Gestion du risque de fraude interne au sein des banques : un panorama des nouvelles pratiques

Marc-Antoine LACROIX*

La révélation récente d'une fraude commise par un *trader* sur des produits dérivés chez UBS a de nouveau questionné la capacité des institutions financières à prévenir de manière efficace le risque de fraude interne.

Depuis les règles de Bâle II, qui reconnaissent la fraude parmi les risques opérationnels, les établissements développent une approche fondée sur le risque au-delà de l'approche traditionnelle en seule conformité. L'état d'avancement des établissements bancaires reste néanmoins assez hétérogène en pratique. La modélisation du risque de fraude interne laisse en effet encore la part belle aux modèles statistiques de type LDA (*loss distribution approach*), adaptés pour modéliser les fraudes répétées et de faible ampleur, mais moins performants pour les risques de fraude extrême.

Le recours complémentaire à l'analyse de scénario, fondée sur une approche en probabilité conditionnelle, peut s'avérer utile pour mieux comprendre l'exposition au risque de fraude interne extrême et tenter d'en réduire les causes.

Au-delà du débat sur le choix des méthodes, la nature des réponses apportées par les établissements reste influencée par les initiatives des régulateurs. Aux États-Unis par exemple, à la suite du scandale Madoff, le Congrès américain a renforcé le statut et le rôle dévolu aux « déclencheurs d'alerte » (*whistleblowers*) pour mieux détecter la fraude interne, soulevant la question de l'articulation souhaitable de ce type d'outils avec les dispositifs de surveillance et de contrôle.

* Directeur, Promontory.

Les points de vue exprimés dans cet article sont ceux de l'auteur et ne sauraient engager la société qui l'emploie.

LA FRAUDE INTERNE : UN ENJEU IMPORTANT

Par nature, la fraude est un acte de mauvaise fois, en général pour un profit personnel au détriment de l'entreprise. L'ACFE (Association of Certified Fraud Examiners) définit la fraude comme « l'utilisation de son propre emploi afin de s'enrichir personnellement tout en abusant ou en détournant délibérément des ressources ou des actifs de l'entreprise ». Les fraudes internes les plus connues vont du faux et usage de faux, vol caractérisé et extorsion de pièces au détournement de gages, usurpation d'identité, actions non autorisée, abus de blanc *seing*, en passant par l'abus de confiance ou la fraude informatique. Au sein des banques, on peut généralement regrouper les types de fraudes internes en cinq catégories : le détournement des avoirs de la clientèle, la « banque dans la banque » où les opérations sont détournées au profit du fraudeur, le détournement des avoirs de la banque, la création des fausses opérations, ou encore la personne qui fausse des objectifs pour augmenter sa rémunération. Les motivations et les conséquences de la fraude interne sont différentes en fonction des auteurs. Pour les employés, l'objectif est d'extraire des valeurs de l'entreprise à leur profit pour en bénéficier à des fins personnelles. Elles touchent l'ensemble des processus de l'entreprise. Pour le management, les fraudes perpétrées le sont en général pour tromper les actionnaires ou les investisseurs sur la situation de l'entreprise. Ce sont celles qui sont d'ailleurs les plus médiatisées (Enron, Worldcom).

Aux États-Unis, le coût annuel de la fraude (interne et externe) est estimé[1] à plus de 200 Md$ et 30 % des faillites de petites et moyennes entreprises résulteraient de la malhonnêteté des salariés. Selon des données recueillies au Canada, le coût de la fraude représenterait 5 % des revenus annuels des entreprises et 90 % des cas de fraude relèveraient de détournement d'actifs, les états financiers frauduleux ne représentant que 11 % des cas (mais beaucoup plus en valeur). Bien entendu, il convient d'ajouter à ces coûts directs les coûts indirects : pertes de parts de marché, baisse des cours de Bourse et d'image de marque auprès des fournisseurs, des clients et des salariés (baisse de la productivité, difficulté de recrutement).

DE L'UNIVERS DE L'AUDITEUR À CELUI DU *RISK MANAGER*

Le cadre prudentiel de Bâle II a reconnu le risque opérationnel comme une classe de risques à part entière à côté du risque de crédit et de marché et a inclus la fraude au sein de l'univers des types de risques opérationnels à prendre en compte.

Cette entrée de la fraude dans l'univers du *risk manager* a eu pour conséquence un élargissement de sa définition par rapport à l'approche traditionnelle de l'auditeur interne ou du commissaire au compte. Par exemple, les incidents de *rogue trading* seront considérés comme des événements de fraude interne en

présence de pertes importantes pour l'établissement, même si l'enrichissement personnel n'est pas toujours avéré (la perte pouvant provenir d'une erreur ou des conditions de débouclage des positions dissimulées, par exemple). La notion de perte pour l'établissement et de défaillance des contrôles se substitue ainsi au caractère intentionnel et malveillant comme critère de définition d'un incident de fraude. La frontière entre la fraude et l'erreur est parfois délicate à définir.

Autre difficulté, le classement entre fraude interne et externe n'est pas toujours aisé à établir. Dans l'affaire Madoff, par exemple, certaines institutions avaient distribué des fonds nourriciers de Madoff. Dans ce cas, il y avait complicité et les banques ont dû classer cet incident comme un événement de fraude interne. Pour les établissements n'ayant pas distribué de fonds nourriciers, à l'inverse, il n'y a pas complicité et l'événement a été classé comme une fraude externe. La notion de fraude interne devient donc une notion relative et dépend de la nature du lien entre l'établissement bancaire et l'événement de fraude.

LA QUALITÉ DE LA GOUVERNANCE : CLÉ DE VOÛTE D'UN DISPOSITIF EFFICACE DE PRÉVENTION

Les régulateurs rappellent régulièrement l'importance du rôle de la gouvernance comme premier rempart contre le risque de fraude interne. Le ton donné au sommet par le conseil d'administration et l'équipe de direction constitue en effet un signal déterminant sur la culture du risque de l'établissement et l'importance donnée aux principes éthiques, la définition des règles de conduite des affaires ainsi que l'adhésion à ces règles. Il appartient également au conseil d'administration d'approuver et de réviser l'appétence au risque opérationnel de l'établissement, celui-ci intégrant le risque de fraude interne. Cette tolérance au risque opérationnel doit articuler les niveaux et les différents types de risques que la banque est prête à prendre. Le conseil doit également vérifier que l'équipe dirigeante assure la bonne mise en œuvre du dispositif de lutte contre la fraude, à la fois sous l'angle des politiques, des procédures, des systèmes et des comportements, et ce, à tous les niveaux de l'organisation. Le recours à des revues externes afin de disposer d'un jugement indépendant sur l'efficacité de l'organisation de leur établissement face au risque de fraude est de plus en plus utilisé par les comités des risques et d'audit, sous l'égide des conseils d'administration.

En termes d'organisation, le dispositif de contrôle repose habituellement sur trois « lignes de défense » : la ligne métier, qui détient la première responsabilité de la détection et de la prévention de la fraude, la fonction risque et les fonctions de contrôle permanent (2ème ligne de défense) et enfin l'audit interne (3ème ligne de défense). Ces lignes de défense ne doivent pas être des lignes Maginot : elles doivent donc être dotées de ressources et d'équipes suffisantes pour opérer leurs missions.

CONNAÎTRE SON RISQUE : LE DÉBAT SUR LES MÉTHODES PERTINENTES DE MODÉLISATION DE LA FRAUDE INTERNE

Dans la gestion du risque opérationnel ouverte par Bâle II, le choix de la méthodologie s'avère crucial. Trois étapes constituent l'ossature habituelle d'une démarche réussie : l'évaluation initiale des risques de fraude (cartographie), la modélisation du risque lui-même et l'identification des leviers de réduction du risque. Si la démarche est robuste, les résultats obtenus doivent ensuite pouvoir servir de base aux plans d'action et être déclinés par l'ensemble des autres fonctions (métiers, contrôle) dans une démarche collaborative.

La démarche a toujours comme point de départ l'évaluation initiale du risque de fraude. La première étape est la production de l'univers des risques de fraude propre à la banque : détournement d'argent, dissimulation de pertes, détournement d'actifs, corruption, systèmes d'information, intelligence économique, falsification d'états financiers… Une fois l'univers des incidents connu, l'enjeu consiste à construire une cartographie de ces risques en les hiérarchisant et surtout en évaluant leur criticité (impact/probabilité). Certains risques de fraude sont des risques de fréquence : ils surviennent assez régulièrement, mais porte sur des petits montants. D'autres sont des risques de gravité : leur occurrence est rare, mais leur survenance peut avoir des conséquences catastrophiques pour la banque (comme le *rogue trading*). La qualité des bases de données utilisées (*reporting* des fraudes, historique de sinistralité, impact et coût financier, coût de gestion des risques) conditionne pour une large part la robustesse de l'exercice.

La modélisation du risque constitue la seconde étape. Il s'agit pour la banque d'évaluer les bons paramètres de risque de chacun des types de sinistres de façon non seulement à calculer l'exposition de la banque et à calculer un montant de fonds propres adéquat pour absorber le risque, mais également à identifier les bons leviers pour réduire ce risque lorsque c'est possible.

On comprend qu'en fonction des modèles utilisés, il est plus ou moins possible d'atteindre l'un ou l'autre de ces objectifs. Dans l'ensemble, pour le calcul des fonds propres au titre du risque opérationnel, nombreuses sont les banques qui utilisent des lois statistiques ajustées à partir d'historique de sinistres pour calculer leur perte « moyenne ». Ces modèles sont sans doute adaptés pour modéliser les fraudes répétées et de faible ampleur, mais s'avèrent moins performants lorsqu'il s'agit de mesurer la fraude de « queue de distribution », c'est-à-dire de faible occurrence, mais à fort impact (comme le *rogue trading*). Faute de pouvoir prendre en compte ces événements rares mais sévères, ces modèles peuvent aboutir à sous-estimer massivement le risque de perte encourue. Par exemple, l'intégration du seul incident Kerviel dans les modèles utilisant la LDA (*loss distribution approach*) aurait conduit à une augmentation des fonds propres réglementaire requis au risque du titre opérationnel de 20 %[2]. L'autre limite de ces modèles

est qu'ils ne sont guère « apprenants » : ils ne disent rien sur la causalité des incidents extrêmes, ni sur leur interaction : quelle est la probabilité pour qu'un risque de fraude apparaisse, sachant qu'une défaillance de système du contrôle interne aurait, par exemple, une chance sur dix d'apparaître au sein d'une banque ?

Le recours à des méthodes complémentaires de type analyse de scénario est alors utile pour éclairer la problématique spécifique des risques extrêmes, en se focalisant moins sur les qualités prédictives des modèles que sur leur qualité descriptive, afin de mieux comprendre les causes. Dans cette approche, le champ des données s'élargit : il s'agit d'utiliser les données internes ou externes, quantitatives, mais aussi qualitatives et de tenir compte non seulement des incidents de sinistres, mais aussi de l'environnement des affaires, de l'expérience des agents ainsi que du contrôle. Il est également possible d'inclure des hypothèses sur la situation future de l'établissement ainsi que des expériences de sinistres connues dans d'autres établissements. La démarche vise, en mobilisant des experts, à s'interroger sur les causes des sinistres pour mieux les prévenir. Dans la chaîne de causalité, les paramètres des modèles vont être par construction les leviers de l'action de réduction : par exemple, les fraudes sont d'abord liées aux personnes qui ont accès aux activités ; donc le nombre de personnes qui ont accès à telle ou telle information est un élément majeur pour évaluer la fréquence du risque de fraude. Ces critères sont d'autant mieux identifiés que les modèles ont bien été construits, avec les paramètres les plus pertinents. *De facto*, le risque diminue quand des actions de réduction sont mises en place et les valeurs des paramètres sont alors actualisées dans les évaluations. L'intérêt de cette approche par les causes est également d'aider à comprendre les interactions d'interdépendance entre les risques : par exemple, le fait qu'il existera une corrélation très forte entre des défaillances de contrôle individuel et la survenance d'un épisode de fraude interne de type *rogue trading*. En identifiant les causes des sinistres plutôt que leur simple distribution historique, cette méthode fournit en même temps les leviers de réduction de ces risques, qu'il s'agisse des *process*, des systèmes ou du facteur humain.

DÉTECTER LA FRAUDE INTERNE : LES BONS OUTILS

Si l'approche en risque décrite plus haut est robuste, il en découle une grande valeur ajoutée pour l'établissement qui pourra décliner les résultats obtenus pour prioriser les actions de contrôle par les unités en charge (conformité, sécurité financière, déontologie, audit interne notamment).

Dans les banques, on entend encore trop souvent dire que le hasard est le plus grand facteur de découverte des fraudes internes. En réalité, l'entreprise pourra aussi apprendre à trouver des solutions et des techniques (*data mining*,

profilage dans le respect des règles CNIL – Commission nationale de l'informatique et des libertés) pour améliorer le « rendement » des politiques de détection. Là encore, ces techniques ne sont efficaces que si leur utilisation s'articule avec une approche « en risque » décrite plus haut.

Ainsi, investir dans des systèmes lourds d'analyse des données n'assurera pas forcément une bonne efficacité si ces systèmes ne savent pas partager leurs informations, et encore moins le faire de manière pertinente, ou s'ils ne savent pas toujours se concentrer sur les zones ou les profils des collaborateurs les plus à risque. Pour prévenir la fraude de manière efficace, les personnes en charge de cela doivent donc être capables d'accéder et de se connecter à une quantité considérable de données, et ce, de manière transversale, dans toute l'entreprise, tant au niveau du service des ressources humaines qu'à celui en charge de la conformité ou à celui des activités de *trading*. Une bonne communication entre les systèmes et les services est absolument essentielle pour détecter une activité frauduleuse quel que soit son type.

FAUT-IL RENFORCER LE RÔLE DES *WHISTLEBLOWERS* ?

Les pratiques de détection évoluent enfin sous les impulsions du régulateur, comme en témoigne l'expérience américaine avec le rôle croissant reconnu aux *whistleblowers*.

Le déclenchement d'alerte est le geste accompli par un individu qui est témoin, dans son activité professionnelle, d'actes illicites et qui, par civisme, décide d'alerter les autorités ayant le pouvoir d'y mettre fin. Les Anglo-Saxons désignent ce geste par l'expression *whistleblowing* (« donner un coup de sifflet »). Le *whistleblowing* s'est vu récemment reconnaître une place accrue au sein de l'arsenal de prévention de la fraude aux États-Unis dans le cadre des suites de l'affaire Madoff.

Le Congrès a en effet adopté un dispositif permettant de rémunérer pour un montant compris entre 10 % et 30 % des sanctions (supérieures à 1 M$) ceux qui alimentent la SEC (Securities and Exchange Commission) en informations pertinentes permettant de détecter des cas de fraude. En créant une incitation monétaire à la dénonciation d'actes suspectés frauduleux, le Congrès va un cran plus loin que la loi Sarbannes-Oxley de 2002 qui se contentait de reconnaître le *whistleblowing* et prévoyait essentiellement des règles de protection. Un aspect délicat de la réforme américaine a trait à l'articulation du *whistleblower* avec les systèmes habituels de contrôle de la banque. En effet, certains observateurs ont pointé le risque d'une déresponsabilisation des structures en charge du contrôle face à l'incitation monétaire à signaler les cas de fraude au régulateur. Pire, compte tenu des seuils en présence, certains observateurs ont avancé que le *whistleblower* pourrait avoir intérêt à laisser le cas de fraude se développer pour ne le dénoncer

que lorsque sa taille permet d'entrer dans les seuils. En réponse à ces critiques, la SEC a mis au point un dispositif en cherchant à contenir ces risques. En particulier, la transmission d'informations ne peut être éligible à la récompense si elle provient des personnes dont la fonction est précisément de prévenir la fraude interne au sein de l'établissement (chargés de conformité, auditeurs internes, par exemple), avec des dérogations extrêmement limitées à ce principe. D'autre part, les règles allongent le délai (à cent vingt jours) sous lequel un *whistleblower* peut attendre avant d'alerter le régulateur sur un cas suspect après avoir averti les instances de contrôle interne. Enfin, le régulateur entend tenir compte de la coopération entre le *whistleblower* et les instances de contrôle comme critère de mesure de la récompense monétaire à attribuer. Un *whistleblower* pourra toucher la prime s'il ne transmet pas lui-même l'information au régulateur, mais si les instances de contrôle interne le font sur la base d'une information obtenue *via* le *whistleblower*.

Malgré les précautions prises par la SEC, le risque demeure d'une mise en concurrence des *whistleblowers* avec les instances de contrôle interne, qui pourrait déboucher sur une allocation inefficiente des efforts de détection si cette concurrence devait conduire à une multiplication des « faux positifs » (cas jugés suspects, mais ne donnant rien après enquête approfondie).

Faut-il recourir davantage aux *whistleblowers* en France et en Europe ? Outre les contraintes spécifiques au respect des données personnelles (CNIL), les obstacles culturels à la dénonciation demeurent nombreux. Une étude du cabinet Ernst & Young réalisée il y a quelques années montrait que seuls quatre salariés sur dix se sentiraient libres de signaler un éventuel cas de fraude à leur employeur. Le phénomène surgit avec vigueur, en revanche, dès que le salarié est licencié. Diverses études et enquêtes ont mis en évidence une faible propension des salariés français à dénoncer des fraudes par peur d'être licencié, voire pour ne pas mettre en péril l'activité économique de l'entreprise. Le recours aux « alertes » suppose de clarifier les questions de légitimité et de proportionnalité. La légitimité tout d'abord concerne la mise en place d'un dispositif d'alerte professionnelle qui doit se faire selon une obligation légale spécifique, ou un intérêt légitime, conforme aux droits et aux libertés fondamentaux (principes rappelés par les directives européennes). Il en découle que les finalités des dispositifs d'alerte professionnelle doivent être spécifiées et explicites. Ces dispositifs doivent établir un équilibre entre proportionnalité, subsidiarité et fiabilité des faits dénoncés. La proportionnalité ensuite renvoie à la question de la limitation possible du nombre de personnes chargées de rapporter les dysfonctionnements ainsi que le type d'informations qui peuvent être dénoncées. Au total, un premier retour d'expérience sur le dispositif américain sera utile pour juger de l'opportunité d'aller plus loin dans le recours à ce type d'outils en France et en Europe.

Ce rapide aperçu des pratiques récentes mises en œuvre au sein des banques

pour prévenir et gérer le risque de fraude interne dessine finalement trois approches assez différentes au sein des banques :

– l'approche traditionnelle « en conformité », qui privilégie le respect des procédures, des contrôles et des règles pour créer un environnement de contrôle efficace ;

– l'approche nouvelle « en risque », qui pousse d'abord à mieux connaître son univers de risque de fraude interne de façon à pouvoir prioriser ses actions de réduction et ses ressources de détection ;

– l'approche par les « incitations », qui consiste à rémunérer les personnes qui transmettent aux régulateurs des informations sur des cas suspects et peuvent se voir récompenser en cas de fraude confirmée.

À terme, il y a sans doute de la place pour chacune de ces trois approches. La question essentielle est surtout celle de leur hiérarchie et de leur articulation au sein de la banque pour assurer au dispositif de prévention une efficacité maximale. De ce point de vue, la responsabilité ultime repose sur la gouvernance de l'établissement et sur ses dirigeants, à qui il revient de doser le juste recours à ces approches de manière optimale en fonction des caractéristiques et des objectifs stratégiques de l'établissement.

Notes

1. Cabinet Grant Thornton.

2. En intégrant le coût lié au débouclage des positions dissimulées.