

LA TARIFICATION ET LE *BIG DATA* : QUELLES OPPORTUNITÉS ?

JEAN-MARC BOYER*

RAPPEL SUR LE *BIG DATA*

O n connaît les « 5 V » des données massives (numériques ou *not only SQL*¹) : le volume qui est le produit du nombre d'observations par le nombre de variables et s'exprime en zettaoctet (soit 10^{21} octets), la vitesse (ou vélocité) définie comme le nombre d'événements par seconde, la variété (complexité ou nombre de variables), la véracité et la valeur. L'important dans l'industrie financière, c'est l'exploitation des données rendue possible par l'accès aux données (par les réseaux Internet ou de type neural), les capacités de stockage (*cloud*), le traitement (puissance de calcul, baisse des coûts des bandes passantes), les accès (tablettes, objets connectés), les logiciels en *open source* plus fiables et plus soutenus, les langages de programmation ouverts (R, Python) et le *cluster* Hadoop (un *framework* en Java libre, qui permet le stockage et le traitement de très gros volumes de données).

Cette période est marquée par une accélération du temps (les chiffres doublent tous les deux ans comme la loi de Moore concernant les puces et les coûts de traitement informatique) et une contraction de l'espace (on est à deux clics du président des États-Unis). C'est une révolution digne de la relativité générale pour les scientifiques. Ne devrait-on d'ailleurs pas envisager un « congrès Solvay » virtuel sur le digital ? Mais si la « toile » est subie, l'Internet ressemblera, pour les plus poètes, au nénuphar des poumons de Chloé dans l'*Écume des jours* qui se réduit jusqu'à la fin du mode de vie actuel.

* Ancien délégué général, Fédération française des sociétés d'assurances (FFSA) ; président, Assurances Conseil Études Boyer (ACEB). Contact : jeanmarc.boyer1@gmail.com.

QUEL USAGE DU *BIG DATA* ? L'EXEMPLE DE L'ASSURANCE

Le champ d'opportunités techniques du *big data* est immense en ce qui concerne la tarification (et le marketing qui lui est lié) même si le *big data* et le digital modifient l'ensemble de la chaîne de valeurs, de la désintermédiation à l'accélération de la gestion.

Ainsi, le *big data* permet une tarification plus fine ainsi qu'un meilleur ajustement des garanties, l'évitement des doubles garanties, mais aussi une estimation des risques en stock en instantané de surcroît (par exemple, dans le domaine de la bancassurance, avec la corrélation du retrait carte bleue le samedi soir et du risque automobile pour un jeune conducteur). Le traitement des données de masse permet de minorer la sélection adverse (le risque est mieux connu) et les fraudes. Si l'on dispose de données fines, on peut améliorer le *zoning* (par exemple, le risque incendie en fonction de la vétusté du secteur) ou retarifier (par des objets connectés embarqués). La seule limite de la segmentation est la nécessaire mutualisation.

Le *big data* peut être utilisé pour l'estimation et la prévention des risques catastrophiques (en cas de pandémie ou d'inondation, par exemple). La prévision nécessite cependant des supercalculateurs hors de portée des assureurs.

82

Il peut permettre des prestations en temps réel (comme des indemnités guidées par les photos du postier envoyées aux réseaux sociaux). Mais attention, dans certains cas comme avec la généralisation du tiers payant dans le cadre de la loi Santé, on passe à un remboursement non incitatif et aveugle.

Il permet la création d'écosystèmes en architecture ouverte (choix de couvertures d'assurances par un intermédiaire ou inversement, choix des prestataires par un assureur). Dans ce « saucissonnage » de la valeur apportée, les assureurs risquent une pression concurrentielle accrue.

La détection et la limitation des fraudes font partie des potentialités liées au *big data* par la détection d'activités pour des assurés à qui l'on verse des indemnités journalières...

En marketing, le *big data* permet de connaître et même d'anticiper les besoins des clients et des prospects (les réseaux sociaux ne s'en privent pas, pour revendre des fichiers n fois) pour prédire les besoins non couverts. Cela implique une refonte des produits, compatible avec ces nouvelles technologies.

En gestion d'actifs, le *big data* peut fournir des enquêtes sur les investissements cibles. En quoi rebat-il les cartes de la tarification du risque de taux ? La tarification en assurance-vie (comme fixer le coût du

taux garanti, fut-il à zéro, en environnement de taux bas) dépend du passif (les engagements), mais aussi de l'actif (allocation de placements en couverture). La situation est symétrique au niveau des banques (tarification du risque de signature, mais aussi du refinancement au passif). La réglementation est antiprudentielle (elle pousse à détenir des titres d'État en considérant le risque souverain nul) et anti-économique : *value at risk* à un an, alors qu'il faut favoriser l'investissement long, comme le rapport de Villeroy de Galhau (2015) l'a compris.

En matière d'assurances, les risques cybernétiques de confidentialité et de sécurité portent sur les personnes et les entreprises. Les cyberattaques sont autant de risques que d'opportunités à couvrir par les assureurs. Mais attention à la confidentialité (cyberattaques ou même traces repérées par Google). On peut tarifer des risques même quand leurs lois statistiques sont inconnues comme pour les cyberrisques. Grâce au théorème central limite qui veut que si l'on assure en grand nombre, la probabilité de n variables indépendantes identiquement distribuées tend vers la loi de Gauss.

La tarification de la « prime pure » (produit de la fréquence et du coût moyen) étant une donnée partagée dans le monde entier, les assureurs devront faire la différence sur le reste : la résilience financière, l'efficacité de la gestion et des services et la fluidité du relationnel. En amont de la tarification, la valeur ajoutée résidera dans l'identification des besoins, les risques déjà assurés et le degré d'aversion selon les risques. L'enjeu est aussi de passer de l'indemnitaire à une combinaison de services et de prévention, et donc, à terme, de réinventer une partie du fonctionnement intrinsèque de l'assurance (y compris l'actuariel).

DES OPPORTUNITÉS

L'exemple de la tarification du risque (en assurance ou pour le risque de crédit)

Les entreprises bancaires et d'assurances sont par essence gestionnaires de données (qui est leur matière première) internes ou professionnelles historiquement (gestion de la relation client – GRC – avec le *data mining* – exploration de données) et désormais globales (*big data*).

L'interfaçage aux objets connectés

Les données peuvent désormais provenir d'objets connectés comme :
– le boîtier des contrats automobiles *Pay as you drive* (on peut ainsi voir se développer une assurance basée sur l'utilisation) ; avec des

véhicules vraiment automobiles (sans chauffeurs, comme annoncé par Carlos Ghosn pour 2016), ce boîtier ne serait plus un mouchard puisque c'est le véhicule (Google Car, par exemple) lui-même qui serait connecté ;

- l'utilisation de drones pour les sinistres catastrophes naturelles (comme l'assureur AIG aux États-Unis) ;
- des photos prises par les postiers pour le règlement des sinistres ;
- les capteurs biométriques de Health Apple, le bracelet Pulse de Withings utilisé par AXA, les lentilles de contact électroniques, les capteurs intracomprimés, les séquenceurs de poches (on peut décoder un ADN pour moins de 1 000 dollars) qui peuvent identifier quatre mille maladies génétiques, mais, aujourd'hui en France, les prédictions de pathologies sont proscrites.

L'interfaçage avec le digital

Avec les comparateurs

L'Association des sites internet communautaires (ASIC) peut s'alarmer des risques de législations contraignantes. Le Groupement des comparateurs d'assurance (GCA) a été lancé au début de 2015. Les sites évoluent à grande vitesse, comme avec les *new best actions* (NBA) : proposer un produit en fonction de quelques clics d'utilisateurs. Des chartes de comparateurs ont émergé à la suite des recommandations du Comité consultatif du secteur financier (CCSF, 2012).

84

Avec Internet

Avec notamment les comparateurs, l'asymétrie d'information entre client et prestataire financier s'estompe, et le client est pratiquement aussi informé que l'assureur ou le banquier. L'asymétrie d'information peut même être inversée dans le cas d'un particulier qui demande un décodage de son génome avec les probabilités de contracter telle ou telle maladie, alors que cela n'est pas accessible à l'assureur.

Avec les cyberrisques

Zurich Re y voit un marché colossal (plus de 1 % du PIB d'ici à 2020, à comparer aux 200 Md€ de chiffres d'affaires des assureurs en 2014 en France), mais les entreprises n'ont pas le réflexe de s'assurer.

Plus généralement, AXA et Microsoft lancent un « hackaton » (événement où des développeurs se réunissent pour faire de la programmation informatique collaborative, sur plusieurs jours) en France.

Avec l'« ubérisation » de l'économie, c'est plus l'usage que la propriété qui devront être assurés dans le cas d'un partage de véhicule (Uber, Blablacar, etc.) et aussi d'une habitation (Air BNB, la plus grande chaîne hôtelière virtuelle du monde), etc.

L'exploitation des données de masse suppose un maximum de variables (vingt mille variables, par exemple), quitte à en ajouter (par exemple, en construisant des agrégats sur fenêtres glissantes de temps ou de zones). Il vaut mieux un modèle simple avec beaucoup de données que l'inverse. Le modèle se construit en deux échantillons, le premier pour l'apprentissage automatique (machine *learning*), puis le deuxième pour la validation de la robustesse de la modélisation (par exemple, polynomiale) des données.

Ces opportunités nombreuses (objets connectés, interfaçage avec le digital « ubérisation » de la société) supposent un peu de technologie, mais que la France semble développer (cursus *big data*, investissements dans les *start-up*, haut niveau scientifique même s'il faut l'adapter, etc.). Quels sont alors les freins qui nous distinguent des Silicon Valleys ?

DES LIMITES MULTIPLES

Des limites éthiques

Sommes-nous dans la prévention (comme les détecteurs de fumée) ou dans la surveillance (avec le risque d'une société *big brother*) ? Nous devons gérer le paradoxe en termes de protection et d'exposition de la vie privée : on va beaucoup plus sur Internet, mais on souhaite plus de confidentialité de ses données.

Comment prendre en compte le droit à l'oubli (comme les cancers de plus de cinq ans dans la convention AERAS – s'assurer et emprunter avec un risque aggravé de santé) ou même le refus de savoir (moratoire sur l'utilisation du génome en France) ?

La prédictibilité des pathologies peut conduire à refuser des risques et, inversement, à refuser de s'assurer (aléa moral). Cette prédictibilité est quand même une opportunité de prévention. Des filles d'Asclépios, la France a préféré Panacée (le curatif) à Hygié (la prévention) par *lobbying* de l'offre sans doute.

Quelles sont les contraintes juridiques (sous l'autorité de la Commission nationale de l'informatique et des libertés – CNIL) et ordinales (confidentialité prônée par le corps médical) ?

La préservation de la confidentialité des données (et de la vie privée) est mise à mal. La supposée muraille de Chine entre les données contractuelles et la connaissance par le prestataire est plutôt une cloison de papier, à la japonaise.

La tarification à la personne (l'hypersegmentation) est contraire à la mutualisation des risques, mais cette segmentation peut avoir des vertus

de prévention. Développer une médecine personnalisée pour des thérapies plus efficaces permet au secteur de la santé de réaliser des économies.

Des limites étatiques

La tentation en France est de se doter d'une législation contraignante (deux lois Macron et Lemaire² sont en discussion sur le numérique, auxquelles s'ajoute la loi Valtier) avec le risque de brider une créativité qu'il faudrait au contraire encourager. Le principe du jardin à la française où il faut tout encadrer fiscalement par des taxes, des contrôles (par l'Autorité de régulation des communications électroniques et des postes – ARCEP – en l'occurrence), des cadres juridiques contraints : comparateurs classés courtiers, plates-formes de *crowdfunding* classées en conseillers en investissement participatif (CIP), prestataires en service d'investissement (PSI) ou intermédiaires en financement participatif (IFP) (ACPR et AMF, 2014).

Entrent également en jeu les conditions administratives pour être fiscalement déductibles (par exemple, concernant l'assurance-santé, les contrats « responsables »).

86 Les directives européennes interdisent de discriminer s'il n'y a pas de base statistique justificative et parfois même quand il y en a, comme l'interdiction de discrimination tarifaire homme/femme, ce qui donne lieu à des absurdités telles que la hausse du tarif automobile et des rentes pour les femmes.

On peut ajouter la tentation d'intervention des organismes parapublics comme l'AFNOR (ISO au plan mondial) qui produit un rapport sur la normalisation en matière de données massives.

Une autre tentation française consiste à demander des rapports (le rapport Lemoine au gouvernement, les plans prioritaires du ministre de l'Économie, les rapports de l'institut Louis Bachelier, etc.) plutôt que de privilégier l'action.

Etalab fournit gracieusement une dizaine de milliers de fichiers en *open source* (www.data.gouv.fr), mais certaines données restent à déverrouiller.

Des limites économiques

La révolution technologique du digital est « disruptive ». Depuis Schumpeter, on sait que cette disruption tend à détruire l'économie existante (*brick and mortar* en l'occurrence et principes actuariels de mutualisation). Un assureur qui investit dans une *start-up* donne l'impression de se tirer une balle dans le pied. En fait, il s'assure car en payant une modeste prime (investissement dans une Assur-

Tech), il se rembourse de la destruction de son *business model* si cette *start-up* devient une licorne (succès de plus de 1 Md\$ de capitalisation).

Des pressions de marché

Les banques auront un avantage concurrentiel encore accru, d'autant plus que la bancassurance est une spécificité française. Non seulement un banquier peut suivre les flux financiers sur un compte et donc proposer des placements au bon moment, mais aussi il pourra utiliser les données bancaires pour affiner son estimation des risques du client. Par exemple, si au distributeur automatique de billets, on constate des retraits en liquide le samedi soir, on peut supposer que le client sortira et sera un sujet à risque pour sa conduite. La Banque postale n'a pas tardé à avoir l'idée. L'histoire montre aussi qu'un mauvais risque assurance est un mauvais risque crédit, et *vice versa*.

Des pressions réglementaires

Solvabilité II exige une analyse plus exhaustive et formalisée des risques, rendue possible avec le *big data*. Dans le domaine bancaire, l'approche bâloise exige une estimation également plus poussée des risques de contrepartie. Avec le *big data*, on peut adjoindre au contenant des échanges de données, l'analyse même du contenu des données.

La directive sur la distribution d'assurance³ (DIA2) oblige à mieux cibler les besoins du client (ce que l'on peut cerner par l'utilisation d'un moteur de recherche). Sigma (Swiss Re Group) a produit à cet égard le rapport *Digital Distribution in Insurance*⁴. Cette connaissance du client s'étend de fait aux conseillers indépendants en gestion de patrimoine (qui subissent aussi la mise en œuvre de la directive Marché d'instruments financiers 2 ou MIF2). Si la transparence s'accroît, cela laisse espérer une plus grande part de conseil versée en honoraires au détriment des commissions qui peuvent biaiser le conseil. Les directives MIF2 et DIA2 vont ainsi modifier le cadre législatif de l'intermédiation (transparence, indépendance, interdiction ou non des rétrocessions suivant les versions de l'AMF – Autorité des marchés financiers – transcrivant celles de l'European Securities and Markets Authority – ESMA). Si un comparateur a un statut de courtier, il doit se plier aux règles de l'intermédiation (livre V du Code des assurances, ORIAS, devoir de conseil, dépendance ou non à un assureur ou à un organisme, etc.). On retombe sur la problématique classique : préconiser le meilleur rapport qualité/prix ou la meilleure commission ?

Les banquiers et les assureurs s'estiment protégés par des barrières à l'entrée (capitaux à mobiliser, agréments). En réalité, ils ne sont protégés que sur la partie de portage de risque. Pour la partie lucrative de la chaîne de valeurs (tenir le client), il suffit d'être courtier, CIF (conseiller en investissements financiers) ou intermédiaire en opérations de banque et de services de paiement (IOBSP). Par sa connaissance de l'internaute, Google pourrait donc déjà envahir le marché de l'intermédiation, mais risquerait de perdre le produit actuel de la vente des connexions.

Les assureurs seront chargés de l'identification des risques par le concept d'exposome, à savoir l'intégration de l'ensemble des expositions pour la vie entière. Il convient de se référer à la multiplicité des facteurs environnementaux (air, eau, sol, expositions professionnelles, alimentation, comportements individuels) susceptibles d'avoir un lien causal avec le développement de certaines pathologies.

L'AssurTech figure parmi les FinTech encouragés par Paris Europlace et une partie des pouvoirs publics. La diffusion de l'AssurTech oscille entre mystérieux (Aztech ?) et indispensable (Hastech ?). AXA (Kamet), Allianz (Riviera), MAIF (économie du partage), SFEREN (innovation) ont démarré leur « incubateur », sans compter les *data labs*, les fonds de *private equity* spécialisés, les partenariats avec les sociétés du digital, à commencer avec le GAFA (Google, Apple, Facebook, Amazon).

88

L'économie de l'assurance s'est transformée (apparition des mutuelles sans intermédiaires, de la bancassurance, explosion de l'assurance-vie, etc.), mais le tournant d'Internet (utilisé plus en consultation tarifaire qu'en souscription) n'a pas engendré de bouleversements. Si le *big data* des financiers ne révolutionnait pas l'écosystème, les intervenants du digital s'en chargerait (Orange s'invite dans les moyens de paiement).

La formation

Pour poursuivre l'essor de cette révolution industrielle au sens de Schumpeter, la formation est indispensable, comme celle de *data scientists* (lancée cette année par l'Institut des actuaires/Institut du Risk Management, l'Ensaë, Telecom Paris Tech, HEC, Dauphine, etc.). Mais il reste toujours la possibilité de se former en ligne grâce aux nombreux cours en ligne accessibles gratuitement.

La France est aussi enviée pour ses matheux que raillée pour ses bureaucrates. Mais les formations scientifiques sont à adapter (cryptage, anonymisation, logiciel *open source*, etc.). L'exploitation du *big data* suppose la conjonction de trois compétences : informatique, méthodes statistiques et connaissance métier. À défaut, on peut travailler en équipe (*pizza team*).

L'un des seuls postes où les assureurs investissent sans compter, c'est le digital (terme venant d'ailleurs de « compter sur ses doigts ») ou, plus largement, le numérique, c'est-à-dire les nombres issus de l'arithmétique des Grecs qui n'aiment plus compter.

CONCLUSION

On assiste à un changement de paradigme, impliquant une réingénierie des processus totale et permanente. Le *big data* et la révolution digitale ne sont pas un simple ajout à la mode, mais une refonte de l'approche des risques (donc du client), de leur appréciation (comité des risques), de leur gestion (en digital et en temps réel) et de leur provisionnement. Cette refonte suppose un changement d'organisation, d'organigramme, moins fréquent que les annonces d'un adjuvant (directeur du digital en plus, alors que ce sujet est transverse). Les services financiers sont plutôt dans leur phase d'apprentissage qu'en excès de traitements (*overfitting*). Le temps presse, dans un champ de prestations mondialisées : les États-Unis détiennent les clés du digital et l'Inde, deuxième marché de l'Internet, sera leur relais.

89

NOTES

1. *Not only* SQL ou NoSQL : systèmes de gestion de base de données qui n'est pas fondée sur l'architecture classique des bases relationnelles ; à l'origine, ces systèmes ont été utilisés pour des bases de données géantes telles celles des sites internet de très grande audience.
2. Voir le projet de loi pour une République numérique : www.republique-numerique.fr/consultations/projet-de-loi-numerique/consultation/consultation#opinion-type--titre-iii-l-acces-au-numerique.
3. Voir le site : http://ec.europa.eu/finance/insurance/consumer/mediation/index_fr.htm.
4. Voir le site : http://media.swissre.com/documents/sigma2_2014_en.pdf.

BIBLIOGRAPHIE

ACPR (Autorité de contrôle prudentiel et de résolution) et AMF (Autorité des marchés financiers) (2014), *S'informer sur le nouveau cadre applicable au financement participatif (crowd funding)*, 30 septembre, https://acpr.banque-france.fr/fileadmin/user_upload/acpr/Communication/Communiqués%20de%20presse/20140930-Sinformer_sur_le_nouveau_cadre_applicable_au_financement_participatif.pdf.

CCSF (Comité consultatif du secteur financier) (2012), « Avis du Comité consultatif du secteur financier pour renforcer la transparence et la qualité des comparateurs d'assurances de dommages sur Internet », 10 mai, www.banque-france.fr/ccsf/fr/publications/telechar/avis_r/avis-du-comite-consultatif-du-secteur-financier-renforcement-comparateurs-assurances-dommages-internet.pdf.

VILLEROY DE GALHAU F. (2015), *Le financement de l'investissement des entreprises*, rapport d'étape de Manuel Valls, Premier ministre de la République française, août.

