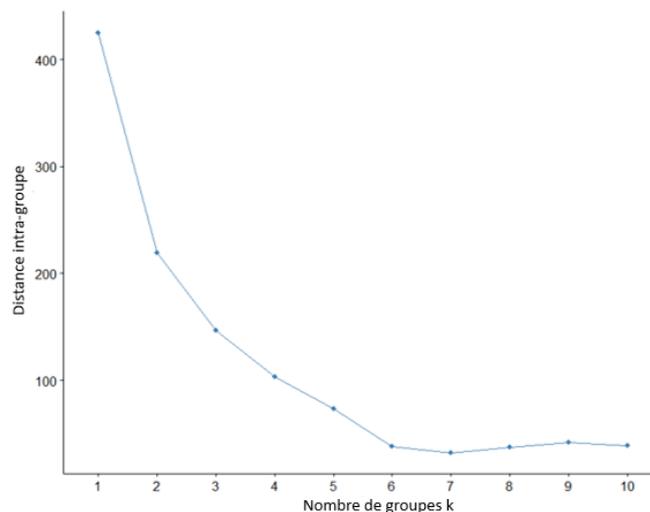


FIGURE 47 – Evolution du SSE sur les 25 variables en fonction du nombre de groupes

On observe que le fléchissement le plus important dans le calcul du SSE se situe entre le passage de $k = 1$ à $k = 2$. Les fléchissements suivants sont aussi conséquents jusqu'à $k = 6$. **Les partitions en 2 ou en 6 groupes peuvent donc se révéler intéressantes.**

2-B Application du partitionnement en 2-moyennes

Le partitionnement en k-moyennes se réalise sur l'ensemble des variables du jeu de données mais une représentation graphique en deux dimensions permet de bien illustrer la classification. Cette représentation est réalisée sur les mêmes axes principaux que l'ACP (cf. Partie 7-I_2D).

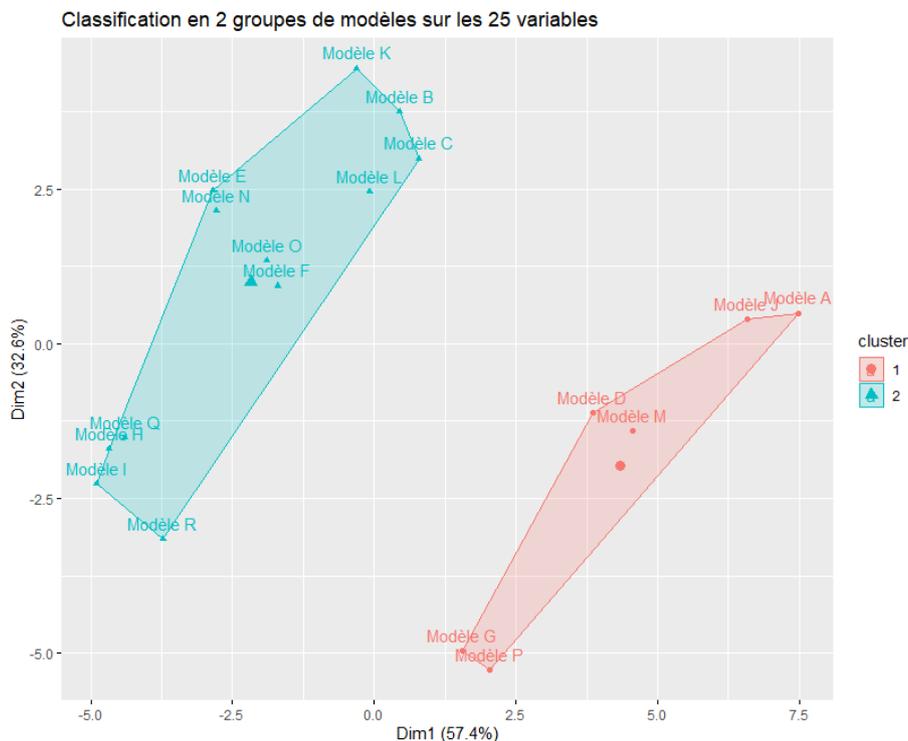


FIGURE 48 – Représentation bidimensionnelle du partitionnement en 2-moyennes

A noter que le sens respectif de chacun des deux axes est inversé par rapport à la représentation de l'ACP, mais la représentation sur R demeure bien la même.

La séparation optimale en deux groupes parmi les 18 modèles qui a été retenue par le partitionnement en 2-moyennes a été de séparer les modèles A, D, G, J, M et P des autres. Ces six modèles du **Groupe 1** ont comme caractéristique commune d'avoir utilisé le coefficient de corrélation brut sur 20 ans (sans pondération). Le **Groupe 2** regroupe donc le reste des modèles. **Le choix du coefficient de corrélation semble donc se confirmer comme l'hypothèse mathématique la plus déterminante des trois hypothèses sur les valeurs des 25 variables de chaque modèle.** La classification en 6 groupes pourra donner des particularités plus fines sur le reste des variables.

2-C Application du partitionnement en 6-moyennes

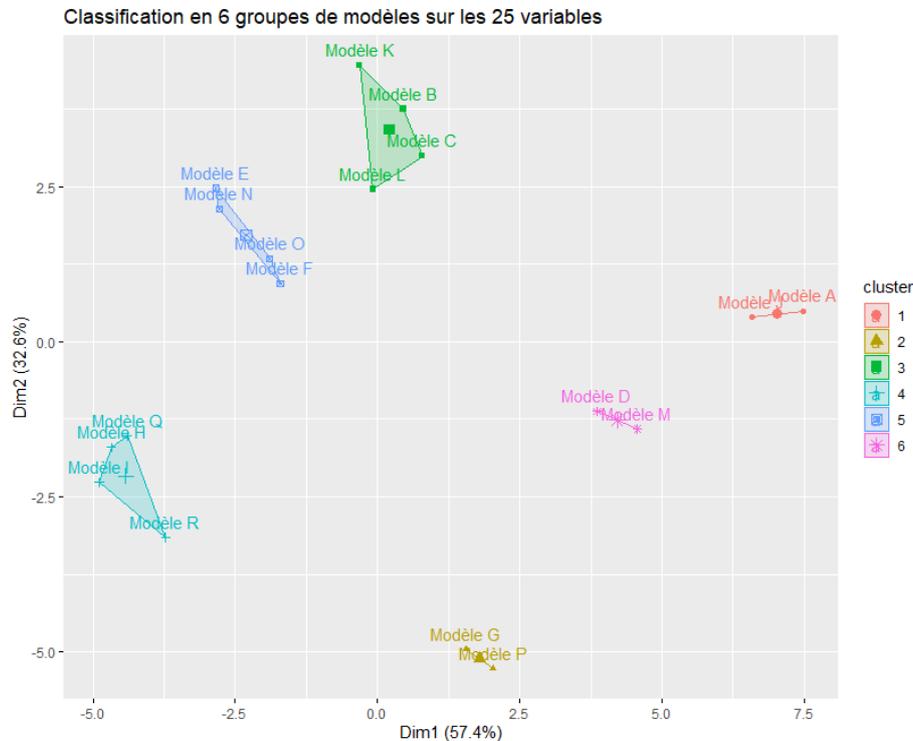


FIGURE 49 – Représentation bidimensionnelle du partitionnement en 6-moyennes

Le partitionnement en 6-moyennes a séparé chacun des deux précédents groupes en trois. Ces sous-groupes étaient soulignés graphiquement par la représentation des différentes variables lors de l'analyse en composantes principales (cf. Partie 7-I_2D). Le groupe constitué des 6 modèles dans le partitionnement en 2-moyennes (cf. groupe de droite Partie 7 - II_2B) s'est scindé, ici, en trois en fonction des lois marginales. En effet, le **Groupe 1** regroupe les modèles A et J utilisant des lois marginales normales, le **Groupe 6** regroupe les modèles D et M utilisant des lois marginales personnalisées et le **Groupe 2** regroupe les modèles G et P utilisant des lois marginales calibrées par estimateur gaussien. Quant au groupe constitué des 12 modèles dans le partitionnement en 2-moyennes (cf. groupe de gauche Partie 7 - II_2B), celui-ci est séparé en 3 groupes également avec cette méthode, les sous-groupes reprennent exactement ce critère de marginales : le **Groupe 3** regroupe les modèles à loi marginale normale, le **Groupe 5** regroupe les modèles à loi marginale personnalisée et le **Groupe 4** regroupe les lois marginales par estimateur gaussien. **Cette classification illustre que l'hypothèse mathématique du choix de la copule est moins impactant que les deux autres hypothèses sur les valeurs des 25 variables.**

A l'instar de l'analyse en composantes principales, il peut être intéressant d'appliquer cette méthode de partitionnement sur la base de données ne prenant en compte que les *outputs* des Reverse Stress Tests qui seront, *in fine*, communiqués à la gouvernance.

3 Application sur le jeu des 12 variables *outputs* des modèles

3-A Détermination du paramètre k

Dans un premier temps, le nombre k doit de nouveau être déterminé sur ce nouveau jeu restreint de variables.

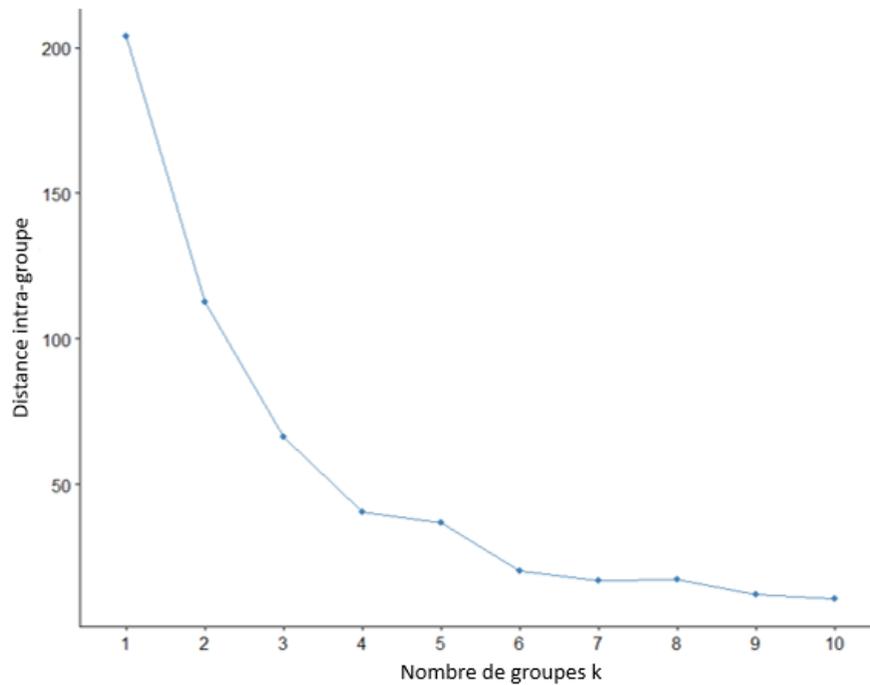


FIGURE 50 – Evolution du SSE sur les 12 *outputs* en fonction du nombre de groupes

Cette fois-ci le choix optimal k semble graphiquement plus visualisable. Le SSE diminue de façon très importante jusqu'à $k = 4$ puis tend à stagner. **La partition en 4 groupes semble donc la plus intéressante.**

3-B Application du partitionnement en 4-moyennes

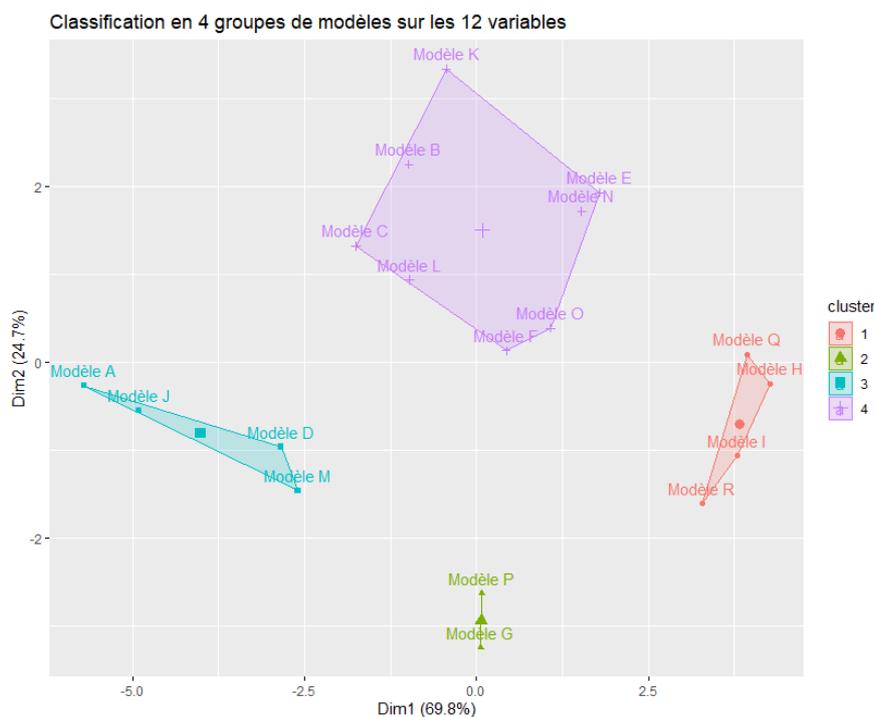


FIGURE 51 – Représentation bidimensionnelle du partitionnement en 4-moyennes

Cette fois-ci la classification en 4-moyennes donne des groupes de tailles assez inégales. Le groupe composé du plus de modèles est le **Groupe 4**, il regroupe l'ensemble des modèles utilisant des lois marginales normales ou personnalisées excepté les modèles utilisant un coefficient de corrélation linéaire brut sur 20 ans. Ces quatre derniers modèles sont eux regroupés au sein du **Groupe 3**. Le reste des modèles concerne donc ceux utilisant des lois marginales calibrées par estimateur gaussien. De la même manière, ceux utilisant des coefficients de corrélation linéaire sur 20 ans sont regroupés au sein du **Groupe 2** et les quatre modèles restants au sein du **Groupe 1**.

Cette classification en 4 groupes, uniquement sur les *outputs* des Reverse Stress Tests, va dans le même sens que les classifications précédentes. **Elle confirme que le choix de la copule comme hypothèse mathématique semble bien moins explicatif des résultats que les deux autres hypothèses.** Mais aucune hiérarchisation ne ressort entre l'hypothèse de marginales et celle des coefficients de corrélation contrairement à la classification précédente sur le jeu de 25 variables. Cependant, **cette classification permet de spécifier que la distinction des modèles par l'hypothèse du choix des marginales se fait avant entre les marginales par estimateur gaussien et les autres marginales.**

Ces différentes classifications permettent bien de regrouper des modèles similaires en fonction des hypothèses mathématiques de chacun. Elles complètent les analyses en composantes principales en séparant de manière optimale les modèles en fonction des différentes variables. Cependant, ces classifications ne permettent toujours pas de dire si les modèles classés dans des groupes différents le sont de façon significative. C'est ce degré de significativité qui permettra de conclure sur la robustesse ou non du module Excel (représenté par le modèle A).

4 Synthèse des trois partitionnements en k-moyennes

Les 2 classifications sur les 25 variables représentatives des modèles de Reverse Stress Test sont résumées dans le tableau suivant, récapitulatif des hypothèses mathématiques des 18 modèles.

	Hypothèses mathématiques			Sur les 25 variables	
	Lois des marginales	Copule	Corrélation de Pearson	k = 2	k = 6
Modèle A	Normales	Gaussienne	Sur 20 ans sans pondération	Groupe 1	Groupe 1
Modèle B	Normales	Gaussienne	Sur 10 ans sans pondération	Groupe 2	Groupe 3
Modèle C	Normales	Gaussienne	Sur 20 ans avec pondération	Groupe 2	Groupe 3
Modèle D	Personnalisées	Gaussienne	Sur 20 ans sans pondération	Groupe 1	Groupe 6
Modèle E	Personnalisées	Gaussienne	Sur 10 ans sans pondération	Groupe 2	Groupe 5
Modèle F	Personnalisées	Gaussienne	Sur 20 ans avec pondération	Groupe 2	Groupe 5
Modèle G	Estimateur à noyau gaussien	Gaussienne	Sur 20 ans sans pondération	Groupe 1	Groupe 2
Modèle H	Estimateur à noyau gaussien	Gaussienne	Sur 10 ans sans pondération	Groupe 2	Groupe 4
Modèle I	Estimateur à noyau gaussien	Gaussienne	Sur 20 ans avec pondération	Groupe 2	Groupe 4
Modèle J	Normales	Student	Sur 20 ans sans pondération	Groupe 1	Groupe 1
Modèle K	Normales	Student	Sur 10 ans sans pondération	Groupe 2	Groupe 3
Modèle L	Normales	Student	Sur 20 ans avec pondération	Groupe 2	Groupe 3
Modèle M	Personnalisées	Student	Sur 20 ans sans pondération	Groupe 1	Groupe 6
Modèle N	Personnalisées	Student	Sur 10 ans sans pondération	Groupe 2	Groupe 5
Modèle O	Personnalisées	Student	Sur 20 ans avec pondération	Groupe 2	Groupe 5
Modèle P	Estimateur à noyau gaussien	Student	Sur 20 ans sans pondération	Groupe 1	Groupe 2
Modèle Q	Estimateur à noyau gaussien	Student	Sur 10 ans sans pondération	Groupe 2	Groupe 4
Modèle R	Estimateur à noyau gaussien	Student	Sur 20 ans avec pondération	Groupe 2	Groupe 4

FIGURE 52 – Synthèse des classifications des 18 modèles sur les 25 variables représentatives

La classification sur les 12 *outputs* des modèles de Reverse Stress Test est résumée dans le tableau suivant, récapitulatif des hypothèses mathématiques des 18 modèles.

	Hypothèses mathématiques			Sur les 12 <i>outputs</i>
	Lois des marginales	Copule	Corrélation de Pearson	k = 4
Modèle A	Normales	Gaussienne	Sur 20 ans sans pondération	Groupe 3
Modèle B	Normales	Gaussienne	Sur 10 ans sans pondération	Groupe 4
Modèle C	Normales	Gaussienne	Sur 20 ans avec pondération	Groupe 4
Modèle D	Personnalisées	Gaussienne	Sur 20 ans sans pondération	Groupe 3
Modèle E	Personnalisées	Gaussienne	Sur 10 ans sans pondération	Groupe 4
Modèle F	Personnalisées	Gaussienne	Sur 20 ans avec pondération	Groupe 4
Modèle G	Estimateur à noyau gaussien	Gaussienne	Sur 20 ans sans pondération	Groupe 2
Modèle H	Estimateur à noyau gaussien	Gaussienne	Sur 10 ans sans pondération	Groupe 1
Modèle I	Estimateur à noyau gaussien	Gaussienne	Sur 20 ans avec pondération	Groupe 1
Modèle J	Normales	Student	Sur 20 ans sans pondération	Groupe 3
Modèle K	Normales	Student	Sur 10 ans sans pondération	Groupe 4
Modèle L	Normales	Student	Sur 20 ans avec pondération	Groupe 4
Modèle M	Personnalisées	Student	Sur 20 ans sans pondération	Groupe 3
Modèle N	Personnalisées	Student	Sur 10 ans sans pondération	Groupe 4
Modèle O	Personnalisées	Student	Sur 20 ans avec pondération	Groupe 4
Modèle P	Estimateur à noyau gaussien	Student	Sur 20 ans sans pondération	Groupe 2
Modèle Q	Estimateur à noyau gaussien	Student	Sur 10 ans sans pondération	Groupe 1
Modèle R	Estimateur à noyau gaussien	Student	Sur 20 ans avec pondération	Groupe 1

FIGURE 53 – Synthèse des classifications des 18 modèles sur les 12 *outputs*

III Analyse qualitative des *outputs* des modèles

Une idée intuitive est de regarder, pour chaque impact cible, si les résultats renvoyés par chacun des 18 modèles sont significativement différents. En effet, en fonction du degré de précision exigée par la gouvernance pour le résultat d'un Reverse Stress Test, l'arrondi peut amener à faire converger les résultats des différents modèles.

1 Analyse qualitative de l'impact à - 3 Md €

	Marginales	Copule	Coefficient corrélation de Pearson	CAC moyen -3Md	OAT moyen -3Md	IPD moyen -3Md
Modele A	Normales	Gaussienne	20 ans non pondere	5 105	0,37	199,19
Modele B	Normales	Gaussienne	10 ans non pondere	5 097	0,42	199,63
Modele C	Normales	Gaussienne	20 ans pondere	5 100	0,41	199,46
Modele D	Personnalisées	Gaussienne	20 ans non pondere	5 102	0,39	199,44
Modele E	Personnalisées	Gaussienne	10 ans non pondere	5 097	0,43	199,58
Modele F	Personnalisées	Gaussienne	20 ans pondere	5 095	0,40	199,79
Modele G	Estimateur gaussien	Gaussienne	20 ans non pondere	5 093	0,37	200,04
Modele H	Estimateur gaussien	Gaussienne	10 ans non pondere	5 089	0,39	200,13
Modele I	Estimateur gaussien	Gaussienne	20 ans pondere	5 086	0,42	200,28
Modele J	Normales	Student	20 ans non pondere	5 104	0,38	199,24
Modele K	Normales	Student	10 ans non pondere	5 100	0,43	199,46
Modele L	Normales	Student	20 ans pondere	5 098	0,39	199,60
Modele M	Personnalisées	Student	20 ans non pondere	5 100	0,38	199,55
Modele N	Personnalisées	Student	10 ans non pondere	5 095	0,42	199,72
Modele O	Personnalisées	Student	20 ans pondere	5 096	0,42	199,81
Modele P	Estimateur gaussien	Student	20 ans non pondere	5 093	0,39	199,94
Modele Q	Estimateur gaussien	Student	10 ans non pondere	5 090	0,42	200,12
Modele R	Estimateur gaussien	Student	20 ans pondere	5 087	0,39	200,32

FIGURE 54 – *Outputs* des 18 modèles de Reverse Stress Test pour un impact cible à - 3 Md €

Pour rappel, le module Excel (avec les mêmes hypothèses mathématiques que le modèle A) pour un impact à - 3 Md € retient l'évènement d'un CAC 40 à 5116 points, des obligations à maturité de l'Etat à un taux de -0,39% et un IPD à 198,51 points. Il est néanmoins logique d'avoir une certaine différenciation dans les résultats entre les deux modules. En effet, le Reverse Stress Test est implémenté sur Excel reviendrait à renvoyer l'évènement le plus fréquent et non l'évènement conjoint des facteurs de risque moyens comme renvoyé en sortie par le module sous R. Mais les résultats des deux implémentations demeurent tout de même du même ordre de grandeur avec des *outputs* du modèle A égaux à un CAC 40 à 5105 points, des OAT 10Y à -0,37% et un IPD à 199,19 points.

Concernant les 18 modèles implémentés sur R, les résultats sont très similaires entre eux. L'*output* minimal du CAC 40 est de 5086 points pour un maximum de 5104 points, soit une amplitude de seulement 18 points (moins de 1%). La valeur minimale est renvoyée pour le modèle I, avec des lois marginales calibrées par estimateur gaussien et un coefficient de corrélation pondéré sur 20 ans. La valeur maximale est renvoyée par le modèle J, avec des lois marginales normales et le coefficient de corrélation brut sur 20 ans. Ces deux modèles confirment les interprétations déduites précédemment de l'ACP (cf. Partie 7-I_2D).

De même, les *outputs* des OAT 10Y varient de $-0,43\%$ à $-0,37\%$ seulement, soit une amplitude maximale de 6 points de base. Cette variation est cependant relativement plus importante, elle représente environ une variation relative de 15% entre les modèles les plus différents sur cet *output*.

Quant aux *outputs* de l'IPD, ils varient de 199,19 points à 200,32 points en opposant les mêmes modèles que le CAC 40. Ces résultats confirment eux aussi l'interprétation réalisée précédemment. L'amplitude de variation est de 1,13 points soit une variation maximale de moins de 1% de la valeur de l'IPD.

Les variations du CAC 40 et de l'IPD peuvent être considérées comme non significatives entre les modèles, *a contrario* de celle des OAT 10Y qui est, elle, relativement plus importante. Cependant, cette conclusion est à tempérer car les taux étant sur des valeurs très basses les variations relatives entre les modèles sont naturellement plus conséquentes que pour le CAC 40 et l'IPD.

2 Analyse qualitative de l'impact à - 5 Md €

	Marginales	Copule	Coefficient corrélation de Pearson	CAC moyen -5Md	OAT moyen -5Md	IPD moyen -5Md
Modele A	Normales	Gaussienne	20 ans non pondere	4 819	0,41	196,31
Modele B	Normales	Gaussienne	10 ans non pondere	4 810	0,47	196,75
Modele C	Normales	Gaussienne	20 ans pondere	4 811	0,45	196,69
Modele D	Personnalisées	Gaussienne	20 ans non pondere	4 809	0,43	196,95
Modele E	Personnalisées	Gaussienne	10 ans non pondere	4 797	0,49	197,46
Modele F	Personnalisées	Gaussienne	20 ans pondere	4 804	0,46	197,21
Modele G	Estimateur gaussien	Gaussienne	20 ans non pondere	4 802	0,40	197,45
Modele H	Estimateur gaussien	Gaussienne	10 ans non pondere	4 788	0,49	198,10
Modele I	Estimateur gaussien	Gaussienne	20 ans pondere	4 793	0,46	197,77
Modele J	Normales	Student	20 ans non pondere	4 814	0,41	196,56
Modele K	Normales	Student	10 ans non pondere	4 808	0,50	196,80
Modele L	Normales	Student	20 ans pondere	4 805	0,46	197,08
Modele M	Personnalisées	Student	20 ans non pondere	4 809	0,42	196,93
Modele N	Personnalisées	Student	10 ans non pondere	4 800	0,49	197,29
Modele O	Personnalisées	Student	20 ans pondere	4 800	0,46	197,44
Modele P	Estimateur gaussien	Student	20 ans non pondere	4 799	0,42	197,55
Modele Q	Estimateur gaussien	Student	10 ans non pondere	4 789	0,47	198,01
Modele R	Estimateur gaussien	Student	20 ans pondere	4 793	0,46	197,89

FIGURE 55 – *Outputs* des 18 modèles de Reverse Stress Test pour un impact cible à - 5 Md €

L'analyse des *outputs* pour un impact à - 5 Md € aboutit aux mêmes conclusions. En effet, les résultats du CAC 40 varient de 4788 points à 4819 points, soit une amplitude de 31 points (moins de 1%). Ceux des OAT 10Y varient de $-0,50\%$ à $-0,41\%$, soit une amplitude de 9 points de base différenciant les résultats des deux modèles les plus divergents d'environ 20%. Et ceux de l'IPD varient de 196,31 points à 198,10 points, soit une amplitude de 1,79 points (un peu moins de 1% de variation maximale).

3 Analyse qualitative de l'impact à - 7 Md €

	Marginales	Copule	Coefficient corrélation de Pearson	CAC moyen -7Md	OAT moyen -7Md	IPD moyen -7Md
Modele A	Normales	Gaussienne	20 ans non pondere	4 535	0,45	193,29
Modele B	Normales	Gaussienne	10 ans non pondere	4 524	0,55	193,77
Modele C	Normales	Gaussienne	20 ans pondere	4 520	0,52	194,08
Modele D	Personnalisées	Gaussienne	20 ans non pondere	4 520	0,46	194,27
Modele E	Personnalisées	Gaussienne	10 ans non pondere	4 508	0,56	194,75
Modele F	Personnalisées	Gaussienne	20 ans pondere	4 511	0,52	194,65
Modele G	Estimateur gaussien	Gaussienne	20 ans non pondere	4 509	0,45	194,88
Modele H	Estimateur gaussien	Gaussienne	10 ans non pondere	4 495	0,55	195,42
Modele I	Estimateur gaussien	Gaussienne	20 ans pondere	4 497	0,50	195,44
Modele J	Normales	Student	20 ans non pondere	4 533	0,45	193,50
Modele K	Normales	Student	10 ans non pondere	4 516	0,58	194,21
Modele L	Normales	Student	20 ans pondere	4 517	0,54	194,18
Modele M	Personnalisées	Student	20 ans non pondere	4 519	0,45	194,33
Modele N	Personnalisées	Student	10 ans non pondere	4 508	0,57	194,79
Modele O	Personnalisées	Student	20 ans pondere	4 507	0,52	194,81
Modele P	Estimateur gaussien	Student	20 ans non pondere	4 511	0,43	194,75
Modele Q	Estimateur gaussien	Student	10 ans non pondere	4 499	0,56	195,30
Modele R	Estimateur gaussien	Student	20 ans pondere	4 500	0,50	195,32

FIGURE 56 – *Outputs* des 18 modèles de Reverse Stress Test pour un impact cible à - 7 Md €

Les amplitudes des 3 facteurs de risque semblent s'amplifier en augmentant la perte cible. En effet, les *outputs* du CAC 40 varient maintenant de 4535 points à 4495 points soit une amplitude de 40 points (représentant relativement quasiment 1%). Ceux des OAT 10Y varient de -0,58% à -0,43%, d'une amplitude donc de 15 points de base (dépassant relativement les 30% de variation maximale). Ceux de l'IPD varient de 193,29 points à 195,44 points, soit une amplitude de 2,15 points (dépassant les 1% de variation relative).

4 Analyse qualitative de l'impact à - 10 Md €

	Marginales	Copule	Coefficient corrélation de Pearson	CAC moyen -10Md	OAT moyen -10Md	IPD moyen -10Md
Modele A	Normales	Gaussienne	20 ans non pondere	4 114	0,51	188,57
Modele B	Normales	Gaussienne	10 ans non pondere	4 087	0,69	189,77
Modele C	Normales	Gaussienne	20 ans pondere	4 093	0,60	189,57
Modele D	Personnalisées	Gaussienne	20 ans non pondere	4 103	0,48	189,20
Modele E	Personnalisées	Gaussienne	10 ans non pondere	4 074	0,67	190,60
Modele F	Personnalisées	Gaussienne	20 ans pondere	4 082	0,62	190,21
Modele G	Estimateur gaussien	Gaussienne	20 ans non pondere	4 082	0,48	190,47
Modele H	Estimateur gaussien	Gaussienne	10 ans non pondere	4 065	0,67	191,12
Modele I	Estimateur gaussien	Gaussienne	20 ans pondere	4 067	0,57	191,17
Modele J	Normales	Student	20 ans non pondere	4 110	0,52	188,69
Modele K	Normales	Student	10 ans non pondere	4 092	0,69	189,56
Modele L	Normales	Student	20 ans pondere	4 093	0,63	189,54
Modele M	Personnalisées	Student	20 ans non pondere	4 099	0,50	189,27
Modele N	Personnalisées	Student	10 ans non pondere	4 077	0,67	190,29
Modele O	Personnalisées	Student	20 ans pondere	4 084	0,62	190,31
Modele P	Estimateur gaussien	Student	20 ans non pondere	4 085	0,48	190,26
Modele Q	Estimateur gaussien	Student	10 ans non pondere	4 067	0,65	191,02
Modele R	Estimateur gaussien	Student	20 ans pondere	4 070	0,59	190,90

FIGURE 57 – *Outputs* des 18 modèles de Reverse Stress Test pour un impact cible à - 10 Md €

L'analyse pour l'impact à - 10 Md € confirme l'accentuation des amplitudes. Les résultats de CAC 40 varient de 4114 points à 4065 points, de 49 points d'amplitude dépassant les 1,2% de variation relative entre les modèles A et I. Les résultats des OAT 10 varient maintenant de -0,69% à -0,48% soit une amplitude de 21 points de base, dépassant les 35% de variation relative. Les résultats de l'IPD varient, quant à eux, de 188,57 points à 191,12 points soit de 2,55 points d'amplitude dépassant les 1,3% de relation relative.

5 Bilan des comparaisons des 18 modèles

Les amplitudes entre les *outputs* demeurent très minimes concernant les facteurs de risque action et immobilier entre les 18 modèles d'un même impact cible. A noter, cependant, que l'amplitude relative tend à s'amplifier pour des pertes cibles plus conséquentes et touchant donc plus aux queues de distribution des marginales des facteurs de risque.

Concernant le facteur de risque des taux, les différences des *outputs* entre les modèles sont beaucoup plus significatives. Pour confirmer l'analyse de l'ACP, il est possible de regarder les modèles tirant les *outputs* extrêmes pour chaque impact afin de confirmer une redondance.

- Pour l'impact à - 3 Md € la valeur des *outputs* maximaux des OAT 10Y provient des modèles A et G. La valeur minimale de taux provient des modèles E et K.
- Pour l'impact à - 5 Md € la valeur des *outputs* maximaux des OAT 10Y provient des modèles A et J. La valeur minimale du taux provient du modèle K.
- Pour l'impact à - 7 Md € la valeur des *outputs* maximaux des OAT 10Y provient du modèle P. La valeur minimale provient du modèle K.
- Pour l'impact à - 10 Md € la valeur des *outputs* maximaux des OAT 10Y provient du modèle D, G et P. La valeur minimale provient du modèle B et K.

Les modèles B, E et K renvoyant les OAT 10Y moyens les plus faibles reposent sur des coefficients de corrélation bruts sur 10 ans. Alors que les modèles A, D, G, J et P renvoyant les OAT moyens les plus hauts, reposent eux sur des coefficients de corrélation bruts sur 20 ans. Cela confirme la corrélation forte entre les *outputs* des taux avec l'axe vertical de l'ACP sur les 12 variables (cf. Partie 7 - I_3B), axe vertical différenciant les modèles selon leur coefficient de corrélation.

Jusqu'ici les coefficients de corrélation bruts sur 10 ans et pondérés sur 20 ans étaient souvent associés dans les mêmes classifications, alors qu'ici seuls les modèles avec des coefficients bruts sur 10 ans sont ressortis. Cela peut s'expliquer par le fait que la corrélation CAC 40/IPD varie peu en valeur entre la corrélation brute sur 10 ans et la corrélation pondérée sur 20 ans, contrairement aux OAT 10Y dont les corrélations avec les autres facteurs de risque évoluent davantage (cf. Partie 5-III_3).

Le choix du coefficient de corrélation est donc bien l'hypothèse mathématique la plus forte et impacte de façon très significative les résultats des OAT 10Y.

Cependant, l'importance des taux dans le Reverse Stress Test reste à relativiser au vu de sa sensibilité qui correspond à une perte de 2 M € par point de base à la hausse. Donc, avec une amplitude maximale pour l'impact à - 10 Md € de 21 points de base, la différence de contribution à l'impact est au maximum 42 M € ce qui est relativement faible sur les milliards d'euros d'impact. Ce point avait été souligné par l'étude des contributions à l'impact cible des différents facteurs de risque au sein du module Excel (cf. Partie 4-VII). Cette faible sensibilité cause une volatilité

dans le tirage du facteur de risque au sein d'un impact cible. En effet, pour les quatre impacts, les écarts-types de variation des taux sont très conséquents, d'environ 66 points de base. Les années simulées des tirages, pour les quatre impacts et quels que soient les modèles, renvoient donc des taux dont l'écart moyen avec l'*output* correspondant est de l'ordre de 0,66 point de pourcentage. Pour des valeurs d'*outputs* des OAT 10Y autour de $-0,50\%$ cette volatilité est gigantesque.

Cette explication permet de comprendre la raison du regroupement des modèles B et K, d'*outputs* des OAT 10Y très différents, au sein de le partitionnement en 6-moyennes sur les 25 variables (cf. Partie 7 - II_2C). L'introduction de ces variables d'écart-type s'est donc révélée être très pertinente.

A titre de comparaison (cf. Annexe 4) pour les quatre impacts, les écarts-types du CAC 40 sont d'environ 150 points et ceux de l'IPD d'environ 8 points. La volatilité est donc, relativement aux valeurs des cours, bien plus faible. Elle s'explique inversement par la forte contribution des facteurs de risque action et immobilier dans l'impact cible.

La forte volatilité des OAT 10Y dans les événements tirés fragilise l'utilisation de l'*output* moyen et sa transmission à la gouvernance. Cependant, les *outputs* du CAC 40 et de l'IPD plus stables et robustes sont eux plus interprétables. Mais la transmission de leurs valeurs sans arrondi pourrait transformer l'interprétation d'évènement le plus probable à évènement certain, il est donc important de conserver une retenue sur la précision des résultats. Il a été décidé, au sein de DFIN, de transmettre un CAC 40 arrondi à la centaine de points et un IPD arrondi au quart de dizaine de points.

Sous de tels arrondis l'ensemble des 18 modèles renvoient alors bien des résultats quasiment identiques. **Le module Excel peut donc être maintenu, il est jugé suffisamment robuste.** De plus, l'hypothèse forte des coefficients de corrélation est modifiable sur Excel (cf. Partie 5-IV). **La méthodologie du tirage est cependant intégrée au module Excel et remplace la fonction de score afin de faciliter la compréhension du sens des *outputs*.**

Conclusion

L'objectif de ce mémoire était de mettre en place un module de Reverse Stress Test pour les besoins de la Direction Financière de la Section Générale de la Caisse des Dépôts et Consignations. Ce module est un outil d'aide à la décision qui permet d'identifier l'évènement le plus probable associé à un objectif fixé, en l'occurrence un impact négatif sur la Caisse des Dépôts.

Le module de Reverse Stress Test devait allier, premièrement, flexibilité dans la modification des hypothèses et l'ajout de facteurs de risque, deuxièmement, simplicité dans l'utilisation et la compréhension de ses résultats, et troisièmement, performance statistique dans la modélisation.

Les trois facteurs (CAC 40, OAT 10Y et IPD) de risque (action, taux et immobilier) retenus dans le mémoire ont été fixés afin d'approximer les principaux risques de solvabilité de la Caisse des Dépôts mais pourront être remplacés ou complétés à l'avenir par d'autres facteurs de risque. Ces trois facteurs ont cependant permis d'explorer des typologies de données différentes ainsi que des granularités et des contraintes différentes sur leurs historiques. L'étude statistique de ces trois facteurs de risque a induit l'utilisation de leurs marginales, représentées différemment selon les facteurs de risque.

Le premier module de Reverse Stress Test mis en place a été réalisé sur Excel afin d'obtenir un module simple d'un point de vue statistique. La conception mathématique a reposé sur une optimisation sous contraintes d'une fonction de densité normale tridimensionnelle des marginales.

Ce module Excel a débouché sur trois problématiques. La première est que les sensibilités de la Caisse des Dépôts aux facteurs de risque sont très grandement impactantes dans la détermination de l'évènement solution du problème d'optimisation. Ces sensibilités ont été transmises par la Direction des Risques du Groupe et prises comme des données exogènes dans ce mémoire mais elles devront être étudiées de manière plus approfondie lors de la mise en production du module de Reverse Stress Test finalement retenu par DFIN. La deuxième problématique concerne la difficulté de présentation à la gouvernance de l'évènement en *output* du module Excel dont les métriques de score n'étaient pas simples de compréhension et dont l'instabilité temporelle a été soulignée par les calculs des sensibilités du module Excel. Quant à la troisième problématique, elle concerne la simplification statistique qui a été réalisée au sein du module Excel, via les hypothèses mathématiques nécessaires à l'utilisation d'une fonction de score fermée, mais dont les impacts sur les résultats du modèle sont difficilement interprétables de manière robuste au sein du module Excel.

La suite du mémoire a porté sur la résolution de cette dernière problématique afin de quantifier de manière plus précise les hypothèses mathématiques du module Excel. Cet approfondissement

avait pour objectif de souligner les performances ou les faiblesses statistiques du module Excel afin d'estimer la pertinence de son utilisation au regard d'autres solutions envisageables.

Afin de *challenger* le module Excel, trois levées d'hypothèses mathématiques ont été étudiées : celle portant sur le choix de loi des marginales des facteurs de risque, celle portant sur la structure de corrélation retenue (copule) et la dernière portant sur la détermination des coefficients de corrélation entre les facteurs de risque. Ces levées d'hypothèses ont nécessité l'utilisation d'un outil statistique plus puissant (R) pour réaliser 18 modèles de Reverse Stress Test, issus des croisements entre ces trois hypothèses mathématiques.

La levée de l'hypothèse de lois marginales normales sur les facteurs de risque a amené à altérer la méthodologie du Reverse Stress Test réalisé sur Excel. L'optimisation sous contrainte d'une loi de densité multidimensionnelle a donc été remplacée par la réalisation d'un tirage d'évènements.

La modélisation par tirage sur R a permis de déboucher sur la résolution de la problématique précédente concernant le *story telling* de l'*output* du Reverse Stress Test. En effet, la fréquence d'apparition dans un tirage, remplaçant les métriques Excel peu précises du score, permet de présenter facilement les *outputs* du Reverse Stress Test comme étant les facteurs de risque moyens d'un tirage.

Les 18 modèles ont pu être comparés à l'aide de méthodes d'analyse de données (analyse en composantes principales et partitionnement en k-moyennes). La comparaison des modèles a permis de hiérarchiser les trois hypothèses mathématiques : entre le choix du coefficient de corrélation qui est l'hypothèse la plus forte, le choix de la marginale et le choix de la copule qui est l'hypothèse la plus faible.

La comparaison des modèles a aussi souligné que la présence d'un facteur de risque peu contributeur à l'impact global par rapport à d'autres facteurs de risque engendrait du bruit dans l'interprétation des résultats. Il est donc d'autant plus opportun de se pencher sur l'étude approfondie des sensibilités ainsi que de réfléchir si au niveau de la direction, il n'est pas plus pertinent d'avoir un Reverse Stress Test avec moins de facteurs de risque mais dont les influences sont équivalentes, ce qui minimise la volatilité des résultats.

DFIN a souhaité appliquer des arrondis aux *outputs* du Reverse Stress Test lors de leur présentation à la gouvernance. En effet, des résultats trop précis pourraient sous-entendre une prévision de l'évènement en sortie du modèle. Etant donné les arrondis réalisés, il est décidé de maintenir l'utilisation du module Excel dont les résultats arrondis sont très similaires aux autres modèles alternatifs implémentés sur R. Cependant, cette étude sur la performance du module Excel demeure une étude sur les données 2001-2020 et sur ces trois facteurs de risque uniquement. Néanmoins, une réflexion plus poussée sur le choix des coefficients de corrélation dans le modèle et sur la détermination des sensibilités de la CDC aux facteurs de risque est souhaitable eu égard à leurs impacts sur le Reverse Stress Test. En fonction des choix de modélisation retenus, ces éléments sont des paramètres facilement modifiables au sein du module Excel.

DFIN a déjà utilisé le module Excel, sur le périmètre action uniquement, avec cinq facteurs de risque ; il a été appliqué à des métriques de solvabilité. La présentation du nouveau module a été appréciée par la gouvernance qui a pu profiter des premiers résultats. De plus, la Direction des Risques, en tant que seconde ligne de défense du risque de modèle, a validé le modèle de Reverse Stress Test inhérent au module Excel.

DFIN a pour objectif de compléter le module Excel pour un Reverse Stress Test plus complet d'ici l'année prochaine. Le module devrait comporter le risque immobilier en plus du risque action (avec potentiellement plusieurs facteurs de risque) mais la question de l'introduction du facteur de risque taux ou non est encore en cours d'étude compte tenu notamment des conclusions du présent mémoire. L'impact cible sera basé sur l'écart à la cible et les sensibilités seront estimées en fonction de ce dernier. Des échanges avec les filiales financières du Groupe ont été entamés notamment sur des partages de la méthodologie de Reverse Stress Tests et sur les synergies à envisager. La prise en compte plus consolidée de certains risques tel que le risque de crédit (risque majeur pour Bpifrance, SFIL et La Banque Postale) est aussi un point à l'étude afin d'obtenir un module de Reverse Stress Tests le plus représentatif du Groupe Caisse des Dépôts.

Glossaire

Sigles de la Caisse des Dépôts

- CDC : Caisse des Dépôts et Consignations
- DFIN : Direction financière du Groupe Caisse des Dépôts en charge du bilan de la Section Générale
- DGAu : Direction de l’Audit interne de l’établissement public et du Groupe Caisse des Dépôts et Consignations
- DRG : Direction des Risques du Groupe Caisse des Dépôts

Sigles concernant les Reverse Stress Tests

- ABE : Autorité Bancaire Européenne
- EBA : European Banking Authority
- RST : Reverse Stress Test

Sigles concernant les facteurs de risque

- CAC : CAC 40
- OAT : Obligation Assimilable du Trésor
- OAT 10Y : OAT de maturité 10 ans
- IPD : Investment Property Databank

Sigles mathématiques

- ACP : Analyse en composantes principales
- NIG : loi normale-inverse gaussienne
- VaR : Value-at-Risk
- CVaR : Conditional Value-at-Risk

Bibliographie

Textes réglementaires

- [1] European Banking Authority, *Orientations sur les tests de résistance des établissements*, EBA/GL/2018/04, 19 juillet 2018.
- [2] Loi n°2019-486 relative à la croissance et la transformation des entreprises dite loi PACTE, Supervision de la Caisse des Dépôts et Consignations par l'ACPR article 112, 22 mai 2019.
- [3] Capital Requirements Regulation (CRR), *Common Equity Tier 1 capital*, partie 2, titre 1, chapitre 2, section 4, article 50, 26 juin 2013.

Mémoires d'actuariat

- [4] BIOLAY Y. (2021), *Intégration des filiales financières régulées dans l'optimisation du couple rendement-risque et déclinaison tactique de la stratégie optimale d'investissement long terme*, Mémoire d'Actuariat, Caisse des Dépôts et Consignations, ISFA.
- [5] DUVAL A. (2012), *Risque de crédit dans un portefeuille obligataire. Modèle de prise en compte du risque de dépendance*, Mémoire d'Actuariat, Altia, ISFA, pages 28-30 et 39-40.
- [6] GALATAUD G.(2010), *Modélisation de la dépendance entre branches d'assurance avec les modèles de chocs communs multiplicatifs*, Mémoire d'actuariat, Groupama, Le CNAM, pages 94-96.
- [7] GRANDPERRIN L. (2019), *Allocation de capital : théorie et pratique de la méthode d'Euler*, Mémoire d'Actuariat, SCOR, ISFA, pages 88-102.
- [8] LATHUILLE D.(2019), *Mise en place de Reverse Stress Tests dans le cadre de l'ORSA*, Mémoire d'Actuariat, BNP Paribas Cardif, ISFA.
- [9] MASIELLO E. (2010), *Modeling dependence by copula functions in an insurance context*, Mémoire d'Actuariat, ISFA, pages 11-16 et 29-32.
- [10] SCHEID E. (2011), *Dépendance entre les principaux risques financiers pour un assureur dans le cadre de Solvabilité II*, Mémoire d'Actuariat, Fixage, Université de Paris Dauphine, pages 32-51.

- [11] TRIEU T. D. (2013), *Une analyse exploratoire et une tarification spécifique aux seniors sur le portefeuille de Swiss Life dans le cadre du programme de recherche « Conséquence du vieillissement de la population sur la sinistralité des assurances »*, Mémoire d'Actuariat, Swiss Life ESSEC Business School, ISFA, pages 32-42.

Articles

- [12] BARNDORFF-NIELSEN O. (1977), « *Exponentially decreasing distributions for the logarithm of particle size* », *The Royal Society*, volume 35, numéro 1674, pages 401-409.
- [13] CHARTIER S. MORISSETTE Laurence (2013), *Tutorials in Quantitative Methods for Psychology*, Vol 9(1), p.15-24.
- [14] JARQUE C. M. et BERA A. K. (1980), « *Efficient tests for normality, homoscedasticity and serial independence of regression residuals* » *Economics Letters*, volume 6, numéro 3, pages 255-259.
- [15] KOLMOGOROV A. N. (1933), « *Sulla determinazione empirica di una legge di distribuzione* », *Giornale dell'Istituto Italiano degli Attuari*, volume 4, pages 83-91.
- [16] PARZEN E. (1962), « *On estimation of a probability density function and mode* », *The annals of mathematical statistics*, volume 33, numéro 3, pages 1065-1076.
- [17] ROSENBLATT M. (1956), « *Remarks on some nonparametric estimates of a density function* », *The annals of mathematical statistics*, volume 27, numéro 3, pages 832-837.
- [18] SHAPIRO S. S. et WILK M. B. (1965), « *An analysis of variance test for normality (complete samples)* », *Biometrika*, volume 52, numéros 3-4, pages 591-611.
- [19] SKLAR A. (1959), « *Fonctions de répartition à n dimensions et leurs marges* », *Publications de l'Institut Statistique de l'Université de Paris*, volume 8, pages 229-231.

Ouvrages

- [20] MARCEAU E. (2013), *Modélisation et évaluation quantitative des risques en actuariat*.

Cours

- [21] CHARPENTIER A. (2013), *Copules et risques multiples*.
- [22] CHARPENTIER A., *Cours de séries temporelles : théorie et applications*, Volume 1 *Introduction à la théorie des processus en temps discret. Modèles ARIMA et méthode Box Jenkins*, pages 79-80.

Source des données

- [23] *Bloomberg*, extractions du CAC 40, des OAT 10Y (janvier 2022) et de l'IEIF (avril 2021).

Table des figures

1	Comparaison entre un Stress Test et un Reverse Stress Test	13
2	Répartition du type d'actions cotées de la Section Générale au 30 juin 2020	16
3	Cours du facteur de risque action sur 21 ans	16
4	Cours du facteur de risque taux sur 21 ans	17
5	Cours du facteur de risque immobilier sur 20 ans	18
6	Représentation graphique des quartiles du CAC 40	20
7	Représentation graphique des principaux <i>drawdowns</i> du CAC 40 sur 20 ans	21
8	Représentation graphique des quantiles des variations relatives du CAC 40	22
9	Représentation graphique en histogramme des log-rendements hebdomadaires du CAC 40	23
10	Fonction d'autocorrélation du CAC 40	25
11	Fonction d'autocorrélation des OAT 10Y	26
12	Fonction d'autocorrélation de l'IPD	26
13	Régression linéaire sur le cours hebdomadaire historique du CAC 40	27
14	Régression linéaire sur le cours hebdomadaire actualisé du CAC 40	28
15	Tableau des paramètres du CAC 40 et des OAT 10Y	33
16	Discrétisation de la fonction de score en 2 dimensions entre le CAC 40 et les OAT 10Y	33
17	Tableau des paramètres du CAC 40, des OAT 10Y et de l'IPD	34
18	Discrétisation de la fonction de score en 2 dimensions entre le CAC 40 et l'IPD	35
19	Discrétisation de la fonction de score en 2 dimensions entre l'OAT 10Y et l'IPD	35
20	Tableau des paramètres des facteurs de risque en entrée du Reverse Stress Test	38
21	Impact cible du Reverse Stress Test sur la variation de valeur du portefeuille d'actifs de la Caisse des Dépôts	40
22	Tableau des sensibilités factices du portefeuille de la Caisse des Dépôts aux facteurs de risque	41
23	Valeurs Récentes des facteurs de risque pour calculer les marginales	43
24	Métrique qualitative du score	43
25	Tableau statistique de la solution du solveur	44
26	Sensibilités du scénario optimal et des métriques associées à un choc sur l'écart-type et/ou la corrélation	45
27	Contributions à différents impacts cibles pour chaque facteur de risque	46

28	p.value des tests statistiques de normalité sur les marginales des trois facteurs de risque	49
29	Illustration graphique du coefficient d'asymétrie	52
30	Valeurs des critères de marginale du CAC 40	54
31	Valeurs des critères de marginale des OAT 10Y	55
32	Valeurs des critères de marginale de l'IPD	55
33	Diagramme quantiles-quantiles entre les marginales du CAC 40 et des OAT 10Y	61
34	Diagramme quantiles-quantiles entre les marginales du CAC 40 et de l'IPD	62
35	Diagramme quantiles-quantiles entre les marginales des OAT 10Y et de l'IPD	62
36	Evolution du coefficient de corrélation linéaire roulant sur 10 ans entre les marginales du CAC 40 et des OAT 10Y	65
37	Evolution du coefficient de corrélation linéaire roulant sur 10 ans entre les marginales du CAC 40 et de l'IPD	65
38	Evolution du coefficient de corrélation linéaire roulant sur 10 ans entre les marginales des OAT 10Y et de l'IPD	66
39	Valeurs des corrélations en fonction de l'historique et de la méthode de calcul retenus	68
40	Caractéristiques des différents Reverse Stress Tests	68
41	Scores d'explicativité des axes de l'ACP sur les 25 variables	76
42	Représentation de la contribution des 25 variables sur les deux axes principaux	77
43	Représentation des 18 modèles sur les deux axes principaux	78
44	Représentation de l'ACP sur 25 variables sur ses deux axes principaux	79
45	Scores d'explicativité des axes de l'ACP sur les 12 variables	81
46	Représentation de l'ACP sur 12 variables sur ses deux axes principaux	82
47	Evolution du SSE sur les 25 variables en fonction du nombre de groupes	84
48	Représentation bidimensionnelle du partitionnement en 2-moyennes	85
49	Représentation bidimensionnelle du partitionnement en 6-moyennes	86
50	Evolution du SSE sur les 12 <i>outputs</i> en fonction du nombre de groupes	87
51	Représentation bidimensionnelle du partitionnement en 4-moyennes	88
52	Synthèse des classifications des 18 modèles sur les 25 variables représentatives	89
53	Synthèse des classifications des 18 modèles sur les 12 <i>outputs</i>	89
54	<i>Outputs</i> des 18 modèles de Reverse Stress Test pour un impact cible à - 3 Md €	90
55	<i>Outputs</i> des 18 modèles de Reverse Stress Test pour un impact cible à - 5 Md €	91
56	<i>Outputs</i> des 18 modèles de Reverse Stress Test pour un impact cible à - 7 Md €	92
57	<i>Outputs</i> des 18 modèles de Reverse Stress Test pour un impact cible à - 10 Md €	92
58	Diagramme quantiles-quantiles empirique entre les marginales du CAC 40 et des OAT 10Y	109
59	Diagrammes quantiles-quantiles théoriques entre les marginales du CAC 40 et des OAT 10Y	109
60	Diagramme quantiles-quantiles empirique entre les marginales du CAC 40 de l'IPD	110
61	Diagrammes quantiles-quantiles théoriques entre les marginales du CAC 40 et de l'IPD	110
62	Diagramme quantiles-quantiles empirique entre les marginales des OAT 10Y de l'IPD	111
63	Diagrammes quantiles-quantiles théoriques entre les marginales des OAT 10Y et de l'IPD	111
64	Estimateurs des marginales du CAC 40	112
65	Estimateurs des marginales des OAT 10Y	112
66	Estimateurs des marginales de l'IPD	113
67	Degrés de liberté de la copule de Student	113
68	Estimateurs des coefficients de corrélation de Pearson	113
69	Portions des évènements tirés d'impacts plus conséquents que -15 Md €	114

70	Variables des 18 modèles sur l'impact à -3 Md €	115
71	Variables des 18 modèles sur l'impact à -5 Md €	115
72	Variables des 18 modèles sur l'impact à -7 Md €	116
73	Variables des 18 modèles sur l'impact à -10 Md €	116

Annexes

- Annexe 1 : *Packages* R utilisés (pages 105 à 108)
- Annexe 2 : Compléments sur le choix de la copule théorique adéquate (pages 109 à 111)
- Annexe 3 : Valeurs de paramètres de calibrage des RST via R (pages 112 à 113)
- Annexe 4 : Jeux de données des 18 modèles (pages 114 à 116)
- Annexe 5 : Coefficients de corrélation de Spearman et de Kendall (pages 117 à 118)
- Annexe 6 : Fonction d'autocorrélation partielle (page 119)
- Annexe 7 : Autres lois marginales alternatives (page 120)

Annexe 1 : *Packages* R et fonctions utilisés

- *Package stats* :
 - "acf" :
Fonction d'autocorrélation.

 - "pacf" :
Fonction partielle d'autocorrélation.

 - "shapiro.test" :
Réalisation du test statistique de Shapiro-Wilk sur un jeu de données.

 - "density" :
Estimateur à noyau d'un jeu de données.

 - "quantile" :
Quantile d'un jeu de données avec fixation de la probabilité en paramètre.

 - "ks.test" :
Test statistique de Kolmogorov-Smirnov entre deux jeux de données.

 - "cor" :
Coefficient de corrélation de Pearson, de Spearman ou de Kendall d'un jeu de données.

 - "rnorm" :
Tirage d'une loi normale à partir des paramètres de la loi.

 - "qnorm" :
Fonction de répartition inverse d'une loi normale à partir d'un quantile et des paramètres de la loi.

 - "rlogis" :
Tirage d'une loi logistique à partir des paramètres de la loi.

 - "qlogis" :
Fonction de répartition inverse d'une loi logistique à partir d'un quantile et des paramètres de la loi.

 - "kmeans" :
Classification des individus par partitionnement en k-moyennes à partir d'un paramètre k (centers) d'un jeu de données de variables.

- *Package tseries* :
 - "jarque.bera.test" :
Réalisation du test statistique de Jarque-Bera sur un jeu de données.

- *Package PerformanceAnalytics* :
 - "skewness" :
Coefficient d'asymétrie d'un jeu de données.
 - "kurtosis" :
Coefficient d'aplatissement d'un jeu de données.
 - "ETL" :
CVaR d'un jeu de données avec fixation du quantile en paramètre.
- *Package copula* :
 - "fitCopula" :
Estimation du degré de liberté d'un jeu de données pour une copule de Student à partir d'un jeu de données.
 - "normalCopula" :
Création d'une copule gaussienne à partir d'un paramètre de corrélation (param) et un paramètre de dimension (dim).
 - "tCopula" :
Création d'une copule de Student à partir d'un paramètre de corrélation (param), un paramètre de dimension (dim) et un paramètre du degré de liberté (df).
 - "gumbelCopula" :
Création d'une copule de Gumbel à partir d'un paramètre de corrélation (param) et un paramètre de dimension (dim).
 - "claytonCopula" :
Création d'une copule de Clayton à partir d'un paramètre de corrélation (param) et un paramètre de dimension (dim).
 - "rCopula" :
Tirage de quantiles à partir d'une copule en paramètre (copula).

- *Package MASS* :
 - "fitdistr" :
Estimation des paramètres de la loi normale ou de la logistique (paramètre dens-fun="normal" ou "logistic") d'un jeu de données.

- *Package sn* :
 - "cp2p" :
Estimation des paramètres de la loi normale asymétrique avec l'espérance, l'écart-type et le coefficient d'asymétrie d'un jeu de données.

 - "rsn" : Tirage d'une loi normale asymétrique avec les paramètres de la loi.

 - "qnorm" :
Fonction de répartition inverse d'une loi normale asymétrique à partir d'un quantile et des paramètres de la loi.

- *Package nigFit* :
 - "nigFit" :
Estimation des paramètres de la loi normale-inverse gaussienne d'un jeu de données.

 - "rnig" :
Tirage d'une loi normale-inverse gaussienne avec les paramètres de la loi.

 - "qnig" :
Fonction de répartition inverse d'une loi normale-inverse gaussienne à partir d'un quantile et des paramètres de la loi.

- *Package SuppDists* :
 - "JohnsonFit" :
Estimation des paramètres de la loi de Johnson's SU d'un jeu de données.

 - "rJohnson" : Tirage d'une loi Johnson's SU avec les paramètres de la loi.

 - "qJohnson" :
Fonction de répartition inverse d'une loi Johnson's SU à partir d'un quantile et des paramètres de la loi.

- *Package* **FactoMineR** :
 - "PCA" :
Analyse en composantes principales d'un jeu de données de variables.

- *Package* **factoextra** :
 - "fviz_cluster" :
Visualisation graphique à partir d'une méthode de classification et d'un jeu de données.

 - "fviz_nbclust" :
Evolution du SSE en fonction du nombre de partitions à partir d'une méthode de classification et d'un jeu de données.

 - "fviz_pca_ind" :
Représentation graphique bidimensionnelle des individus d'un jeu de données à partir de son analyse en composantes principales et des axes à représenter.

 - "fviz_pca_var" :
Représentation graphique bidimensionnelle des variables d'un jeu de données à partir de son analyse en composantes principales et des axes à représenter.

 - "fviz_pca" :
Représentation graphique bidimensionnelle des individus et des variables d'un jeu de données à partir de son analyse en composantes principales et des axes à représenter.

Annexe 2 : Comparaison des diagrammes quantiles-quantiles théoriques et empiriques

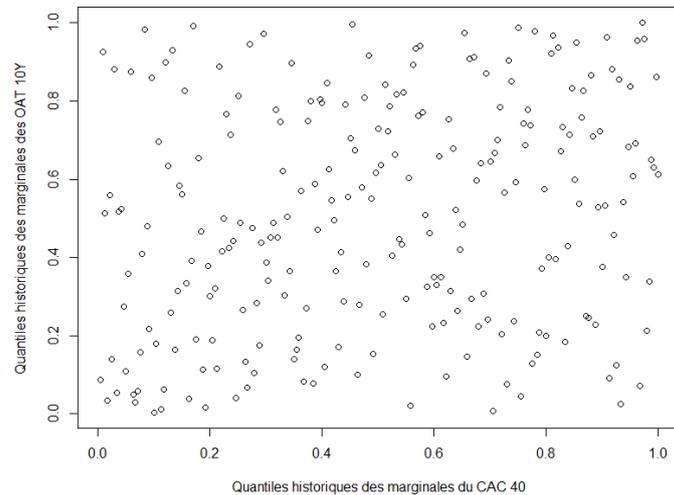


FIGURE 58 – Diagramme quantiles-quantiles empirique entre les marginales du CAC 40 et des OAT 10Y

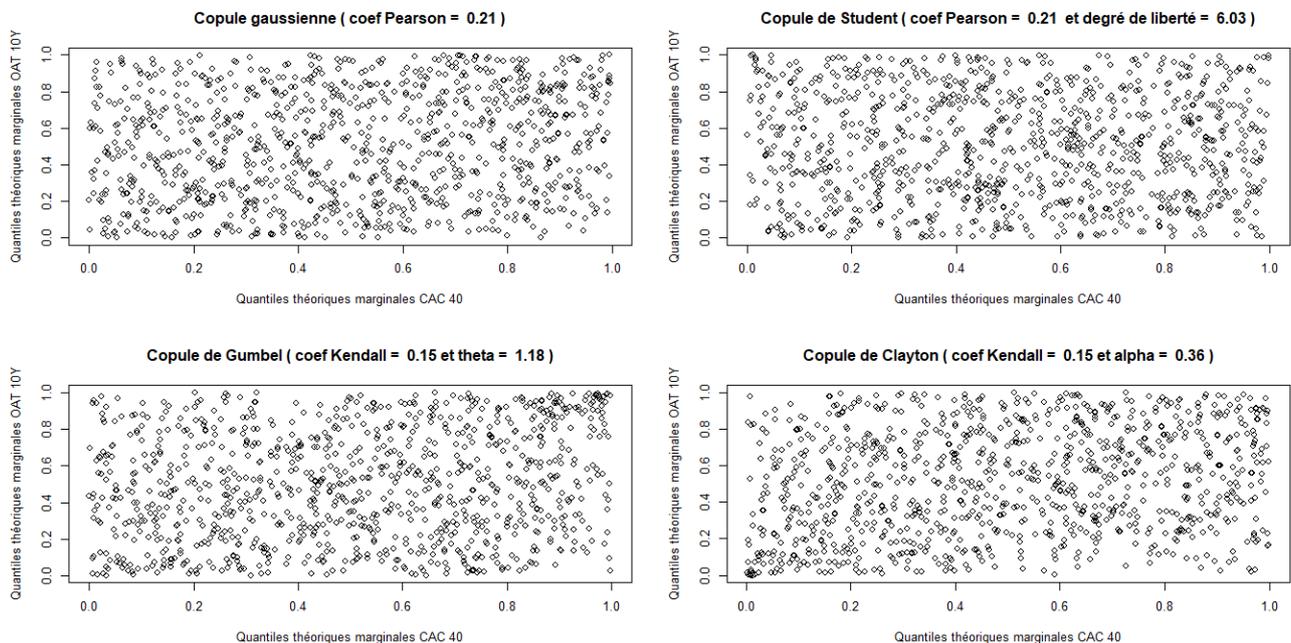


FIGURE 59 – Diagrammes quantiles-quantiles théoriques entre les marginales du CAC 40 et des OAT 10Y

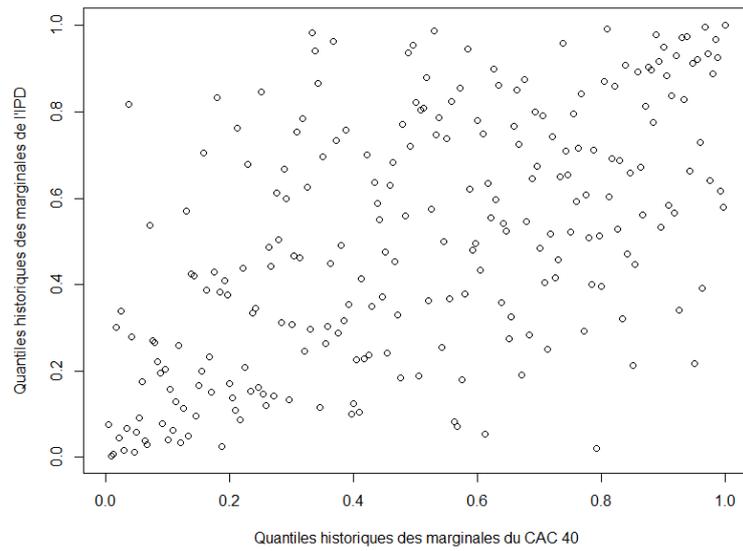


FIGURE 60 – Diagramme quantiles-quantiles empirique entre les marginales du CAC 40 de l'IPD

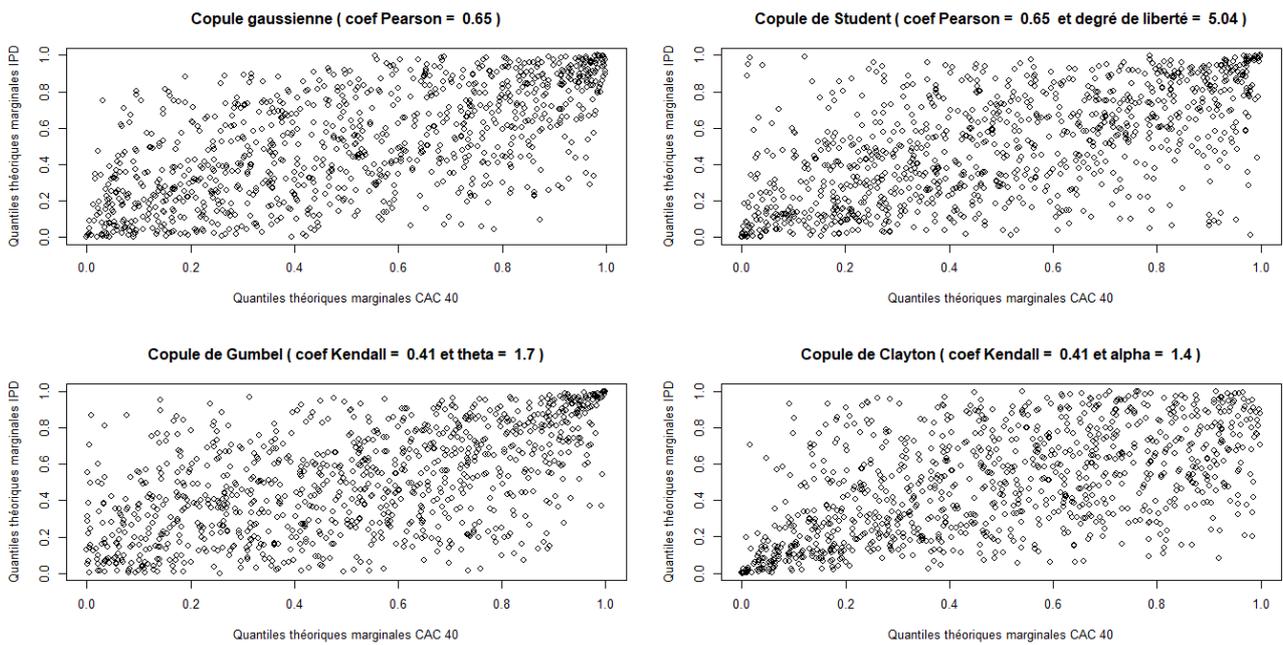


FIGURE 61 – Diagrammes quantiles-quantiles théoriques entre les marginales du CAC 40 et de l'IPD

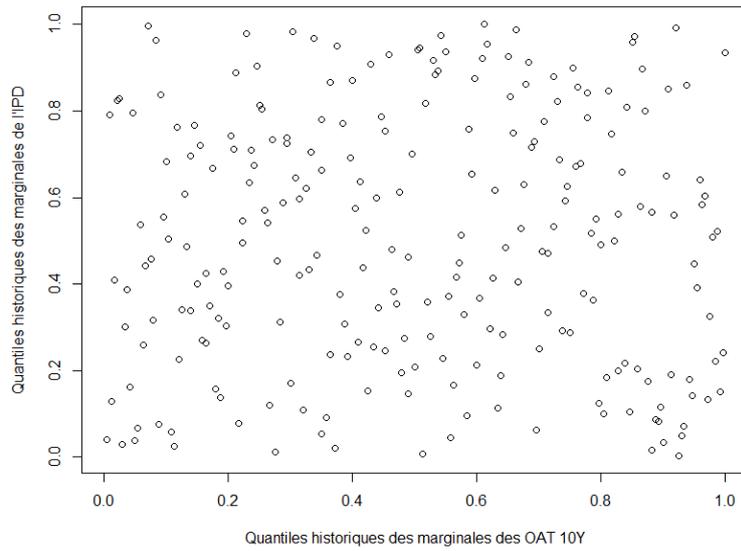


FIGURE 62 – Diagramme quantiles-quantiles empirique entre les marginales des OAT 10Y de l'IPD

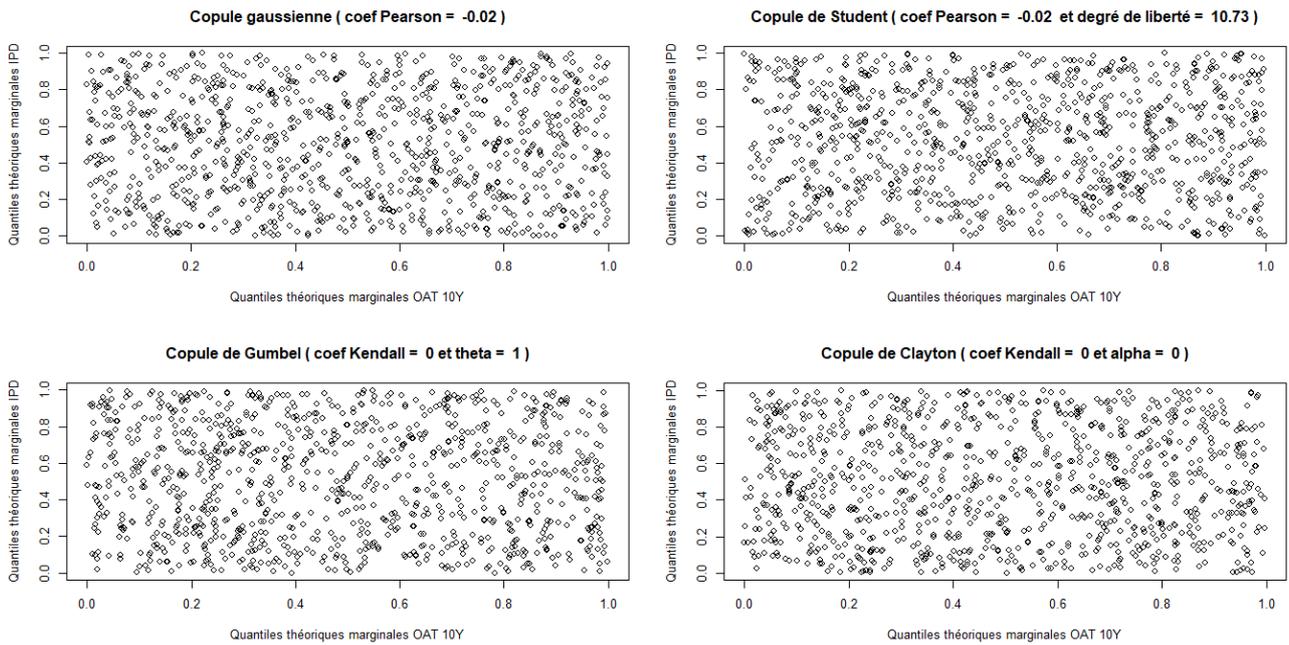


FIGURE 63 – Diagrammes quantiles-quantiles théoriques entre les marginales des OAT 10Y et de l'IPD

Annexe 3 : Valeurs des paramètres de calibrage des Reverse Stress Tests via R

Estimateurs des lois marginales théoriques

Lois marginales	Estimateurs des marginales du CAC 40			
	Paramètres des lois			<i>Drifts</i>
Loi normale	Espérance	Ecart-type		0,0011
	-0,0003	0,0529		
Loi normale asymétrique	Position	Echelle	Forme	0,0011
	0,0574	0,0782	-2,4127	
Estimateur gaussien	Fenêtre			0,0012
	0,1406			

FIGURE 64 – Estimateurs des marginales du CAC 40

Lois marginales	Estimateurs des marginales des OAT 10Y	
	Espérance	Ecart-type
Loi normale	-0,0223	0,1933
Estimateur gaussien	Fenêtre	
	0,0566	

FIGURE 65 – Estimateurs des marginales des OAT 10Y

Lois marginales	Estimateurs des marginales de l'IPD				Drifts
	Paramètres des lois				
Loi normale	Espérance		Ecart-type		0,0025
	0,0023		0,0172		
Loi NIG	Queue	Asymétrie	Echelle	Position	0,0025
	51,4228	-14,6834	0,0122	0,0060	
Estimateur gaussien	Fenêtre				0,0025
	0,0035				

FIGURE 66 – Estimateurs des marginales de l'IPD

Estimateurs des copules théoriques

Copule de Student	
Historique du coefficient de corrélation de Pearson	Estimateur
	Dégré de liberté
20 ans	4,3330
10 ans	9,3981

FIGURE 67 – Degrés de liberté de la copule de Student

Estimateurs des coefficients de corrélation de Pearson

Corrélation de Pearson des marginales		Facteurs de risque 2 à 2		
Historique	Calcul	CAC 40 / OAT 10Y	CAC 40 / IPD	OAT 10Y / IPD
20 ans	Brut	21,32%	64,54%	-2,26%
10 ans	Brut	35,42%	55,16%	11,97%
20 ans	Pondéré	30,59%	55,63%	5,03%

FIGURE 68 – Estimateurs des coefficients de corrélation de Pearson

Annexe 4 : Ensemble des 25 variables des 18 modèles

	Marginales	Copule	Coefficient corrélation de Pearson	Rang -15Md
Modele A	Normales	Gaussienne	20 ans non pondere	0,31%
Modele B	Normales	Gaussienne	10 ans non pondere	0,25%
Modele C	Normales	Gaussienne	20 ans ponderes	0,26%
Modele D	Personnalisées	Gaussienne	20 ans non pondere	0,52%
Modele E	Personnalisées	Gaussienne	10 ans non pondere	0,41%
Modele F	Personnalisées	Gaussienne	20 ans ponderes	0,41%
Modele G	Estimateur gaussien	Gaussienne	20 ans non pondere	0,62%
Modele H	Estimateur gaussien	Gaussienne	10 ans non pondere	0,55%
Modele I	Estimateur gaussien	Gaussienne	20 ans ponderes	0,53%
Modele J	Normales	Student	20 ans non pondere	0,33%
Modele K	Normales	Student	10 ans non pondere	0,26%
Modele L	Normales	Student	20 ans ponderes	0,26%
Modele M	Personnalisées	Student	20 ans non pondere	0,50%
Modele N	Personnalisées	Student	10 ans non pondere	0,38%
Modele O	Personnalisées	Student	20 ans ponderes	0,45%
Modele P	Estimateur gaussien	Student	20 ans non pondere	0,70%
Modele Q	Estimateur gaussien	Student	10 ans non pondere	0,52%
Modele R	Estimateur gaussien	Student	20 ans ponderes	0,58%

FIGURE 69 – Portions des évènements tirés d'impacts plus conséquents que -15 Md €

	Marginales	Copule	Coefficient corrélation de Pearson	CAC moyen -3Md	CAC écart type -3Md	OAT moyen -3Md	OAT écart type -3Md	IPD moyen -3Md	IPD écart type -3Md
Modele A	Normales	Gaussienne	20 ans non pondere	5 105	143	0,37	0,66	199,19	7,85
Modele B	Normales	Gaussienne	10 ans non pondere	5 097	152	0,42	0,64	199,63	8,64
Modele C	Normales	Gaussienne	20 ans ponderes	5 100	156	0,41	0,65	199,46	8,82
Modele D	Personnalisées	Gaussienne	20 ans non pondere	5 102	141	0,39	0,66	199,44	7,76
Modele E	Personnalisées	Gaussienne	10 ans non pondere	5 097	153	0,43	0,63	199,58	8,64
Modele F	Personnalisées	Gaussienne	20 ans ponderes	5 095	154	0,40	0,64	199,79	8,64
Modele G	Estimateur gaussien	Gaussienne	20 ans non pondere	5 093	142	0,37	0,69	200,04	7,86
Modele H	Estimateur gaussien	Gaussienne	10 ans non pondere	5 089	153	0,39	0,67	200,13	8,67
Modele I	Estimateur gaussien	Gaussienne	20 ans ponderes	5 086	153	0,42	0,66	200,28	8,63
Modele J	Normales	Student	20 ans non pondere	5 104	142	0,38	0,65	199,24	7,87
Modele K	Normales	Student	10 ans non pondere	5 100	153	0,43	0,63	199,46	8,74
Modele L	Normales	Student	20 ans ponderes	5 098	157	0,39	0,64	199,60	8,88
Modele M	Personnalisées	Student	20 ans non pondere	5 100	140	0,38	0,66	199,55	7,71
Modele N	Personnalisées	Student	10 ans non pondere	5 095	152	0,42	0,63	199,72	8,54
Modele O	Personnalisées	Student	20 ans ponderes	5 096	153	0,42	0,64	199,81	8,61
Modele P	Estimateur gaussien	Student	20 ans non pondere	5 093	140	0,39	0,69	199,94	7,64
Modele Q	Estimateur gaussien	Student	10 ans non pondere	5 090	151	0,42	0,66	200,12	8,58
Modele R	Estimateur gaussien	Student	20 ans ponderes	5 087	154	0,39	0,67	200,32	8,64

FIGURE 70 – Variables des 18 modèles sur l'impact à -3 Md €

	Marginales	Copule	Coefficient corrélation de Pearson	CAC moyen -5Md	CAC écart type -5Md	OAT moyen -5Md	OAT écart type -5Md	IPD moyen -5Md	IPD écart type -5Md
Modele A	Normales	Gaussienne	20 ans non pondere	4 819	138	0,41	0,66	196,31	7,65
Modele B	Normales	Gaussienne	10 ans non pondere	4 810	151	0,47	0,63	196,75	8,52
Modele C	Normales	Gaussienne	20 ans ponderes	4 811	151	0,45	0,64	196,69	8,58
Modele D	Personnalisées	Gaussienne	20 ans non pondere	4 809	143	0,43	0,66	196,95	7,90
Modele E	Personnalisées	Gaussienne	10 ans non pondere	4 797	152	0,49	0,64	197,46	8,62
Modele F	Personnalisées	Gaussienne	20 ans ponderes	4 804	153	0,46	0,65	197,21	8,60
Modele G	Estimateur gaussien	Gaussienne	20 ans non pondere	4 802	143	0,40	0,68	197,45	7,90
Modele H	Estimateur gaussien	Gaussienne	10 ans non pondere	4 788	154	0,49	0,67	198,10	8,76
Modele I	Estimateur gaussien	Gaussienne	20 ans ponderes	4 793	157	0,46	0,67	197,77	8,87
Modele J	Normales	Student	20 ans non pondere	4 814	141	0,41	0,65	196,56	7,84
Modele K	Normales	Student	10 ans non pondere	4 808	151	0,50	0,64	196,80	8,53
Modele L	Normales	Student	20 ans ponderes	4 805	153	0,46	0,65	197,08	8,61
Modele M	Personnalisées	Student	20 ans non pondere	4 809	138	0,42	0,66	196,93	7,61
Modele N	Personnalisées	Student	10 ans non pondere	4 800	153	0,49	0,65	197,29	8,62
Modele O	Personnalisées	Student	20 ans ponderes	4 800	154	0,46	0,64	197,44	8,65
Modele P	Estimateur gaussien	Student	20 ans non pondere	4 799	142	0,42	0,68	197,55	7,83
Modele Q	Estimateur gaussien	Student	10 ans non pondere	4 789	155	0,47	0,66	198,01	8,82
Modele R	Estimateur gaussien	Student	20 ans ponderes	4 793	153	0,46	0,67	197,89	8,66

FIGURE 71 – Variables des 18 modèles sur l'impact à -5 Md €

	Marginales	Copule	Coefficient corrélation de Pearson	CAC moyen -7Md	CAC écart type -7Md	OAT moyen -7Md	OAT écart type -7Md	IPD moyen -7Md	IPD écart type -7Md
Modele A	Normales	Gaussienne	20 ans non pondere	4 535	134	0,45	0,66	193,29	7,39
Modele B	Normales	Gaussienne	10 ans non pondere	4 524	149	0,55	0,62	193,77	8,35
Modele C	Normales	Gaussienne	20 ans pondere	4 520	150	0,52	0,65	194,08	8,37
Modele D	Personnalisées	Gaussienne	20 ans non pondere	4 520	141	0,46	0,66	194,27	7,86
Modele E	Personnalisées	Gaussienne	10 ans non pondere	4 508	154	0,56	0,64	194,75	8,71
Modele F	Personnalisées	Gaussienne	20 ans pondere	4 511	152	0,52	0,64	194,65	8,63
Modele G	Estimateur gaussien	Gaussienne	20 ans non pondere	4 509	147	0,45	0,70	194,88	8,15
Modele H	Estimateur gaussien	Gaussienne	10 ans non pondere	4 495	153	0,55	0,67	195,42	8,72
Modele I	Estimateur gaussien	Gaussienne	20 ans pondere	4 497	156	0,50	0,68	195,44	8,78
Modele J	Normales	Student	20 ans non pondere	4 533	137	0,45	0,67	193,50	7,60
Modele K	Normales	Student	10 ans non pondere	4 516	151	0,58	0,63	194,21	8,47
Modele L	Normales	Student	20 ans pondere	4 517	150	0,54	0,64	194,18	8,49
Modele M	Personnalisées	Student	20 ans non pondere	4 519	140	0,45	0,65	194,33	7,79
Modele N	Personnalisées	Student	10 ans non pondere	4 508	155	0,57	0,63	194,79	8,77
Modele O	Personnalisées	Student	20 ans pondere	4 507	151	0,52	0,64	194,81	8,52
Modele P	Estimateur gaussien	Student	20 ans non pondere	4 511	143	0,43	0,69	194,75	7,87
Modele Q	Estimateur gaussien	Student	10 ans non pondere	4 499	155	0,56	0,67	195,30	8,77
Modele R	Estimateur gaussien	Student	20 ans pondere	4 500	155	0,50	0,68	195,32	8,76

FIGURE 72 – Variables des 18 modèles sur l'impact à -7 Md €

	Marginales	Copule	Coefficient corrélation de Pearson	CAC moyen -10Md	CAC écart type -10Md	OAT moyen -10Md	OAT écart type -10Md	IPD moyen -10Md	IPD écart type -10Md
Modele A	Normales	Gaussienne	20 ans non pondere	4 114	134	0,51	0,67	188,57	7,32
Modele B	Normales	Gaussienne	10 ans non pondere	4 087	143	0,69	0,66	189,77	8,04
Modele C	Normales	Gaussienne	20 ans pondere	4 093	143	0,60	0,64	189,57	7,98
Modele D	Personnalisées	Gaussienne	20 ans non pondere	4 103	147	0,48	0,65	189,20	8,22
Modele E	Personnalisées	Gaussienne	10 ans non pondere	4 074	156	0,67	0,64	190,60	8,88
Modele F	Personnalisées	Gaussienne	20 ans pondere	4 082	159	0,62	0,66	190,21	9,06
Modele G	Estimateur gaussien	Gaussienne	20 ans non pondere	4 082	148	0,48	0,69	190,47	8,29
Modele H	Estimateur gaussien	Gaussienne	10 ans non pondere	4 065	158	0,67	0,65	191,12	9,03
Modele I	Estimateur gaussien	Gaussienne	20 ans pondere	4 067	165	0,57	0,66	191,17	9,45
Modele J	Normales	Student	20 ans non pondere	4 110	131	0,52	0,68	188,69	7,23
Modele K	Normales	Student	10 ans non pondere	4 092	141	0,69	0,63	189,56	7,95
Modele L	Normales	Student	20 ans pondere	4 093	146	0,63	0,67	189,54	8,20
Modele M	Personnalisées	Student	20 ans non pondere	4 099	141	0,50	0,66	189,27	7,85
Modele N	Personnalisées	Student	10 ans non pondere	4 077	158	0,67	0,65	190,29	9,05
Modele O	Personnalisées	Student	20 ans pondere	4 084	153	0,62	0,64	190,31	8,75
Modele P	Estimateur gaussien	Student	20 ans non pondere	4 085	148	0,48	0,70	190,26	8,33
Modele Q	Estimateur gaussien	Student	10 ans non pondere	4 067	158	0,65	0,68	191,02	8,94
Modele R	Estimateur gaussien	Student	20 ans pondere	4 070	159	0,59	0,70	190,90	8,96

FIGURE 73 – Variables des 18 modèles sur l'impact à -10 Md €

Annexe 5 : Coefficients de corrélation de Spearman et de Kendall

Rho de Spearman

Il est l'analogie du coefficient de Pearson mais appliqué au rang des données. Il permet de mesurer si les variables X et Y de même taille n ont tendance à être inférieures ou supérieures à leur médiane respective plus ou moins au même moment.

Soient $R_X = (r_1^X, \dots, r_n^X)$ et $R_Y = (r_1^Y, \dots, r_n^Y)$ les variables représentant respectivement les rangs de $X = (x_1, \dots, x_n)$ et $Y = (y_1, \dots, y_n)$.

$$\rho_S(X, Y) = \frac{\text{Cov}(R_X, R_Y)}{\sigma_{R_X} \cdot \sigma_{R_Y}} = \frac{\sum_{i=1}^n (r_i^X - \mu_{R_X})(r_i^Y - \mu_{R_Y})}{\sqrt{\sum_{i=1}^n (r_i^X - \mu_{R_X})^2 \times \sum_{i=1}^n (r_i^Y - \mu_{R_Y})^2}}$$

avec μ_{R_X} et μ_{R_Y} pouvant être estimés respectivement par les estimateurs $\widehat{\mu}_{R_X}$ et $\widehat{\mu}_{R_Y}$:

$$\widehat{\mu}_{R_X} = \frac{1}{n} \sum_{i=1}^n r_i^X \quad \text{et} \quad \widehat{\mu}_{R_Y} = \frac{1}{n} \sum_{i=1}^n r_i^Y$$

- Avantages :

1. Ne capte pas l'impact des valeurs aberrantes
2. Stable par transformation monotone
3. Capte les corrélations même si pas affines

- Inconvénients :

1. Nécessite de calculer le rang des données
2. Ne capte pas l'impact des valeurs extrêmes

Tau de Kendall

Tout comme le coefficient de corrélation de Spearman, c'est une mesure du rang des données mais de méthodologie de calcul différente. Il se base sur les paires concordantes et discordantes entre deux jeux de données.

Soient $X = (x_1, \dots, x_n)$ et $Y = (y_1, \dots, y_n)$

$$\tau_K(X, Y) = \frac{(\text{Nb paires concordantes} - \text{Nb Paires discordantes})}{\binom{n}{2}}$$

$$\tau_K(X, Y) = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=i+1}^n \text{Signe}(x_i - x_j) \times \text{Signe}(y_i - y_j)$$

- Avantages :

1. Ne capte pas l'impact des valeurs aberrantes
2. Stable par transformation monotone
3. Capte les corrélations même si pas affines
4. Coefficient utilisé, entre-autres, dans les copules de Gumbel et de Clayton

- Inconvénients :

1. Nécessite de compter le nombre de paires concordantes ou discordantes
2. Ne capte pas l'impact des valeurs extrêmes
3. Difficilement implémentable sur Excel

Annexe 6 : Fonction d'autocorrélation partielle

Fonction d'autocorrélation partielle

La fonction d'autocorrélation partielle complète la fonction d'autocorrélation d'un jeu de données. Elle permet d'observer chaque degré indépendamment des influences des autres. Elle renvoie, elle aussi, des coefficients de corrélation compris entre -1 et 1, pour un pas de temps k fixé.

Soit un jeu de données de taille $N : X_1, \dots, X_N$.
Pour rappel le k^{eme} degré d'autocorrélation vaut :

$$r_k = \frac{\sum_{i=1}^{N-k} (X_i - \bar{X})(X_{i+k} - \bar{X})}{\sum_{i=1}^{N-k} (X_i - \bar{X})^2} \quad \text{avec} \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Le k^{eme} degré d'autocorrélation partielle vaut :

$$r'_k = \frac{\det(R^*(k))}{\det(R(k))}$$

où

$$R(k) = (r_{i-j})_{\substack{i \in \llbracket 1, k \rrbracket \\ j \in \llbracket 1, k \rrbracket}}$$

et $R^*(k)$ la même matrice où la colonne k est remplacée par le vecteur $(r_1 \dots r_k)$

Ce degré d'autocorrélation partielle revient à calculer la fonction d'autocorrélation sur des données neutralisées des effets des $k-1$ degrés d'autocorrélation précédents.

De même que pour la fonction d'autocorrélation, il est important de regarder l'ensemble des valeurs prises pour chaque pas de temps k . A noter que r'_0 est nécessairement égal à 1. Les packages statistiques informatiques permettent parfois d'intégrer un seuil de significativité au delà duquel la valeur est jugée significativement non-nulle.

Annexe 7 : Autres lois marginales alternatives

Loi logistique

La loi logistique est une loi semblable à la loi normale. Elle est aussi symétrique mais son coefficient d'aplatissement est légèrement supérieur à 1. Cette loi est composée de deux paramètres : un paramètre de position (noté μ) et un paramètre d'échelle (noté s).

La fonction de densité est définie telle que :

$\forall x \in \mathbb{R}$

$$f_{\mu,s}(x) = \frac{e^{-\frac{(x-\mu)}{s}}}{s \left(1 + e^{-\frac{(x-\mu)}{s}}\right)^2}$$

avec

$$\mu \in \mathbb{R}$$

$$s \in \mathbb{R}_+^*$$

Les deux paramètres s'estiment trivialement par la méthode des moments.
En effet, pour une variable aléatoire X :

$$\mu = E(X)$$

$$s = \sqrt{\frac{3\text{Var}(X)}{\pi^2}}$$

Loi de Johnson's SU

La loi de Johnson's SU est une loi liée à la loi normale par la fonction inverse du sinus hyperbolique. Cette fonction inverse permet de rendre la fonction asymétrique. Cette loi est composée de 4 paramètres : deux paramètres de position (notés ξ et λ) et deux paramètres de forme (notés γ et δ).

La fonction de densité est définie telle que :

$\forall x \in \mathbb{R}$

$$f_{\gamma,\xi,\delta,\lambda}(x) = \frac{\delta}{\lambda\sqrt{2\pi}} \frac{1}{\sqrt{1 + \left(\frac{x-\xi}{\lambda}\right)^2}} e^{-\frac{1}{2}(\gamma + \delta \sinh^{-1}\left(\frac{x-\xi}{\lambda}\right))}$$

avec

$$\gamma, \xi \in \mathbb{R}$$

et

$$\delta, \lambda \in \mathbb{R}_+^*$$

Les quatre paramètres ne possèdent pas de formules fermées mais sont estimables numériquement.