

# EXEC SUM

## **SOLVABILITÉ 2 EST-IL RISK BASED ?** QUAND LE BRUIT NOIE LE SIGNAL



Sylvestre Frezal

**PARI**

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

## Abstract

L'exigence de capital sous Solvabilité 2 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 internes portant sur l'évolution dans le temps des calibrages de la formule standard puis d'un modèle interne, nous montrons que cette mesure n'est pas fiable, le bruit la parasitant étant d'une amplitude comparable à celle du signal qu'on cherche à capter.

En termes de niveau relatif, d'outil de hiérarchisation des risques, nos résultats battent également en brèche le sentiment diffus et omniprésent du « *mieux que rien* », d'une supériorité *par principe* du résultat d'un calcul sur une approche qualitative. Nous montrons ainsi que, au fil des calibrages de Solvabilité 2, les mesures censées nous éclairer sur le fait que telle branche serait plus risquée que telle autre sont, dans 80% des cas, démenties par une étude ultérieure, avant le cas échéant de se retourner de nouveau : l'information sur le niveau relatif de risque apportée par Solvabilité 2, loin d'éclairer par « *un bon proxy* », se contredit elle-même et induirait souvent en erreur.

Ceci remet en cause la pertinence de ce système pour fonder des décisions de régulation prudentielle ou éclairer les choix d'orientation stratégique ou de gestion des risques des organismes d'assurance.

EXEC  
SUM

**SOLVABILITÉ 2 EST-IL RISK BASED ?**  
QUAND LE BRUIT NOIE LE SIGNAL

## So what ?

- > Au niveau macro-économique, les calibrages de Solvabilité 2 n'étant pas techniquement fiables, ils n'ont pas de légitimité pour fonder les exigences de capital des assureurs.  
Ceci plaide pour une reprise en main de ces exigences à un niveau politique, à même d'articuler l'ensemble des enjeux (e.g. financement de l'économie, maintien d'un marché, etc.), au-delà d'un calibrage *risk based* factice, uniquement formel.

---
- > Au niveau microéconomique, le danger principal est que ces outils soient effectivement utilisés comme indicateurs de risque pour piloter les entreprises.  
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 donc de dissocier, autant que possible, l'appréhension des risques (indépendante de Solvabilité 2) des quantifications de Solvabilité 2 (qui ont un impact financier mais ne permettent pas d'évaluer le risque).

---
- > Les études de sensibilité qui, oubliées aussitôt présentées, servent souvent de bonne conscience scientifique, de quitus à la présentation et l'utilisation d'un résultat non fiable, ne sauraient se substituer à la mesure des marges d'erreur qui devraient accompagner systématiquement un chiffre.

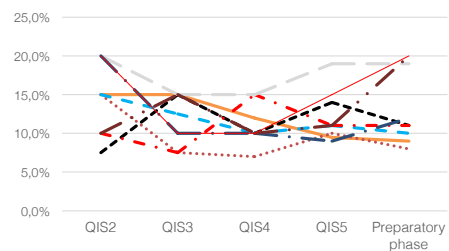
## SOLVABILITÉ 2 EST-IL RISK BASED ? QUAND LE BRUIT NOIE LE SIGNAL

Solvabilité 2 se veut *risk based* : les exigences de capital y sont réputées refléter la prise de risque, envoyant des messages adaptés aux régulateurs (un ratio de solvabilité articulant *level playing field* et protection des assurés) et aux dirigeants et investisseurs (des décisions fondées sur une rentabilité ajustée aux risques).

Toutefois, les mesures de risque sur lesquelles cette réglementation repose ont fait l'objet de fortes variations. Par exemple, en 2011, alors que son chantier d'élaboration a débuté depuis une décennie et que plusieurs études de calibrage des paramètres ont été menées, une fédération professionnelle apprécie « *une avancée positive sur le scénario catastrophe RC spécialisée, calibré à 100% au lieu de 225% proposé au départ* ». A l'inverse, l'arrivée d'un nouveau calibrage du risque tempête génère des protestations : « *cela double le coût de la pire tempête de période de retour 200 ans* ». 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 ?

Les réponses usuellement apportées à cette critique sont du type « *il ne faut pas jeter le bébé avec l'eau du bain* », « *ce n'est pas parfait mais*

### Évolution des calibrages du risque de provision



Chaque trait représente une branche ; chaque intersection marque un changement d'opinion quant à la hiérarchie des risques

*c'est mieux que rien* ». Pour apprécier si ces réponses sont fondées, il nous faut évaluer si ces indicateurs sont suffisamment robustes pour fonder une décision *opérationnelle* dans les champs qu'ils prétendent éclairer. Cela suppose d'éclaircir deux points : (1) les observations *supra* sont-elles singulières ou générales, et (2) 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 ?

### De nombreuses sources de pollution inévitables

Lorsqu'on quantifie les risques, la mesure doit nous renseigner sur la hiérarchie et la proportionnalité des prises de risques de différentes compagnies, branches, ou actifs. Cette information est le *signal* qu'on cherche à isoler.

Mais, face à cette ambition réglementaire, plusieurs difficultés émergent :

- *Une difficulté scientifique* : on cherche à appréhender des événements rares, sur lesquels on dispose de peu ou pas de statistiques. Il est donc nécessaire d'extrapoler, de façon nécessairement subjective et peu robuste.
- *Une difficulté interprétative* : la VaR visée par Solvabilité 2 doit-elle être interprétée comme « chaque année, une entreprise sur 200 fait faillite », ou bien « tous les deux siècles, le marché fait faillite » ? Dans un cas, il faudrait ne quantifier que le risque intrinsèque, et dans l'autre, que le risque systémique. Les calibrages qui en résultent sont différents.
- *Une difficulté politique* : la réglementation prudentielle est immergée dans un contexte socio-économique plus large, ce qui exige des

arbitrages et peut conduire à la distorsion d'un calibrage réputé purement technique. L'exemple le plus emblématique est le postulat de risque nul sur certaines dettes souveraines.

- *Une difficulté opérationnelle* : les risques peuvent évoluer plus rapidement que la fréquence de révision des calibrages, et la maille d'analyse (e.g. niveau national ou européen) peut être inadaptée, rassemblant des ensembles hétérogènes, introduisant un écart entre la mesure et la « vraie » valeur.

Ces sources de *bruit* viennent polluer le signal fourni par la mesure et posent de sérieuses limites à la prétention de Solvabilité 2 d'offrir un cadre « fondé sur les risques ». Afin d'éclairer dans quelle mesure on peut se fonder sur les exigences de capital de Solvabilité 2 comme mesure de risque, nous évaluons, premièrement, les amplitudes relatives du signal et du bruit (appréciation de la capacité de ces mesures à segmenter l'ampleur des risques portés) et, deuxièmement, la stabilité de la hiérarchie des risques (appréciation de la fiabilité d'un arbitrage reposant sur ces mesures).

### Les données utilisées

*Formule standard* - Nous utilisons pour mener l'étude cinq jeux de calibrages de la formule standard : les QIS successifs (2 à 5) et les mesures d'implémentation de niveau 2. Afin d'avoir un jeu de mesures du signal significatif et méthodologiquement homogène, nous nous intéressons aux risques de souscription non vie (calibrages des risques de prime et de provision des neuf LoB S2 d'assurance directe).

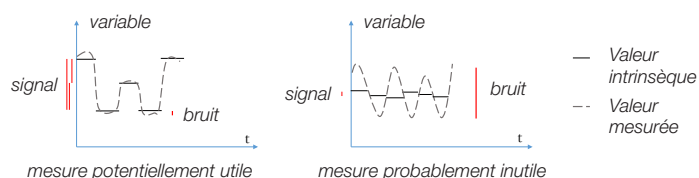
*Modèle interne* - Dans un second temps, nous poursuivons l'étude en l'étendant aux jeux de calibrages successifs d'un modèle interne : nous disposons de données exploitables pour le risque de provision pour lesquels cinq branches étaient disponibles (auto RC, auto matériel, marine-aviation-transport, dommages et RC), sur six années (2009 – 2014).

### Comment apprécier le caractère opérationnel d'une mesure ?

L'appréciation relative du bruit et du signal doit nous permettre d'apprécier si la mesure fournit une information opérationnellement utilisable. Prenons un exemple : supposons que nous souhaitons utiliser un altimètre pour piloter un avion. Si, au début du XXème 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 à *minima* pour éviter une montagne).

### Le signal et le bruit : des calibrages n'apportant guère d'information

Nous mesurons l'amplitude relative du signal et du bruit afin de déterminer si la mesure véhicule une information exploitable ou bien si elle est noyée dans le bruit.



Pour cela, nous comparons :

- un indicateur de *bruit* reflétant la dispersion des calibrages alternatifs d'une même LoB
- à un indicateur de *signal* captant l'écart de calibrage entre les différentes LoB.

#### Exemple d'estimation du signal et du bruit

	QIS 2	QIS 3	QIS 4	...
LoB 1	1	0,9	1,1	...
LoB 2	3			
LoB 3	2			
...	...			

Amplitude caractéristique du bruit : 0,1

Amplitude caractéristique du signal : 1

Signal, net du bruit :  $1 - 0,1 = 0,9$   
 Ratio signal sur bruit :  $0,9/0,1 = 9$

Le bruit n'altère ici guère le signal

Nous retenons deux indicateurs de dispersion : l'un captant des mesures du signal et du bruit « extrêmes », en mesurant le ratio entre les valeurs les plus éloignées l'une de l'autre, et l'autre mesurant l'amplitude du signal et du bruit « courant », à l'aide de l'écart-type.

#### Ratio signal sur bruit

	amplitude moyenne	amplitude extrême
formule standard - risque de prime	28%	99%
formule standard - risque de provision	12%	47%
modèle interne - risque de provision	68%	106%

Les résultats synthétisés dans le tableau ci-dessous montrent que le signal, c'est-à-dire l'information apportée par la mesure, est noyé dans le bruit : il ne représente, pour les variations courantes captées par l'écart-type, que 28% du bruit pour le risque de prime mesuré par la formule standard, 12% du bruit pour le risque de provision calibré par la formule standard, et 68% du bruit pour le risque de provision capté par un modèle interne.

### Hiérarchie des risques : une instabilité trop importante pour que le classement soit fiable

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 branches les moins risquées. D'autant plus qu'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. Dans un tel cas de figure, la capacité des mesures de risque à 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 ?

Nous avons construit un indicateur de stabilité de la hiérarchie des risques. Pour cela, (1) nous recensons toutes les combinaisons de deux branches possibles (e.g dommage vs. santé), puis (2) nous regardons pour chaque couple si la hiérarchie mesurée (quelle branche est la plus risquée ?) est stable (indépendante de l'approche retenue) et (3) nous mesurons la proportion de couples dont le positionnement relatif est stable.

#### Taux de stabilité de la hiérarchie des risques perçus

	brut	enrichi
formule standard - risque de prime	31%	19%
formule standard - risque de provision	31%	19%
modèle interne - risque de provision	40%	30%

Nous constatons (première colonne) que dans un tiers des cas seulement (31%), le jugement porté sur le caractère plus risqué d'une branche par rapport à une autre, ne s'est pas inversé au gré des études (tant vis-à-vis du risque de prime que vis-à-vis du risque de provision). Si on prend en compte une version enrichie de l'indicateur (deuxième colonne) où l'on comptabilise non seulement les inversions de jugement, mais également les cas de figure où deux risques ont été jugés parfois identiques et parfois différents de plus d'un facteur deux, alors un cinquième seulement des jugements restent stables (19%).

Dans le cas du modèle interne, l'indicateur de stabilité de la hiérarchie des cinq branches sur la période de six années considérée est de 40% (30% avec la notion d'instabilité élargie). Ceci signifie que le jugement servant le cas échéant à fonder une décision aura été inversé au moins une fois dans la majorité des cas.

### Le risk based, une utopie ?

A tort ou à raison, Solvabilité 2 souhaitait être *risk based*. Cet objectif n'est pas atteint : les calibrages sur lesquels sont fondées les exigences de capital, réputés mesurer le risque, ne sont pas fiables. On ne peut donc pas s'appuyer sur une telle base pour fonder de décision.

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. De même, il n'est guère permis d'espérer que des données sur les chocs bicentennaux vont spontanément apparaître dans un avenir proche. 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.

# PARI

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

PARI (Programme sur l'Appréhension des Risques et des Incertitudes) est une chaire de recherche portée par l'Institut Europlace de Finance, l'ENSAE, Sciences Po et Datastorm. Son objectif est d'identifier le champ de pertinence de nos concepts et outils d'appréhension des risques, ainsi que leurs modalités d'émergence et d'utilisation.

Elle est financée par Actuaris, la Financière de la cité, Generali et le Groupe Monceau, et est 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).

Ses recherches sont diffusées via des séminaires mensuels ouverts aux universitaires et aux professionnels, la publication de working papers, des interventions lors de rencontres de place ainsi que des prises de position dans des revues professionnelles. Vous pouvez rester au contact de ces travaux par son site ou sa newsletter bimestrielle.

[www.chaire-pari.fr](http://www.chaire-pari.fr)



SciencesPo

