

Working paper



06.1

Solvabilité 2 n'est pas *risk based*

Sylvestre Frezal

Mars 2016

PARI

PROGRAMME DE RECHERCHE
SUR L'APPRÉHENSION DES RISQUES
ET DES INCERTITUDES

Solvabilité 2 n'est pas *risk based*

Sylvestre Frezal ¹

Abstract :

L'ambition de Solvabilité 2 est d'être *risk based* : l'exigence de capital est réputée être fondée sur une mesure de risque. En nous concentrant sur les risques de souscription non vie, à l'aide de tests de robustesse interne, nous montrons que cette mesure n'est pas fiable. Elle ne permet pas d'apprécier dans l'absolu le risque porté par une branche, le bruit parasitant la mesure étant d'une amplitude comparable en ordre de grandeur à l'amplitude du signal qu'on cherche à capter. Elle ne permet pas non plus, en relatif, d'apprécier la hiérarchie des risques entre branches. Ceci souligne le besoin d'une analyse aussi large que possible des marges d'erreur des mesures de risque portées par Solvabilité 2, et remet en cause la pertinence de ce système pour fonder des décisions de régulation prudentielle ainsi que les choix d'orientation stratégique ou de gestion des risques des organismes d'assurance.

¹ sylvestre.frezal@datastorm.fr,

DataStorm, 40 rue Etienne Dolet, 92 240 Malakoff, France

Chercheur affilié LFA (Laboratoire de Finance Assurance, CREST, Paris, France)

Co-porteur de PARI (programme sur l'Appréhension des Risques et des Incertitudes)

www.chaire-pari.fr

Introduction

La directive Solvabilité 2 (EPC, 2009), qui refonde la réglementation prudentielle des assurances en Europe, se veut *risk based*: les exigences de capital qu'elle impose sont réputées refléter le risque auquel l'organisme est exposé. Ainsi, l'article 26 indique :

Le point de départ, en ce qui concerne l'adéquation des exigences quantitatives dans le secteur de l'assurance, est le «capital de solvabilité requis» (SCR). [...] La formule standard de calcul du capital de solvabilité requis vise à refléter le profil de risque de la plupart des entreprises d'assurance et de réassurance.

En 2011, alors que son chantier d'élaboration a débuté depuis une décennie, alors que quatre études de calibrage des paramètres ont été menées (*Quantitative Impact Studies - QIS 2 à 5*, EIOPA, 2006 à 2010), une fédération professionnelle apprécie « *une avancée positive sur le scénario catastrophe RC spécialisée, calibré à 100% au lieu de 225% proposé au départ* »². Deux interprétations de cette révision sont possibles : soit on avait fait auparavant une erreur de calibrage d'un facteur 2, soit vient de décider, contrairement à la position officielle de la directive, de retenir un calibrage deux fois plus faible que celui qui reflèterait les risques.

A l'inverse, à la même époque, RMS, un logiciel de simulation de catastrophes naturelles utilisé pour alimenter les négociations tarifaires entre assureurs et réassureurs, et comme input dans certains calculs d'exigence de capital, a ajouté dans sa version 11 la simulation d'un phénomène de « clustering », pour capter le fait que les tempêtes puissent arriver par grappe. Un assureur proteste : « *cela double le coût de la pire tempête de période de retour 200 ans* » (celle retenue pour la détermination des exigences de capital dans Solvabilité 2). Cette évolution montre, ici encore, que l'on se fonde sur des grandeurs semblant avoir une marge d'erreur significative.

Si on révisé ainsi ses estimations, considérant demain que tel risque est deux fois plus important qu'on le pensait et que tel autre est deux fois moins important qu'on le croyait, peut-on raisonnablement considérer qu'on capte le risque ? Y répondre suppose d'éclairer deux points : (1) de tels écarts sont-ils significatifs d'un point de vue opérationnel, c'est-à-dire au regard de ce qu'on cherche à capter et (2) ces observations sont-elles singulières ou générales ? L'objet de cet article est de répondre à ces deux questions. Il est possible de le faire de façon très simple.

² Commission économique et financière d'une fédération professionnelle nationale, 7 juin 2011.

La section I présente le contexte et les analyses réalisées dans des champs connexes. La section II présente les métriques développées. La section III présente les données utilisées. La section IV présente les résultats. La section V présente les conséquences à en tirer.

Section I : le cadre

Le contexte

D'un point de vue théorique, le principe d'une réglementation où les exigences de capital sont fondées sur des mesures de risque plutôt que sur des indicateurs de volume est débattu, certains auteurs estimant cela efficace³, d'autres étant plus circonspects⁴. Cet article ne vise pas à se positionner directement dans ce débat sur la structure du design de la réglementation : nous partons du fait établi que, en assurance, l'objectif de Solvabilité 2 est d'être *risk based*.

Solvabilité 1 était considéré comme trop fruste, ne distinguant par exemple en non-vie que deux catégories d'activité et pondérant forfaitairement la plus risquée d'un facteur 1,5 là où, en comparaison, Solvabilité 2 distingue 12 branches et calibre pour chacune d'elle des chocs comportant plus de deux chiffres significatifs. Le schéma ci-dessous synthétise la façon dont, pour les principaux postes du bilan, Solvabilité 2 introduit un calcul *risk-based* des exigences de capital associées, en regard de Solvabilité 1⁵.

Tableau 1 : Solvabilité 2, ou l'introduction d'une exigence de capital *risk based*

Solvabilité 1	Solvabilité 2
Non vie	
<ul style="list-style-type: none"> ▪ Fondée sur le max d'une règle liée au stock et d'une liée au flux ▪ Deux catégories d'activité (écart max : 1,5) ▪ Diversification prise en compte uniquement par effet taille (1 seuil) 	<ul style="list-style-type: none"> ▪ Fondée sur l'agrégation de plusieurs règles liées au stock et au flux ▪ 12 catégories d'activité (écart max : facteur 2 à 2,5⁶) ▪ Effet taille non pris en compte mais la diversification entre branches l'est (matrices de corrélation)
Vie	

³ e.g. Cummins, J., & Phillips, R. (2009) ; Eling, M., & Holzmüller, I. (2008), Weber, R., & Darbellay, A. (2008).

⁴ e.g. Repulo & Martinez-Miera (2014) en banque, Frezal, (2016) en assurance.

⁵ Voir Frezal (2015), pour une présentation plus générale de Solvabilité 2.

⁶ Ce facteur dispersif serait plus important si on prenait en compte les calibrages liés à la réassurance non proportionnelle, mais les réassureurs ont tendance à développer systématiquement un modèle interne. Le facteur indiqué correspond uniquement aux 9 branches d'assurance directe et de réassurance proportionnelle, sur lesquelles se concentre cet article.

<ul style="list-style-type: none"> ▪ Deux catégories selon que capitaux garantis ou non (écart d'un facteur 4), ▪ <i>Add on</i> forfaitaire pour les capitaux décès 	<ul style="list-style-type: none"> ▪ Fondée sur une modélisation police par police (prise en compte des taux garantis, des taux de participation aux bénéfices, des caractéristiques de l'assuré) et l'adossement A/P
Actifs	
<ul style="list-style-type: none"> ▪ Pas d'impact sur les exigences de capital ▪ Règles limitant la concentration 	<ul style="list-style-type: none"> ▪ Exigences de capital dépendant de la catégorie (nature), du rating, de la duration, etc.
Correctifs	
<ul style="list-style-type: none"> ▪ Vie : marge disponible diminuée si adossement A/P inadapté au taux garantis – mais pas d'impact sur l'exigence de capital ▪ Non vie : réassurance prise en compte <i>au prorata</i> de son impact passé 	<ul style="list-style-type: none"> ▪ Simulation traité par traité de l'impact de la réassurance sur les événements extrêmes ▪ Possibilité de développer des calibrages internes (<i>Undertaking specific parameters, USP</i>), ou des modèles internes

Cette ambition revendiquée d'être *risk based* se heurte à trois types de difficultés :

Scientifiques. Ainsi, les modèles de risque n'ont jusqu'à présent pas fait la preuve de leur robustesse. En banque, une étude menée par la FSA a montré que les modèles internes des banques divergeaient d'un facteur 6 dans l'estimation du risque pondéré d'un même portefeuille représentatif d'actifs (Samuel & Harrison, 2011). Danielsson (2002, 2008) a parallèlement montré que, en retenant une palette très usuelle de profondeurs d'historique et un jeu restreint de modèles mathématiques sous-jacents, on observait jusqu'à un facteur deux d'écart entre les différentes estimations possibles d'une VaR quotidienne à 99% d'un titre vanille liquide. En assurance, dans le cadre de Solvabilité 2, ce risque de modèle pourrait être beaucoup plus important que dans ce dernier exemple. De fait, (1) la VaR visée et les données utilisées ne sont pas quotidiennes, mais annuelles : il est donc davantage nécessaire d'extrapoler sur la base d'un échantillon de taille réduite, ce qui limite d'autant la robustesse des statistiques qui peuvent être estimées, et (2) certaines données, par exemple issues de l'interprétation de guidelines labiles, doivent être systématiquement retraitées avant d'être exploitées⁷, ce qui accroît le recours aux jugements d'experts, et élargit donc la palette de résultats finaux non illégitimes.

⁷ Par exemple, la vision d'une charge de sinistres à l'ultime nécessite de retraiter les données de provisionnement fournies par les services d'indemnisation d'un ensemble de données qui biaiseraient les extrapolations en distordant les cadences de règlement : révision des politiques

Interprétatives. En effet, si la définition mathématique de la mesure de risque est univoque (une VaR annuelle à 99,5%), le sens physique associé à cette grandeur, et donc les modalités de calibrage associées, sont plus ambigus (Leroy et Planchet, 2010). Ainsi, on ne sait si cette probabilité doit être interprétée comme « chaque année, une entreprise sur 200 fait faillite », ou bien « tous les deux siècles, le marché fait faillite ». Ces deux formulations représentent des cas polaires qui conduiraient, dans un cas, à ne quantifier que le risque intrinsèque et dans l'autre, que le risque systémique. Ils constituent des vues de l'esprit puisque ni les différents exercices comptables ni les différentes entreprises ne sont indépendantes, mais le fait de privilégier l'une ou l'autre interprétation conduit à une quantification différente, soucieuse dans un cas de plutôt différencier fortement les comportements des différents organismes, et dans l'autre de plutôt prévenir l'impact d'une crise systémique. Ainsi, Leroy et Planchet (op.cit.) interprètent une phase de révision à la hausse des calibrages comme la marque d'un déplacement de l'interprétation vers la lutte contre le risque systémique⁸.

Politiques. Si les calibrages sont réputés être techniques, ils incorporent également des arbitrages politiques. L'exemple le plus flagrant, et faisant régulièrement l'objet de débat, est celui de l'absence de risque postulée des dettes souveraines. Par exemple, la dette publique grecque était, et reste à ce jour, réputée ne pas porter de risque de crédit : sa VaR annuelle à 99,5% est considérée comme nulle. Certes, sur un cas aussi patent, la distorsion politique d'un calibrage qui devrait être technique est régulièrement critiquée. Toutefois, de façon plus générale, le fait que les calibrages soient, au-delà d'une simple mesure économétrique, la résultante d'arbitrages politiques est largement accepté. Ainsi, alors que la probabilité de faillite cible était *déjà* fixée par les textes et que le calibrage était supposé n'être que la simple déclinaison technique de cette probabilité, les OIS n'ont pas servi uniquement à anticiper l'impact de la nouvelle réglementation,

de provisionnement dossier/dossier, IBNR atypiques liés à une vague de sinistres précédant de peu la clôture, retraitement des graves, etc.

⁸ On peut formuler différemment cette ambiguïté d'interprétation concrète de la VaR et du calibrage associé en considérant par exemple la probabilité d'une faillite de l'Etat français. Si on considère qu'il est impossible qu'il fasse défaut dans l'année, mais nullement à exclure qu'il fasse défaut au cours des 200 prochaines années, doit-on considérer une VaR annuelle à 99,5% nulle (cette année, l'organisme ne sera pas affecté par ce risque), ou bien non nulle (sur les 200 prochaines années, le marché risque d'être affecté par un tel événement) ? Si la définition mathématique semble univoque, nous sommes ici dans une impasse lorsqu'il s'agit d'interpréter la grandeur considérée, et partant de relier les mathématiques aux faits, c'est-à-dire de savoir comment la mesurer.

mais également à permettre de réviser les calibrages pour que leurs impacts deviennent acceptables.⁹

Ces trois raisons posent de sérieuses limites à la prétention de Solvabilité 2 d'offrir un cadre « fondé sur les risques », mais les réponses usuellement apportées à de telles critiques qualitatives sont du type « *il ne faut pas jeter le bébé avec l'eau du bain* », « *c'est un proxy, un bon proxy* », « *l'ordre de grandeur est le bon* », ou encore « *ce n'est pas parfait mais c'est mieux que rien* ».

La problématique

Dans ce cadre, mon enjeu est d'apprécier si ces réponses, elles aussi qualitatives, sont fondées, c'est-à-dire d'éclairer la capacité *opérationnelle* de ces calibrages. En d'autres termes, je souhaite évaluer dans quelle mesure ces calibrages sont suffisamment robustes pour fonder une décision dans les deux champs qu'ils prétendent éclairer :

- le champ réglementaire, puisque ces mesures impactent directement le dénominateur du ratio de solvabilité. L'enjeu y est double : pour les régulateurs souhaitant garantir un niveau de solvabilité minimum équivalent et pour les assureurs ayant besoin d'un *level playing field*, la hiérarchie et la proportionnalité des prises de risques de différentes compagnies doivent être respectées par ces mesures ;
- le champ du pilotage, puisque ces mesures impactent directement le dénominateur du ratio de rentabilité. L'enjeu y est également double : pour les assureurs comme pour les investisseurs souhaitant arbitrer entre différentes lignes d'activité, la hiérarchie et la proportionnalité des risques de chaque branche ou produit, et partant leur rentabilité ajustée du risque, doivent être reflétées par ces mesures.

Danielsson (op.cit.) dans le domaine des VaR bancaires, ou Planchet et Kamega (2013) et El Karoui et al. (2015) pour le provisionnement vie (donc, par complémentarité, la marge disponible) et les fonds propres économiques, se sont intéressés à la quantification des marges d'erreur dans l'absolu *via* le test de

⁹ Ainsi, la Commission Européenne demande explicitement à l'EIOPA de se prononcer sur le caractère approprié des calibrages à l'occasion des études d'*impact* (e.g. « *The analysis performed for the report regarding the appropriateness of the calibration of QIS4 [...]* ». - CfA.27, EC 2008, MARKT/2504/08, *Call for Advice from CEIOPS*). Qu'on se soucie des impacts d'une exigence ayant des conséquences massives sur le financement de l'économie et la stabilité d'un marché avant de l'entériner se comprend tout à fait, est sain et nécessaire, mais il a été choisi, plutôt que de réviser la probabilité de faillite cible, de modifier directement au cas par cas ces calibrages. Ceux-ci, incorporant des arbitrages, ne sont donc pas, contrairement au discours officiel, les simples mesures économétriques des risques.

différentes hypothèses théoriques alternatives légitimes en entrée du modèle. Je développe une méthodologie alternative, simple et robuste, reposant sur quelques ratios sans nécessiter de test d'hypothèse, et fondée directement sur les données de sortie de modèles officiels, en me fondant sur l'observation et la comparaison de différents jeux de calibrages élaborés par les régulateurs ou un acteur de l'industrie.

Afin d'éclairer dans quelle mesure on peut, dans le cadre de prises de décision opérationnelles, se fonder sur les exigences de capital de Solvabilité 2 comme mesure de risque,

- j'évalue à partir de l'ensemble des calibrages obtenus les amplitudes relatives du signal et du bruit afin d'apprécier la capacité de ces mesures à segmenter l'ampleur des risques portés,
- puis j'évalue la stabilité de la hiérarchie des risques afin d'apprécier la fiabilité d'un arbitrage reposant sur ces mesures.

Cette méthodologie rustique ne permettrait pas de valider la pertinence opérationnelle des calibrages, mais elle permet, le cas échéant, de l'invalidier. En l'occurrence, elle nous permettra de conclure.

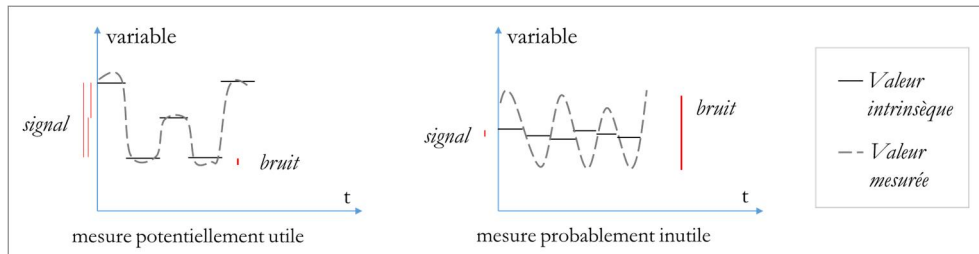
Section II : les métriques

Afin d'apprécier la qualité des mesures dans deux de leurs dimensions : leur capacité de dispersion (vision cardinale), et leur capacité de classification (vision ordinale), je considère deux types d'indicateurs.

Dispersion. L'objectif est de mesurer l'amplitude relative du bruit et du signal. Comme dans l'expression *signal-noise ratio* utilisée par Sims (2003) ou Alesina et Tabellini (2007), le signal correspond à la grandeur caractéristique de l'amplitude de l'information qu'une mesure cherche à capter, et le bruit à la grandeur caractéristique des perturbations qui viennent flouter cette information.

L'appréciation relative du bruit et du signal doit me permettre d'apprécier si la mesure fournie une information opérationnellement utilisable. Prenons un exemple : supposons que nous souhaitions utiliser un altimètre pour piloter un avion. Si, au début du XX^{ème} siècle, du fait des caractéristiques de l'avion, il ne monte pas à plus de 1 000 m et que l'altimètre est précis à 10 000 m près, alors l'amplitude du signal recherché est faible par rapport au bruit parasite et l'altimètre n'a pas d'utilité opérationnelle. Inversement, si l'avion peut monter à 10 000 m et que l'altimètre est précis à 1 000 m près, alors il sera utile dans certaines circonstances (non pour se poser, mais *a minima* pour éviter une montagne).

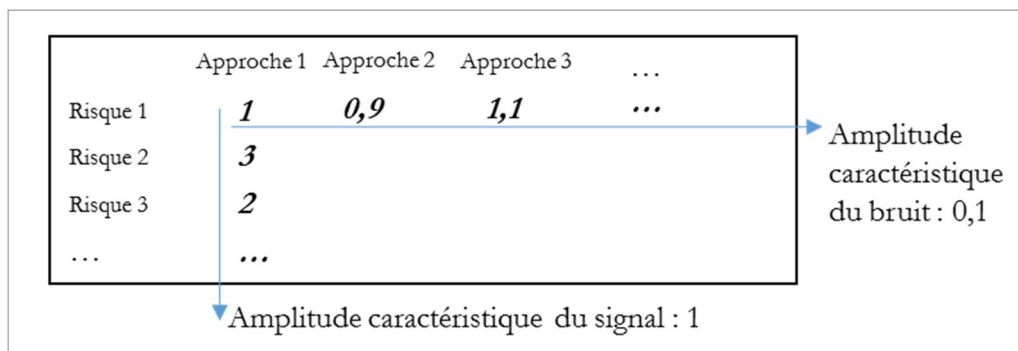
Figure 1 :
visualisation signal / bruit de la pertinence opérationnelle d'une mesure



Pour apprécier la dispersion des mesures relative au bruit et celle relative au signal, je compare :

- un indicateur de *bruit* reflétant la dispersion des calibrages alternatifs d'une même grandeur. Il s'agira d'une grandeur caractéristique de l'écart entre les évaluations d'un *risque donné* réalisées par *différentes approches*,
- à un indicateur de *signal brut* captant l'écart de calibrage entre différentes grandeurs. Il s'agira d'une grandeur caractéristique de l'écart entre l'évaluation par *une approche donnée* de *différents risques*.

Figure 2 : Illustration du signal et du bruit



Nous en déduisons un ratio signal sur bruit (SSB) suivant les modalités présentées en annexes 1 (exemple, indicateur et intuition) et 2 (formalisation).

Classification. Dans son dessein d'amélioration de la gestion des risques et des processus décisionnels, Solvabilité 2 souligne l'importance d'une « *capacité à classer*

les risques »¹⁰. De fait, si les risques sont dans l'absolu mal appréciés mais que le positionnement relatif des risques est correctement mesuré, alors le niveau de sécurité global offert par le système prudentiel ne sera certes pas celui visé mais l'allocation du capital (entre branches, entre compagnies, etc.) sera tout de même plus efficace qu'en l'absence de quantification. En revanche, si la hiérarchie des risques n'est pas captée par la mesure utilisée, alors la quantification n'améliorera pas les arbitrages, voire les dégradera en imposant la substitution de quantifications erronées aux analyses alternatives potentiellement existantes.

Je m'intéresse donc également à la stabilité de la hiérarchie des n risques tels qu'appréhendée par les différents calibrages. L'objectif est de déterminer dans quelle mesure un changement d'approche modifie la perception d'un risque comme plus risqué (ou moins risqué) qu'un autre risque. Je compare ainsi, deux à deux, un ensemble de paires de risques afin de déterminer leur positionnement relatif dans une approche donnée ; puis, je regarde, pour un ensemble d'approches, si cette comparaison est stable en mesurant la proportion de couples dont le positionnement relatif est indépendant de l'approche retenue.

Cette mesure est formalisée en annexe 3.

Caveat. Notons que, par rapport aux analyses de l'impact de choix d'hypothèses sur des résultats de calibrage, telles que celles proposées dans les études susmentionnées, notre méthodologie elle présente l'inconvénient de ne mesurer que la crédibilité interne du système de mesure, et non sa validité¹¹ : en cas de réussite à ces tests (amplitude du bruit faible par rapport à celle du signal, stabilité de la hiérarchie des risques), je n'aurai pas montré le bien-fondé des calibrages et ne pourrai pas conclure. En revanche, en cas d'échec à ce test, j'aurai montré que le système n'est pas crédible.

Section III : les données

Pour la mesure d'un risque donné, quelles sont les différents types d'approche envisageables ? Nous retiendrons deux catégories de d'approches.

¹⁰ « Indépendamment de la méthode de calcul retenue, la capacité du modèle interne [il s'agit d'un article sur les modèles internes] à classer les risques est suffisante pour garantir qu'il est largement utilisé et qu'il joue un rôle important dans le système de gouvernance de l'entreprise d'assurance ou de réassurance concernée, et notamment dans son système de gestion des risques et ses processus décisionnels, ainsi que dans l'allocation de son capital ». (EPC, 2009), Article 121.4

¹¹ En revanche, elle présente l'avantage de caractériser de façon tangible, très intuitive, l'apport opérationnel des mesures de risque pour prendre une décision relative aux risques.

Les différents QIS. Solvabilité 2 a fait l'objet d'une série d'études quantitatives d'impact (QIS) destinées à mesurer la capacité des organismes à réaliser les calculs, à apprécier l'impact de la mise en œuvre de cette réglementation sur leur besoin de financement, et d'affiner les calibrages. Quatre de ces études¹² ont ainsi proposé des calibrages de VaR à 99,5% à un an d'une série de chocs, et les mesures d'implémentation de niveau 2 ont conduit à un dernier jeu de calibrages. Nous disposons donc d'un ensemble de cinq jeux de mesures de risques, représentant chacun une « approche possible », jugée légitime à la date de sa publication, proposée par l'EIOPA, le superviseur européen.

Comme évoqué, plusieurs sources de diversité sont susceptibles de faire varier les calibrages d'un exercice à l'autre :

- une évolution d'un arbitrage politique, liée par exemple à la pression d'une industrie, à la volonté d'un Etat membre de ne pas défavoriser « ses » acteurs, à la crainte des conséquences d'un calibrage précédent sur un marché, etc. Ces appréciations politiques ont été favorisées par le fait que chaque QIS apportait de nouvelles informations sur l'impact en termes d'exigences de capital des calibrages précédents et, potentiellement, par le fait que la dégradation parallèle de l'environnement financier ajoutait une tension sur la situation capitaliste des entreprises et la capacité à lever des fonds si la nouvelle réglementation venait à l'exiger.
- une évolution technique, soit (i) liée aux données elles-mêmes, par exemple à l'occasion d'un raffinement des données et partant des classes d'analyses, d'un accroissement de la profondeur d'historique ou au contraire d'une évolution des produits qui rend caduque un calibrage passé, soit (ii) liée à un arbitrage d'expert, par exemple dans le cadre de l'exploitation des données, du fait de l'utilisation d'un nouveau type de modèle (e.g. changement de fonction de distribution sous-jacente), ou simplement au changement du responsable du calcul¹³.

La raison de cette volatilité, qu'elle soit de nature technique ou politique, nous importe peu. Si elle est politique, elle vient de toute évidence distordre la crédibilité technique pure de l'analyse. Si elle est technique, il ne semble guère soutenable de considérer que les données des assureurs et la science actuarielle

¹² Les QIS 2, 3, 4 et 5. Le QIS1, focalisé sur l'évaluation du bilan, ne comportait pas de test des exigences de capital, ni de calibrage des exigences associées.

¹³ A titre d'exemple concernant ce dernier point, un manager connaissant son équipe peut généralement savoir, en observant quels points « aberrants » ont été exclus d'une analyse statistique, qui parmi l'ensemble des personnes de son équipe a mené l'analyse.

ont fait un saut qualitatif entre 2006 (date du QIS 2) et 2009 (date du QIS 5) ou 2014 (publication des mesures d'implémentation) et que 2006 correspondrait à une date où l'assurance n'avait pas de données ni ne savait les exploiter, ou bien que 2009 ou 2014 correspondrait à une date représentant l'aboutissement de la science actuarielle et des données disponibles.

Deux arguments pourraient cependant expliquer une dispersion qui serait alors susceptible d'être interprétée à tort comme du bruit :

- d'une part, le fait que les risques évoluent, et qu'il est alors légitime que les calibrages s'adaptent. Cet argument ne peut être écarté par les analyses que nous menons. Afin de limiter l'impact d'un tel risque, nous nous intéresserons uniquement aux risques de souscription, excluant les risques de crédit et de marché susceptibles d'avoir évolué à l'occasion de la crise financière. Qui plus est, s'il est pertinent et que l'évolution passée des calibrages reflète l'évolution du risque sous-jacent, alors leur cristallisation générera dans le futur un désadossement entre la mesure et le risque d'un ordre de grandeur comparable. Or, hormis une clause de revoyure en 2018, il n'est pas prévu de réviser régulièrement les calibrages de la directive.
- d'autre part, certains des calibrages des premiers QIS ont été fondés sur la base des données de certains marchés nationaux de l'Union Européenne (CEIOPS, 2010b, p189), et l'élargissement de la base de données peut expliquer les variations. Là encore, la pertinence de cet argument ne peut être écartée par les analyses que nous menons. Toutefois, si là est la source des variations, cela signifie que l'hétérogénéité des risques sous-jacents entre marchés nationaux est telle qu'imposer des mesures uniformes génère un désadossement entre le risque sous-jacent et sa mesure d'un ordre de grandeur comparable à celui du bruit mesuré. Il s'agit donc effectivement d'un bruit lorsqu'on cherche à capter le profil de risque d'une entreprise donnée, ce qui est l'objectif et la revendication de Solvabilité 2.

Ainsi, dans le cadre de la directive Solvabilité 2 dont les calibrages seront fixes et homogènes, les deux arguments évoqués ne sont pas de nature à remettre en cause le fait de considérer la dispersion entre QIS (et avec les mesures d'implémentation de niveau 2) comme relevant du bruit, d'un écart indu entre l'exigence de capital d'une compagnie et son risque propre, et non comme un signal sur le risque pris.

Périmètre de risques analysés.

L'une des difficultés auxquelles cette approche empirique nous expose est l'existence d'un conformisme (phénomène de mimétisme via les « standards de marché » véhiculés par les consultants, conformisme réglementaire¹⁴) qui biaise l'analyse en restreignant le champ des calibrages par rapport à l'ensemble de ceux qui seraient « scientifiquement légitimes ». La dispersion observée est ainsi sous-estimée, générant une convergence factice des résultats et favorisant ainsi une image de crédibilité interne du système. Le fait de ne pas nous intéresser aux calibrages des risques de marché, où l'usage de la VaR est répandu de longue date et où les comportements ont donc davantage eu l'occasion d'être standardisés, limite ce biais.

Au sein des risques de souscription, ceux sur lesquels nous disposons d'un nombre conséquent de risques standardisés, nous permettant donc de mener notre étude comparative, est celui des chocs de primes et de réserve en non vie. Ces deux risques cumulés représentent 40% de l'exigence de capital globale des organismes d'assurance non vie¹⁵.

12 branches sont ainsi identifiées, chacune ayant un calibrage associé, nous retenons les 9 caractérisant l'assurance directe et la réassurance proportionnelle. Trois autres lignes concernent la réassurance non proportionnelle, mais les réassureurs ayant abondamment communiqué sur le caractère inadapté des calibrages Solvabilité 2 à leur activité et sur leur développement par principe de modèle interne, ces calibrages ne devraient pas être *de facto* utilisés par l'industrie, aussi nous ne retiendrons pas les calibrages associés à ces activités dans notre échantillon.

Modèle interne. Les QIS ont permis de déterminer les calibrages de la formule standard, mais les entreprises d'assurance qui le souhaitent ont la possibilité d'utiliser un modèle interne en lieu et place de cette formule. Nous disposons des calibrages mesurés en interne, à différentes dates, par un groupe d'assurance, pour chacun des chocs considérés. Il s'agit de la filiale française d'un des quatre

¹⁴ Au niveau des principes, la directive souligne explicitement la nécessité qu'en termes de calibrage, les « spécifications correspondent bien aux pratiques de marché généralement admises » (EPC, 2009, Art. 122.4). A titre d'exemple en termes de mise en oeuvre, le *QIS 3 calibration paper* (CEIOPS, 2007b, p12) explique qu'on retiendra pour estimer la VaR à 99,5% des risques de prime non vie une valeur de 3σ , « *assuming a lognormal distribution of the underlying risk* ». Cette hypothèse s'applique quelle que soit la branche considérée et tend de ce fait à guider et figer les choix de loi retenus parmi l'ensemble des extrapolations possibles.

¹⁵ ACPR, 2011, p16. Les risques de souscription représentent 63,2% de l'exigence de capital d'un organisme d'assurance non vie. Nous en déduisons les 36% correspondant aux risques catastrophe, qu'il serait plus difficile d'inclure dans cette étude puisque leur mesure est contingente à la répartition géographique du portefeuille assuré.

groupes d'assurance européens de taille mondiale, développant un modèle interne et présent en France sur l'ensemble des branches.

Nous disposons ici de données exploitables

- pour le risque de provision¹⁶,
- pour les cinq principales branches (Auto RC, Auto matériel, Marine-Aviation-Transport, dommages, et RC),
- sur six années, de la clôture 2009 à la clôture 2014. L'organisme d'assurance avait avant 2009 développé un modèle depuis plusieurs années. Les méthodologies ont ainsi eu le temps de cristalliser avant la plage de données dont nous disposons, ce qui est susceptible de créer une stabilité artificielle. Nous devrions donc sous-estimer la marge d'erreur, de façon plus marquée qu'à partir des données des QIS.

Ici encore, les sources de variations, au cours du temps d'une part et vis-à-vis de la formule standard d'autre part, peuvent être dues :

- soit à des raisons politiques (e.g. étayer un choix stratégique ou piloter une communication, interne ou vis-à-vis de la holding par exemple),
- soit à des raisons techniques (spécificités du portefeuille et des données associées, des jugements d'expert, etc.).

Ici encore, les raisons de la dispersion de paramètres réputés être fondés et pertinents techniquement importe peu. Les mêmes arguments pour justifier techniquement d'une éventuelle dispersion comme d'un signal pourraient être soulevés. Les mêmes réponses leurs sont opposables et permettent de classer en bruit la dispersion des observations.

Section IV : les résultats

Formule standard - capacité de dispersion absolue

Le tableau 3 (resp. 5) présente, pour le risque de prime (resp. provision), l'ensemble des calibrages déterminés par l'EIOPA à l'occasion de chacune des études d'impacts, ainsi que la proposition finale retenue pour l'entrée en vigueur de la directive.

¹⁶ Pour le risque de prime, la méthodologie diffère de celle de la formule standard en ce qu'elle se focalise directement sur le quantile extrême et non sur l'écart entre le quantile extrême et le *best estimate*. L'interprétation de la mesure de risque ne correspond donc pas à une dispersion et les comparaisons entre branches ne sont plus directement valables, puisque chacune à une espérance différente. Nous ne disposons pas de ces espérances pour nous ramener à des indicateurs de type « volatilité » comparables.

Les deux dernières colonnes présentent, pour chaque branche, la valeur des indicateurs de dispersion entre QIS. Elles synthétisent donc, pour chaque branche, une estimation de l'amplitude du bruit parasitant la mesure. En observant le second indicateur, on notera par exemple que, pour le risque de prime comme pour le risque de provision, les deux tiers des branches voient leur estimation du risque associé évoluer d'un facteur deux ou plus selon la mesure.

Les deux dernières lignes présentent, pour chaque QIS, la valeur des indicateurs de dispersion entre branche. Elles synthétisent donc un ordre de grandeur de la capacité de dispersion de la mesure, c'est-à-dire l'amplitude de l'information fournie.

Le tableau 4 (resp. 6) synthétise ces résultats, en présentant les moyennes de ces indicateurs de bruit (sur l'ensemble des branches) et de signal (sur l'ensemble des QIS et mesures d'implémentation). Les éléments suivants se dégagent :

- Sur la base du premier indicateur de dispersion (écart-type), le bruit est d'un ordre de grandeur comparable au signal. Pour la mesure du risque de prime (3,2% vs. 4,1%), l'amplitude du signal ajusté sur le bruit est inférieure à 30%. Pour la mesure du risque de provision (3,4% vs. 3,7%), ce ratio signal sur bruit devient totalement négligeable.
- Les résultats obtenus à partir du second indicateur de dispersion (ratio max-min/min), sans être aussi marqués, tendent à confirmer cette première observation. Ainsi, pour le risque de prime (amplitude du bruit de 1,4 vs. 2,7 pour le signal brut) le ratio signal sur bruit est de 1 et, pour le risque de provision (facteur 0,9 vs. 1,3), d'un demi.

Tableau 3 : calibrages associés au risque de prime et indicateurs de signal brut et de bruit associés

	2006	2007	2008	2010	2014	indicateurs de bruit	
	QIS2	QIS3	QIS4	QIS5	Preparatory phase	stdev	max-min/min
1 Motor, third party liability	5,0%	7,5%	9,0%	10,0%	10,0%	2,1%	1,0
2 Motor, other classes	12,5%	3,0%	9,0%	7,0%	8,0%	3,4%	3,2
3 Marine, aviation and transport	7,5%	5,0%	12,5%	17,0%	15,0%	5,0%	2,4
4 Fire and other damage to property	15,0%	10,0%	10,0%	10,0%	8,0%	2,6%	0,9
5 Third-party liability	10,0%	10,0%	12,5%	15,0%	14,0%	2,3%	0,5
6 Credit and suretyship	25,0%	12,5%	15,0%	21,5%	12,0%	5,8%	1,1
7 Legal expenses	10,0%	10,0%	5,0%	6,5%	7,0%	2,2%	1,0
8 Assistance	15,0%	10,0%	7,5%	5,0%	9,0%	3,7%	2,0
9 Miscellaneous non-life insurance	10,0%	12,5%	11,0%	13,0%	13,0%	1,3%	0,3
indicateurs de signal brut	stdev	5,8%	3,2%	3,0%	5,4%	2,9%	
	max-min/min	4,0	3,2	2,0	3,3	1,1	

Tableau 4 : indicateurs moyens de bruit et de signal brut pour le calibrage du risque de primes et ratio ajusté signal sur bruit

tous les calibrages	avec QIS 2		hors QIS2	sans QIS 2	
	stdev	max/min		stdev	max/min
signal brut	4,1%	2,7		3,6%	2,4
bruit	3,2%	1,4		2,4%	0,9
SSBa	28%	99%		50%	156%

Tableau 5 : calibrages associés au risque de provision et indicateurs de signal brut et de bruit associés

	2006	2007	2008	2010	2014	indicateurs de bruit	
	QIS2	QIS3	QIS4	QIS5	Preparatory phase	stdev	max-min/min
1 Motor, third party liability	15,0%	15,0%	12,0%	9,5%	9,0%	2,9%	0,7
2 Motor, other classes	15,0%	7,5%	7,0%	10,0%	8,0%	3,3%	1,1
3 Marine, aviation and transport	7,5%	15,0%	10,0%	14,0%	11,0%	3,0%	1,0
4 Fire and other damage to property	15,0%	12,5%	10,0%	11,0%	10,0%	2,1%	0,5
5 Third-party liability	10,0%	7,5%	15,0%	11,0%	11,0%	2,7%	1,0
6 Credit and suretyship	20,0%	15,0%	15,0%	19,0%	19,0%	2,4%	0,3
7 Legal expenses	20,0%	10,0%	10,0%	9,0%	12,0%	4,5%	1,2
8 Assistance	10,0%	15,0%	10,0%	11,0%	20,0%	4,3%	1,0
9 Miscellaneous non-life insurance	20,0%	10,0%	10,0%	15,0%	20,0%	5,0%	1,0
indicateurs de signal brut	stdev	4,8%	3,3%	2,6%	3,2%	4,9%	
	max-min/min	1,7	1,0	1,1	1,1	1,5	

Tableau 6 : indicateurs moyens de bruit et de signal brut pour le calibrage du risque de provision et ratio ajusté signal sur bruit

tous les calibrages		avec QIS 2		hors QIS2	
		stdev	max/min	stdev	max/min
signal brut		3,7%	1,3	3,5%	1,2
bruit		3,4%	0,9	2,6%	0,6
SSBa		12%	47%	33%	96%

L'EIOPA ayant particulièrement insisté, lors de la publication des calibrages du QIS 2, sur le fait qu'il s'agissait de calibrages préliminaires¹⁷, nous avons mené la même étude en retirant le QIS 2 du champ de l'analyse (troisième et quatrième colonnes des tableaux 4 et 6). Si l'amplitude du bruit par rapport à celle du signal brut n'est plus aussi élevée, elle reste forte, avec un ratio signal sur bruit presque toujours inférieur à 1. Qui plus est, nous pensons qu'il est plus pertinent d'inclure le QIS 2 dans l'analyse. D'une part parce qu'au-delà du QIS 2, chacun des exercices a été accompagné par l'EIOPA de précautions discursives quant au caractère provisoire des calibrages, ce qui, indépendamment de tout aspect technique, était sans doute nécessaire pour des raisons politico-juridiques (éviter de se lier les mains et d'outrepasser formellement son rôle d'expert). D'autre part parce que, comme nous l'avons souligné dans la partie précédente, si les calibrages de certains chocs des premiers QIS se fondaient sur des jeux partiels de données (marché français, britannique ou allemand selon les branches) et que l'extension de la base de données à l'ensemble du marché européen modifie significativement les calibrages, alors cela signifie que l'écart entre les données globales utilisées *in fine* et les données nationales constituent un bruit venant distordre l'appréciation du risque porté par chaque entreprise (dont les bilans restent nationaux).

A ce stade, il apparaît donc douteux de considérer que les mesures de risque proposées par Solvabilité 2 offrent une capacité de dispersion significativement supérieure à celle d'un aléa parasite (quelle que soit l'origine de cet aléa : contrainte technique ou choix politique).

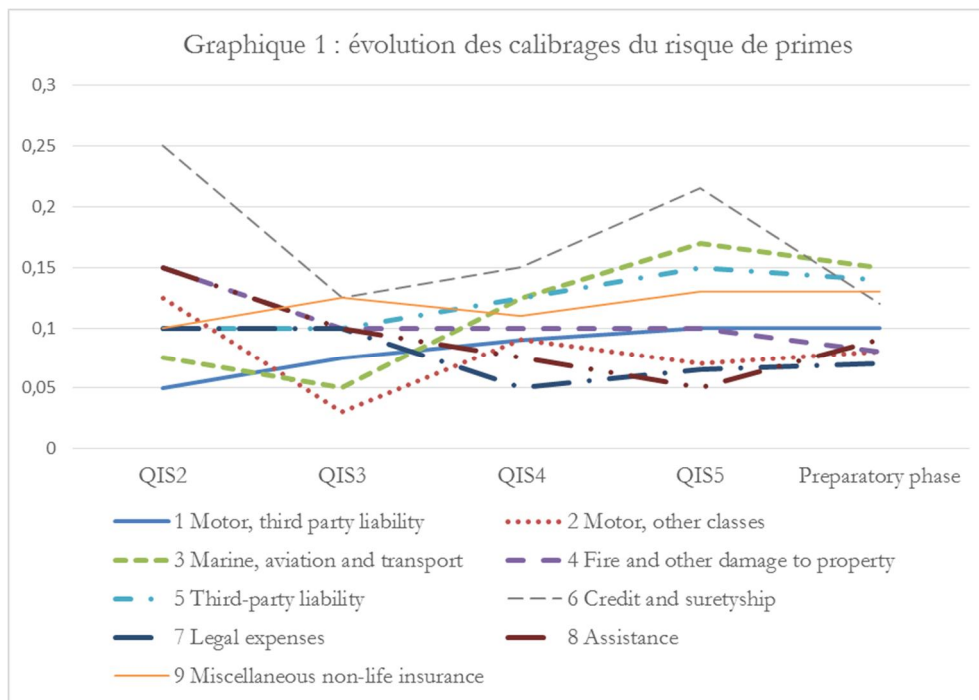
Toutefois, si la dispersion moyenne (écart-type) est totalement noyée dans le bruit, la dispersion extrême (ratio max -min/min) lui est légèrement supérieure. Ainsi, on pourrait espérer que ces calibrages permettent de dissocier les branches les plus risquées des moins risquées. Qui plus est, il est à ce stade possible d'espérer que, à défaut de caractériser les risques dans l'absolu, les mesures de risques ont une portée relative : si les mesures de risques de toutes les branches tendaient à évoluer de façon homothétique, obéissant à des enjeux politiques de capitalisation globale de l'industrie, susceptibles d'impacter les différentes branches de façon homogène, alors les indicateurs que nous avons retenus ne les capteraient pas. Or, dans un tel cas de figure, la capacité des mesures de risque à

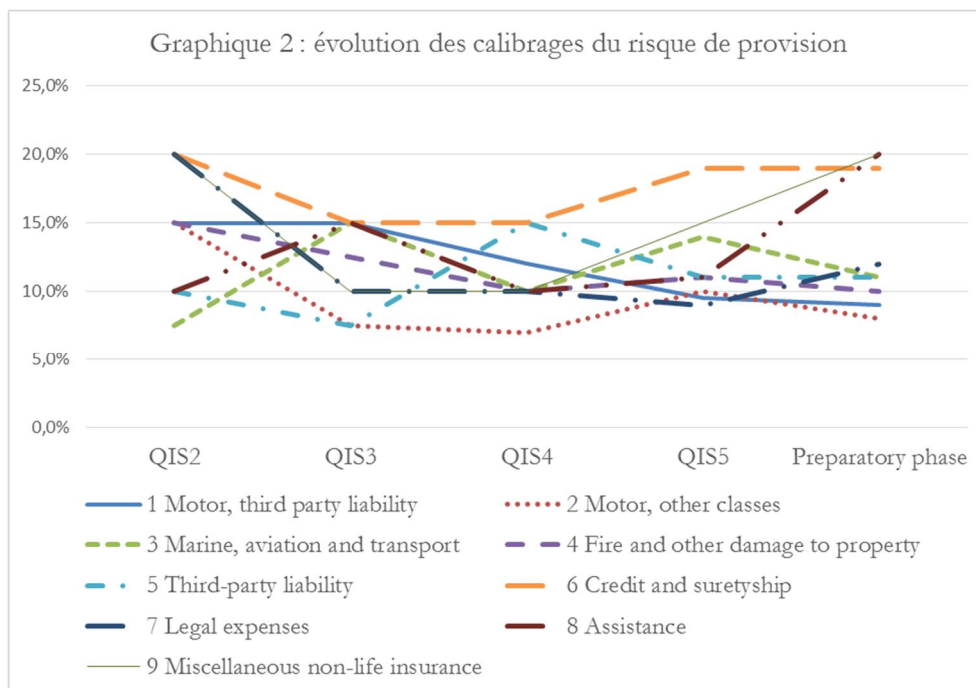
¹⁷ " *The parameters and assumptions used reflect only an initial and very tentative calibration. Parameters and shocks have been selected with the aim of approximating for 1/200 year events. However, the focus for this exercise is on methodological / design issues. CEIOPS recognises that further calibration work will be required at a later stage in the Solvency II project to fully reflect the prudential objectives*", CEIOPS, 2006, p16

juger du caractère *relativement* risqué d'une branche par rapport à une autre serait utile, tant pour le pilotage de l'entreprise que pour sa régulation. Qu'en est-il ?

Formule standard - hiérarchie des risques

Les deux graphiques ci-dessous reproduisent, toujours pour les risques de prime et de provision, l'évolution des calibrages de chaque branche au fil des QIS. Chaque intersection entre deux courbes marque donc un changement d'opinion quant au risque relatif de deux branches.





Visuellement, il semble difficile de considérer que la hiérarchie des risques telle que perçue d'après ces mesures serait stable. Afin de caractériser cela de façon quantitative, présentons les résultats de notre indicateur de stabilité.

Tableau 5 : indice de stabilité relative des risques perçus

	ensemble des calibrages		hors QIS 2	
	brut	enrichi	brut	enrichi
risques de prime	31%	19%	53%	36%
risques de provision	31%	19%	36%	28%

Nous constatons dans la première colonne que dans un tiers des cas seulement (31%), le jugement porté sur le caractère plus risqué d'une branche, tant vis-à-vis du risque de prime¹⁸ que vis-à-vis du risque de provision, ne s'est pas inversé

¹⁸ On pourrait remarquer que les courbes des branches *Marine, Aviation, Transport* et *Assistance* croisent toutes les autres ou presque, et croire de ce fait qu'elles sont la source d'un constat abusivement généralisé à l'ensemble des branches. Le calcul de ce ratio sur l'ensemble des

au gré des études. Si on prend également en compte une version enrichie de l'indicateur où l'on comptabilise non seulement les inversions de jugement, mais également les cas de figure où deux risques ont été jugé parfois strictement identiques et parfois que l'un était au moins deux fois plus risqué que l'autre (deuxième colonne), alors un cinquième seulement des jugements restent stables (19%).

En restreignant le champ de l'analyse aux calibrages post QIS2 (troisième et quatrième colonnes), on constate que, si le taux de stabilité est (par construction) légèrement supérieur, la fiabilité reste douteuse : un acteur qui prendrait une décision de développement ou d'immobilisation de capital en privilégiant une branche plutôt qu'une autre sur la base des indicateurs changerait d'opinion près de la moitié du temps s'il se référait au risque de prime, et près des deux tiers du temps s'il se référait au risque de provision.

Une analyse de certains couples permet de mieux souligner l'absence totale de stabilité des calibrages. Prenons quelques exemples à titre d'illustration :

- Risque de prime : lors du QIS 2, l'assistance était considérée comme une fois et demi plus risquée que la responsabilité civile (RC) (15% vs. 10%), lors du QIS 3, elles étaient aussi risquées l'une que l'autre (10%), lors du QIS 5, c'était la RC qui était devenue trois fois plus risquée que l'assistance (15% vs. 5%), et finalement, on a considéré qu'elle était environ une fois et demi plus risquée, à peu près comme lors du QIS 4.
- Risque de prime toujours : d'après le QIS 2, le crédit/caution était deux fois et demi plus risqué que le *divers* (25% vs. 10%), lors du QIS 3, ils étaient devenus aussi risqués l'un que l'autre (12,5%), le QIS 5 estimait de nouveau que le crédit/caution était bien plus risqué (21,5% vs. 13%), et finalement, on a considéré qu'il était moins risqué (12% contre 13%).
- Risque de provision : lors du QIS 2, la RC était considérée comme plus risquée que l'assurance transport, lors du QIS 3, c'était le contraire (avec un facteur deux d'écart), lors du QIS4, on revient sur la hiérarchie initiale (facteur 1,5 d'écart), le QIS 5 renverse de nouveau cet ordre... avant que la version finale ne les considère comme équivalents.

Modèle interne

On pourrait s'attendre à ce que les calibrages issus d'un modèle interne soient significativement plus stables que ceux des différents tests de la formule standard, en premier lieu en raison de la stabilité souhaitable du cadrage

branches hormis celles-ci ne remet cependant pas en cause l'observation : le taux de stabilité relative resterait faible, à 48 % au lieu de 31 % (10 cas stables sur 21, au lieu de 11 sur 36).

méthodologique en vue d'une approbation par le superviseur, en deuxième lieu en raison de leur immunisation relative aux aléas des orientations politiques auxquels sont soumis des textes européens en construction et en troisième lieu, enfin, parce que les données sur lesquelles nous avons pu nous appuyer débutent plusieurs années après la mise en place d'un modèle interne et la production de chiffres associés par l'entité considérée, permettant ainsi un début de cristallisation. Ainsi, sur ces données, le biais de mesure de la marge d'erreur lié au fait que notre méthodologie ne capte que le bruit interne et considère une pratique stable comme fondée pourrait être suffisamment fort pour que la marge d'erreur perçue soit très faible.

Toutefois, la reproduction des mêmes études sur le calibrage du risque de provision tel que calculé par le biais d'un modèle interne, conduisent à des résultats qui, s'ils sont effectivement légèrement meilleurs, montrent que le bruit interne reste considérable :

Tableau 6 : dispersion moyenne du signal et du bruit à partir d'un modèle interne (2009-2014)

	stdev	max-min/min
signal brut	10,9%	1,8
bruit	6,5%	0,9
SSBa	68%	106%

On constate ainsi que :

- L'estimation du risque associé à une branche donnée évolue en moyenne d'un facteur 2 au cours de la période (max-min/min=0,9),
- l'amplitude du signal interne est inférieure ou comparable à celle du bruit interne.

L'indicateur de stabilité de la hiérarchie des cinq branches sur la période de six années considérée est de 40% (30% si on élargit la notion d'instabilité au fait que deux risques sont jugés parfois égaux et parfois différents de plus d'un facteur deux). Ceci signifie que le jugement servant le cas échéant à fonder une décision aura été inversé dans la majorité des cas.

Cette dispersion peut être de nature politique, technique, ou les deux. Si elle est de nature politique, alors il s'agit d'une nature politique différente de celle de la formule standard : ce serait ici des enjeux internes à l'entreprise (e.g. confirmation de la pertinence du choix d'une orientation stratégique) et non de

défense d'une industrie par des pays ou des fédérations professionnelles. Si cette dispersion est de nature technique et liée à des choix d'experts, alors cela détruit l'ambition même d'une vision *risk based* scientifiquement fondée ; si elle est de nature technique et liée à une évolution dans les données, alors cela détruit l'idée d'une pertinence possible d'une formule standard stable dans le temps. Qui plus est, dans ce cas, il apparaît que la période d'obsolescence des informations fournies par ces calibrages n'est pas supérieure à la durée des produits commercialisés, ce qui conduit à douter de la pertinence de leur utilisation à des fins de pilotage par les entreprises.

Section V : les conséquences

On peut débattre du fait que, dans un contexte global où co-existent non seulement des enjeux de protection des assurés mais aussi de développement de marchés d'assurance et de financement de l'économie, les exigences financières imposées aux entreprises le soient en fonction de seuls critères prudentiels. On peut ensuite débattre du fait que, d'un strict point de vue prudentiel, pour protéger les assurés, éviter les faillites et optimiser l'allocation des fonds propres, il soit souhaitable de retenir une réglementation *risk based*, c'est-à-dire de fonder les exigences de capital sur une mesure de risque. Quoi qu'il en soit, c'est ce double choix qui a été retenu par Solvabilité 2. Or, il apparaît que cet objectif n'est pas atteint : les calibrages sur lesquels sont fondées les exigences de capital, réputés mesurer le risque, ne sont pas fiables.

Qui plus est, il est permis de douter qu'un tel objectif de régulation *risk based* soit atteignable. De fait, il n'est pas possible de considérer que les technologies de mesure des risques sont aujourd'hui inexistantes. Or les éléments présentés dans cet article montrent que l'enjeu n'est pas dans *un raffinement* de la précision, mais bien dans *l'incapacité* actuelle de caractériser le risque relatif d'une activité par rapport à une autre : on ne peut donc pas s'appuyer sur cette base pour fonder de décision.

Dans ce cadre, que faire du point de vue de chacune des sources de bruit, c'est-à-dire du point de vue des limitations techniques et des arbitrages politiques ?

Mieux cerner les limitations techniques

L'étude que nous avons menée peut être enrichie de deux façons. D'une part, en élargissant le champ de l'analyse, depuis les seuls risques de prime et provision en non vie, à l'ensemble des risques. D'autre part ne testant pas que la crédibilité interne mais également la marge d'erreur dans l'absolu des calibrages via des jeux d'hypothèses alternatives.

Sans prétendre dresser un panorama exhaustif des champs de recherche à ouvrir, il est possible de citer à titre d'exemple, toujours en ne testant qu'une cohérence interne aux pratiques en vigueur, donc sans aller jusqu'à identifier le spectre de l'ensemble des calibrages légitimement crédibles, les champs suivants :

- *Impact de la classification des risques.* Par exemple, une même entreprise, commercialisant le même produit d'assurance perte de bagages, se voit recommander selon les régulateurs nationaux de le classer en *pertes pécuniaires* (France), *fire and damages* (UK), *assistance* (Allemagne), ou *Marine Aviation Transport* (Italie), faisant évoluer d'un facteur deux les mesures de risque associées. Un autre exemple est celui de la prévoyance/santé, qui peut être modélisée à partir de méthodologies vie ou non vie à la discrétion des organismes. Si la comparaison n'est ici pas aussi triviale que les simples comparaisons de chocs que nous avons menées, puisque les exigences de capital sont fondées sur des structures de calcul différentes, les études menées par des consultants pour des organismes font également apparaître un facteur deux d'écart *a minima* entre ces deux types de méthodologies ouverts par la réglementation.
- *Impact des choix de modèle interne.* Nous avons évoqué l'étude menée par la FSA auprès de différentes banques consistant à leur demander de mesurer le risque associé à un portefeuille d'actifs à partir de leur modèle interne pour évaluer la façon dont les différents choix de modélisation conduisaient à apprécier différemment le risque d'un même produit. Une telle étude, envisagée mais considérée à ce stade comme non prioritaire par les régulateurs (qui ont validé plusieurs modèles internes) pourrait être utilement menée auprès des différents assureurs utilisant un modèle interne. Pour les assureurs vie, la simple application de la formule standard nécessitant à elle seule l'élaboration de modèles actif/passif *ad hoc*, cette étude gagnerait à être menée auprès de l'ensemble des acteurs.

Plus globalement, si des études de sensibilité à des hypothèses ont été menées¹⁹, serait utile de dresser une revue de ces analyses afin de déterminer, mesure de risque par mesure de risque, composante de l'exigence de capital par composante, la marge d'erreur associée. Toutefois, insistons sur le fait que l'enjeu n'est pas tant les sensibilités, qui sont oubliées aussitôt présentées et

¹⁹ Et, si elles sont totalement éparses, elles sont nombreuses : impact du choix de telle ou telle référence de courbe des taux, du choix de tel ou tel calibrage de stress de marché, du choix de tel modèle Cat plutôt que tel autre, de telle *management rule* plutôt que telle autre, d'une référence aux notations ou aux *spreads* de marché pour évaluer le risque de défaut souverain d'un portefeuille, etc. tant en interne des entreprises qu'à des niveaux agrégés par exemple au niveau de fédérations professionnelles ou de régulateurs.

servent souvent de bonne conscience scientifique, de quitus à la présentation et l'utilisation d'un résultat non fiable, que la marge d'erreur. Dans ce cadre, la sensibilité à une hypothèse n'est qu'un moyen de déterminer l'impact de la marge d'erreur des paramètres en entrée sur la marge d'erreur de la variable de sortie.

Piloter en conscience

Surtout, les études de sensibilité menées n'ont jamais eu pour objectif, à notre connaissance, de déterminer si les mesures étaient suffisamment précises pour servir de socle à une prise de décision. Toujours, elles ont visé à permettre de choisir une hypothèse plutôt qu'une autre. Trois raisons peuvent être avancées pour expliquer cela : le fait qu'un système *risk based* allait être mis en œuvre, et que l'enjeu portait donc sur ses caractéristiques et non sur son bien-fondé ; la volonté des modélisateurs de ne pas dévaloriser leur propre travail et risquer d'être perçus comme inutiles²⁰ ; le sentiment diffus et omniprésent du « mieux que rien », d'une supériorité *par principe* du résultat d'un calcul sur un montant forfaitaire. Concernant cette dernière explication, nous avons montré qu'il n'en était rien puisqu'un calcul censé nous éclairer sur le fait que telle branche serait plus risquée que telle autre est, dans 80% des cas démenti par une étude ultérieure, avant le cas échéant de se retourner de nouveau. L'information sur les risques apportée par Solvabilité 2, si elle se contredit elle-même et est alors susceptible d'avoir induit en erreur, n'est pas « mieux que rien », bien au contraire.

En l'état actuel des connaissances, l'exigence de capital d'une société d'assurance sous Solvabilité 2 est l'agrégation de plusieurs dizaines de composantes dont la marge d'erreur de chacune est inconnue. Sur le périmètre analysé, on constate que la marge d'erreur minimale envisageable est supérieure à un facteur deux et, globalement, on ne peut donc pas savoir si les ordres de grandeurs relatifs sont respectés. La résultante est donc totalement dépourvue de sens, même relatif : le fait qu'une entreprise ait une exigence de capital supérieure à une autre ne signifie *a priori* en rien qu'elle est plus risquée. Les conséquences de cet enseignement sur le pilotage actuel doivent être tirées tant au niveau macro-économique qu'au niveau microéconomique.

²⁰ Ainsi, interrogé par un consultant pour une enquête sur les pratiques ALM et les méthodes de place d'allocation stratégique d'actifs, j'eus l'échange suivant :

- Moi : l'important, c'est de connaître et de communiquer sur les marges d'erreurs.
- Lui : Oui, il faut présenter des analyses de sensibilité.
- Moi : Non, les sensibilités, c'est anecdotique. Ce qu'il faut présenter, c'est la marge d'erreur.
- Lui : Mais si on présente des marges d'erreur, les modèles vont perdre toute crédibilité !

Au niveau macro-économique, dès lors que ces calibrages apparaissent comme n'étant pas techniquement fiables, ils n'ont plus de légitimité pour fonder les exigences de capital des assureurs. Ceci plaide pour une reprise en main de ces exigences à un niveau ayant une superficie de compétences globale, à même d'articuler l'ensemble des enjeux du contexte (e.g. financement de l'économie), au-delà d'un calibrage *risk based* factice. Certes, on pourrait considérer que les enjeux politiques ont bien été pris en compte puisque l'une des sources d'évolution des calibrages d'une étude à l'autre est sans doute justement à rechercher, au-delà des difficultés techniques intrinsèques, dans les jeux d'acteurs des différentes parties prenantes, mais l'EIOPA, dont la compétence est exclusivement technique²¹ et qui est ici l'une des parties prenantes, n'a pas la légitimité pour arbitrer ces jeux d'acteurs.

Au niveau microéconomique, le danger principal est que ces outils soient effectivement utilisés comme indicateurs de risque pour piloter les entreprises. Ainsi, l'expression « le risque action » désigne désormais « l'exigence de capital associée au risque action »²², tendant à faire assimiler cette mesure sans valeur à la réalité sous-jacente et à faire gérer les risques en fonction de cet indicateur. Les entreprises les plus fragiles, donc celles pour lesquelles la qualité de la gestion des risques est la plus sensible, y sont d'autant plus exposées que la pression d'un ratio de solvabilité faible les conduira à aligner leur gestion sur cette grille de lecture afin d'optimiser l'image qu'elles renverront au régulateur. En termes de discours comme d'implémentation, vis-à-vis des exigences de use-test et au niveau des méthodologies d'ORSA et de gestion des risques, il importe de dissocier, autant que possible, l'appréhension des risques des quantifications de Solvabilité 2.

Conclusion

S'il est illusoire d'espérer que ce système sera abandonné à court terme, il est souhaitable, dans l'attente de progrès technologiques ou de sa refonte, d'assumer le fait que, en termes de mesure de risque, les calibrages de Solvabilité 2 *ne sont*

²¹ Nous n'entrons pas ici dans la sociologie institutionnelle de Solvabilité 2. Certes, formellement, les calibrages de l'EIOPA ne sont que des conseils à la Commission Européenne, mais ceux-ci sont en pratique suivis en raison de la légitimité technique relative de la première par rapport à la seconde. Ce poids de l'EIOPA est renforcé par la portée de ses recommandations vis-à-vis desquelles les Trésors nationaux sont dans une posture « *comply or explain* », inversant ainsi les positions entre l'autorité technique de jugement censée appliquer les règles et l'autorité politique censée fixer le cadre.

²² On peut citer par exemple, le passage suivant : « *La lucidité c'est d'être vigilant sur les exigences de capital pour les actifs les plus utiles à l'économie, en particulier sur le calibrage du risque de crédit. Ceci suppose aussi [...] de diminuer le calibrage des risques de responsabilité civile de long terme, le calibrage du risque de longévité [...]* » (Spitz, 2011).

pas « mieux que rien ». L'ensemble des acteurs de l'assurance européenne, assureurs comme régulateurs, entre dans une ère de « *damage control* ».

Bibliographie

ACPR (Autorité de Contrôle Prudentiel et de Résolution), 2011, Solvabilité 2 : principaux enseignements de la cinquième étude quantitative d'impact (QIS5), *Analyses et synthèses* n°1, 24p.

Alesina, A. and Tabellini, G., 2007, Bureaucrats or Politicians? Part I: A Single Policy Task. *American Economic Review*, 97(1): 169-179.

CEIOPS, 2006, QIS2 Technical Specification, <https://eiopa.europa.eu/Publications/QIS/QIS2TechnicalSpecification.pdf>

CEIOPS, 2007, QIS3 Technical specifications, Part I: Instructions, <https://eiopa.europa.eu/Publications/QIS/QIS3TechnicalSpecificationsPart1.PDF>

CEIOPS, 2007b, QIS3 Calibration of the underwriting risk, market risk and MCR, <https://eiopa.europa.eu/Publications/QIS/QIS3CalibrationPapers.pdf>

CEIOPS, 2008, QIS4 Technical Specifications (MARKT/2505/08), <https://eiopa.europa.eu/publications/qis/insurance/insurance-quantitative-impact-study-4> (A.3)

CEIOPS, 2010a, QIS5 Technical Specifications, https://eiopa.europa.eu/Publications/QIS/QIS5-technical_specifications_20100706.pdf

CEIOPS, 2010b, Solvency II calibration paper, <https://eiopa.europa.eu/Publications/QIS/CEIOPS-Calibration-paper-Solvency-II.pdf>

Cummins, J., & Phillips, R. (2009). Capital adequacy and insurance risk-based capital systems. *Journal of Insurance Regulation*, 28(1), 25-72.

Danielsson J., 2008, Blame the models, *Journal of Financial stability*, 4 (4). pp. 321-328

Danielsson J., 2002 The emperor has no clothes: Limits to risk modelling, *Journal of Banking and Finance*, vol. 26, issue 7, pages 1273-1296

El Karoui, N., Loisel, S., Prigent, J.-L., & Vedani, J. (2015), Market inconsistencies of the market-consistent European life insurance economic

valuations: pitfalls and practical solutions, *working paper*, <https://hal.archives-ouvertes.fr/hal-01242023>

Eling, M., & Holz Müller, I. (2008). An overview and comparison of risk-based capital standards. *Journal of Insurance Regulation*, 26(4), 31-60

European Commission, 2008, MARKET/2504/08, Call for Advice from CEIOPS, <https://eiopa.europa.eu/Publications/QIS/Call%20for%20Advice%20from%20EC%20to%20CEIOPS.pdf>

European Parliament and Council, 2009, Directive 2009/138/CE du Parlement européen et du Conseil du 25 novembre 2009 sur l'accès aux activités de l'assurance et de la réassurance et leur exercice (solvabilité II) *OJ L 335*, 17.12.2009, p. 1–155

Frezal, S., 2015, de quoi Solvabilité 2 est-il le nom ? *working paper PARI*, <http://www.chaire-pari.fr/wp-content/uploads/2015/12/De-quoi-Solvabilit%C3%A9-2-est-elle-le-nom-dec-2015.pdf>

Harrison, M. & Samuel, S., 2011, Two hundreds millions inputs. Can you trust risk weightings at European banks? *Barclays Capital Equity Research*

Kamega, A. & Planchet F., 2013, Construction de tables de mortalité prospectives sur un groupe restreint : mesure du risque d'estimation, *Bulletin français d'Actuariat*, *Bulletin Français d'Actuariat*, vol. 13, n°25.

Leroy, G. & Planchet, F. 2010, Que signifie la ruine dans Solvabilité 2 ? *La Tribune de l'assurance*, n° 147, mai 2010, 55-56

Repulo & Martinez-Miera, 2014, Comparing Flat and Risk-based Capital Requirements, https://acpr.banque-france.fr/fileadmin/user_upload/acp/Seminaires_de_recherche/Chaire_ACP/02RR-KNS.pdf

Sims, C., 2003, Implications of rational inattention, *Journal of Monetary Economics*, Volume 50, Issue 3, April 2003, Pages 665–690

Spitz, B., 2011, Solvabilité II, cinq conditions, *Risques* n°85, mars 2011, 39-41.

Weber, R., & Darbellay, A. (2008). The regulatory use of credit ratings in bank capital requirement regulations. *Journal of Banking Regulation*, 10(1), 1-16.

Annexe 1 : exemple, indicateurs et intuition du ratio signal sur bruit

Le tableau ci-dessous présente un exemple de signal et de bruit : afin de ne pas être trop stressé, je souhaite savoir avant de regarder un James Bond lequel des principaux personnages a le moins de chances de mourir. Je demande à deux amis, l'un ayant vu la bande annonce, l'autre ayant vu plusieurs épisodes de James Bond. Chacun me fournit une mesure du risque de décès.

Tableau 2 : exemple d'indicateur de bruit et de signal

	Approche 1 (dangers auquel le personnage est confronté)	Approche 2 (statistique sur les films précédents)	bruit
Risque A (proba de décès de James Bond)	0,9	0	0,9
Risque B (proba de décès d'une James Bond girl)	0,6	0,7	0,1
signal brut	0,3	0,7	

Nous avons ici deux risques, chacun d'entre eux pouvant être évalué à partir de deux méthodes. La dispersion, pour une approche donnée, entre les estimations de chaque risque (0,3 pour l'approche 1 et 0,7 pour l'approche 2) reflète la capacité de discrimination de cette approche : elle correspond à une appréciation de l'amplitude du signal brut (perception de l'information fournie par l'analyse). La dispersion, pour un risque donné, de la mesure suivant les différentes approches (0,9 ou 0,1) me fournit une estimation de l'amplitude du bruit (ma marge d'erreur, ce qui parasite le signal).

Ces écarts seront mesurés pour un ensemble de risques et un ensemble d'approches, à l'aide de l'un ou l'autre de ces indicateurs intuitifs et usuels de dispersion :

$$\text{Dispersion standard} = \text{écart} - \text{type}$$

et

$$\text{Dispersion extrême} = \frac{\max - \min}{\min}$$

Ce deuxième indicateur ne peut être que positif et, s'il est supérieur à 1, cela signifie qu'il y a plus d'un facteur deux d'écart entre les estimations la plus faible et la plus élevée.

A partir de ces indicateurs, je construis le ratio ajusté signal sur bruit (SSB_a) de la façon suivante :

$$SSB_a = \frac{\text{signal brut} - \text{bruit}}{\text{bruit}}$$

Considérer un tel ratio plutôt qu'un simple ratio signal brut /bruit permet de prendre en compte le fait que l'indicateur de signal brut, correspondant à la mesure de dispersion au sein d'une approche donnée, est lui-même pollué par le bruit. Ainsi, mon indicateur ajusté est tel que, si toutes les cases de notre tableau étaient remplies uniformément par un générateur aléatoire de même loi et même paramètre, les indicateurs de signal brut et de bruit étant asymptotiquement identiques, le ratio signal sur bruit considéré serait nul en espérance, ce qui reflète le fait que le signal brut perçu ne serait que du bruit vide d'information²³. Cet ajustement repose sur une hypothèse d'isotropie de la matrice des erreurs, hypothèse qui est formalisée et discutée en annexe 2, et dont les conséquences ne modifient pas la nature du résultat obtenu.

Annexe 2 : fonctionnement et modalités de l'ajustement du ratio signal sur bruit.

Nous formalisons et discutons dans cette annexe les raisons et conséquences du choix d'un ratio signal sur bruit ajusté plutôt qu'un signal brut.

Formalisation

On peut noter le calibrage du risque $R_{i,j}$ associé à la branche (*line of business*, LoB) i dans le jeu de calibrages (e.g. QIS) j sous la forme

$$R_{i,j} = V_i + n_{i,j}$$

Où V_i représente la valeur du paramètre de risque recherché pour la branche i et $n_{i,j}$ représente le bruit associé à la branche i pour le jeu de mesures j .

Ce bruit $n_{i,j}$ peut comporter :

- des composantes spécifiques à chaque branche. Par exemple, une branche a un déroulement plus long que les autres et des techniques extrapolant insuffisamment le risque de dérive conduiraient à sous-estimer le risque associé, puis cela est revu à la hausse ; une branche est stratégique pour la puissance publique qui souhaite développer ce marché (ou bien pour une entreprise qui en avait un axe de sa communication financière), et son calibrage doit être revu à la baisse ; etc.

²³ alors que, si on retenait dans ce cas un ratio brut (signal brut sur bruit), on obtiendrait un résultat trompeur de 1, sous-entendant qu'il existe un signal de même amplitude que le bruit.

- des composantes spécifiques à chaque jeu de calibrage. Par exemple, l'utilisation d'une nouvelle méthodologie prenant en compte la possibilité de lois à queues épaisses qui conduirait à augmenter l'ensemble des calibrages ; ou bien une décision politique de une baisse transversale des mesures retenues résultant du constat que les calibrages sont tels que l'industrie de l'assurance européenne serait, dans son ensemble, exposée à des exigences trop pénalisantes ; etc.
- un résidu, correspondant par exemple au fait qu'à la suite d'un événement particulier, on révisé son jugement sur une branche donnée, avant le cas échéant qu'une nouvelle information vienne infirmer ce jugement.

Dès lors, il peut être modélisé à partir de trois composantes. Par exemple, si elles interagissent de façon additive, sous la forme :

$$n_{i,j} = L\varepsilon_i + Q\sigma_j + B\tau_{i,j}$$

Où ε_i , σ_j et $\tau_{i,j}$ sont des bruits blancs normés reflétant respectivement la composante aléatoire :

- pour ε , liée à la branche,
- pour σ , liée au jeu de calibrage
- et pour τ , non spécifique,

et L , Q et B représentent l'amplitude de chacune de ces composantes²⁴.

Signification d'un ratio non ajusté

Si on supposait L et B nuls, c'est-à-dire en considérant que la seule source de bruit vient des différents jeux de mesure et qu'elle affecte de façon homogène toutes les branches, alors on aurait :

$$R_{i,j} = V_i + Q\sigma_j$$

Dans ce cas, notre indicateur de bruit correspond à \bar{Q} , et l'entièreté de la dispersion sur une colonne correspond à de l'information. Ainsi, notre indicateur de signal brut, qui correspond à $\overline{dispersion}_i(R_{i,j}) = \overline{dispersion}(V_i)$, capte effectivement l'amplitude du signal.

²⁴ Ou, de façon plus fine, $n_{i,j} = L_j\varepsilon_i + Q_i\sigma_j + B\tau_{i,j}$, avec L_j et Q_i représentant respectivement, l'amplitude pour chaque calibrage j du bruit affectant chaque branche, et l'amplitude pour chaque branche i du bruit issu du changement de QIS.

Ainsi, ce ratio *non ajusté* signal brut /bruit, qui détermine le ratio signal/bruit dans le cas où L et B sont nuls, est un majorant du ratio signal /bruit.

Ce ratio brut correspond à $1 + SSB_{\alpha}$ et, appliqué aux données de l'article, il indique que l'amplitude du bruit est, en ordre de grandeur, comparable à celle du signal :

Tableau 7 : ratios signal brut sur bruit (non ajustés)

	stdev	max/min
Formule standard - primes	1,3	2,0
Formule standard - provision	1,1	1,5
Modèle interne - provision	1,7	2,1

Signification du ratio ajusté

Si, en l'absence de possibilité d'estimer L ni de postuler une hiérarchie entre L et Q , on suppose $L = Q$, avec B quelconque, alors on peut réécrire $n_{i,j}$ sous la forme $n_{i,j} = N \nu_{i,j}$ où $\nu_{i,j}$ est un bruit blanc normé. Il convient alors de corriger la dispersion correspondant à l'indicateur de signal brut par la dispersion induite par le bruit d'amplitude caractéristique N , telle que mesurée par l'indicateur de bruit. Ceci correspond au ratio ajusté signal sur bruit tel que présenté dans le corps du papier.

Ce résultat tient, de façon plus générale, avec L quelconque et $Q \ll B$, où on peut alors écrire $n_{i,j}$ sous la forme $n_{i,j} \cong N \nu_{i,j}$.

Enfin, dans le cas où $Q \ll L$, l'ajustement auquel nous procédons est insuffisant, puisque le bruit mesuré entre QIS est faible par rapport au bruit entre branches assimilé à tort à de l'information dans la mesure du signal brut. Le ratio ajusté signal sur bruit tel que proposé surestime alors l'information apportée relativement au bruit.

Annexe 3 : mesure de stabilité de la hiérarchie des risques

Formellement, la mesure de stabilité du positionnement relatif des risques peut être représentée ainsi :

Soit $R_{i,j}$ la mesure du risque i dans l'approche j ,

$$\text{Soit } C_{i,i',j} = \begin{cases} 1 & \text{si } R_{i,j} > R_{i',j} \\ 0 & \text{si } R_{i,j} = R_{i',j} \\ -1 & \text{si } R_{i,j} < R_{i',j} \end{cases} \text{ exprimant la position relative des risques } i$$

et i' vue par l'approche j ,

$$\text{Soit } T_{i,i'} = \begin{cases} 1 & \text{si } \text{signe}(C_{i,i',j}) = \text{constante } \forall j \\ 0 & \text{sinon} \end{cases}$$

$$\text{Stabilité} = \frac{\sum_i \sum_{i'} T_{i,i'}}{n(n-1)/2}$$

Cet indicateur est compris entre 0 et 1 et représente, parmi l'ensemble des couples de risques envisageables, la proportion de ceux dont la hiérarchie de la mesure de risque reste stable au fil des études.

Si nous reprenons notre exemple *supra*, où nous n'avons que deux risques, donc une seule paire à comparer, $C_{A,B,1}=1$, $C_{A,B,2}=-1$, $T_{A,B}=0$, notre indicateur de stabilité est donc nul : ma vision du risque relatif a évolué lorsque j'ai changé d'approche.

PARI

PROGRAMME DE RECHERCHE
SUR L'APPRÉHENSION DES RISQUES
ET DES INCERTITUDES

PARI, placé sous l'égide de la Fondation Institut Europlace de Finance en partenariat avec l'ENSAE/Excess et Sciences Po, a une double mission de recherche et de diffusion de connaissances.

Elle s'intéresse aux évolutions du secteur de l'assurance qui fait face à une série de ruptures : financière, réglementaire, technologique. Dans ce nouvel environnement, nos anciens outils d'appréhension des risques seront bientôt obsolètes. PARI a ainsi pour objectifs d'identifier leur champ de pertinence et de comprendre leur émergence et leur utilisation.

L'impact de ses travaux se concentre sur trois champs :

- les politiques de régulation prudentielle, l'optimisation de leur design technique et leur appropriation pour le pilotage, dans un contexte où Solvabilité 2 bouleverse les mesures de solvabilité et de rentabilité ;
- les outils d'allocation stratégique d'actifs des investisseurs institutionnels, dans un environnement combinant taux bas et forte volatilité ;
- les solutions d'assurance, à l'heure où le big data déplace l'assureur vers un rôle préventif, créant des attentes de personnalisation des tarifs et de conseil individualisé.

Dans ce cadre, la chaire PARI bénéficie de ressources apportées par Actuaris, la Financière de la Cité, Generali et le Groupe Monceau.

Elle est co-portée par Pierre François, directeur du département de sociologie de Sciences Po et Sylvestre Frezal, directeur à Datastorm, la filiale de valorisation de la recherche de l'ENSAE.

PARTENAIRES

