

**Mémoire présenté devant l'Université de Paris-Dauphine
pour l'obtention du Certificat d'Actuaire de Paris-Dauphine
et l'admission à l'Institut des Actuaire**

le

Par : Pauline CHAUVÉAU

Titre : Impact de la conjoncture économique sur l'activité d'un assureur-crédit

Confidentialité : ☒ Non ☐ Oui (Durée : ☐ 1 an ☐ 2 ans)

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

*Membres présents du jury de l'Institut
des Actuaire :*

Entreprise :
Nom : Forvis Mazars
Signature :

*Membres présents du Jury du Certificat
d'Actuaire de Paris-Dauphine :*

Directeur de Mémoire en entreprise :
Nom : SPEISSER Romain
GUENDEHOU Godfried
Signature :

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

Secrétariat :

Signature du responsable entreprise

Signature du candidat

Bibliothèque :

Résumé

L'environnement économique actuel, marqué par une forte inflation, des taux d'intérêts qui, après avoir été très bas pendant plusieurs années, ont atteint un niveau plus vu depuis plus de 10 ans, et des conflits géopolitiques, ainsi que le rattrapage du « retard de défaillances » observé pendant la période Covid, met aujourd'hui à l'épreuve les assureurs-crédit.

Les méthodes actuarielles traditionnellement utilisées pour prédire la survenance des défauts se basent principalement sur des scores affectés aux entreprises, censés refléter la solidité financière de celles-ci. Cependant, ces modèles peinent à intégrer en temps réel les évolutions de l'environnement économique. Dans ce mémoire, nous proposons une modélisation de la fréquence des défauts par secteur d'activité en utilisant les processus de Hawkes, permettant de modéliser des effets de contagion tant intra-sectoriel qu'inter-sectoriels.

Après une présentation de l'assurance-crédit et de ses spécificités, nos travaux se concentreront sur la capacité des processus de Hawkes à modéliser la structure de dépendance des défaillances entre secteurs d'activité. Nous développerons ensuite un modèle utilisant ces processus pour modéliser les pertes attendues d'un assureur-crédit fictif. Enfin, à l'aide du modèle développé, nous étudierons l'impact de la conjoncture économique sur les pertes de l'assureur en fonction de la répartition de son activité dans les différents secteurs via divers scénarios économiques. Une attention particulière sera portée sur la capacité du modèle à évaluer l'ampleur de l'effet de rattrapage de la crise Covid sur le nombre de défaillances attendues d'ici la fin de l'année. Les résultats obtenus permettront de mettre en évidence l'intérêt de la prise en compte des effets de contagion sectoriels dans un cadre de gestion des risques.

Mots-clés : Assurance-crédit, défaillances d'entreprise, dépendance sectorielle, processus de Hawkes, risque de contagion.

Abstract

The current economic environment, marked by high inflation, interest rates that, after being very low for several years, have reached levels not seen in over 10 years, and geopolitical conflicts, as well as the catch-up of the “backlog of defaults” observed during the Covid period, is currently testing credit insurers.

The actuarial methods traditionally used to predict defaults are primarily based on scores assigned to companies, which are supposed to reflect their financial strength. However, these models struggle to integrate real-time changes in the economic environment. In this paper, we propose a modeling of the frequency of defaults by sector using Hawkes processes, which allow for the modeling of contagion effects both within and between sectors.

After presenting credit insurance and its specificities, our work will focus on the ability of Hawkes processes to model the dependency structure of defaults between sectors. We will then develop a model using these processes to estimate the expected losses of a fictitious credit insurer. Finally, using the developed model, we will study the impact of the economic situation on the insurer’s losses based on the distribution of its activities across different sectors through various economic scenarios. Particular attention will be paid to the model’s ability to assess the extent of the catch-up effect of the Covid crisis on the expected number of defaults by the end of the year. The results obtained will highlight the importance of considering sectoral contagion effects within a risk management framework.

Keywords : Credit Insurance; business defaults; sectoral dependance; Hawkes processes; contagion risk.

Note de Synthèse

Les récents bouleversements économiques, marqués par une inflation persistante, une remontée des taux d'intérêt, des tensions géopolitiques et les conséquences de la pandémie de Covid-19, ont fragilisé le tissu économique mondial. Ces évolutions ont mis en lumière la vulnérabilité des entreprises face à des risques financiers croissants, influençant directement l'activité des assureurs-crédits. Ces derniers doivent anticiper et gérer des risques dans un contexte incertain, dont les effets sont amplifiés par l'interdépendance entre secteurs économiques.

Ce mémoire propose d'exploiter les processus de Hawkes pour modéliser la fréquence des défaillances d'entreprise par secteur d'activité, en intégrant les dépendances intra et intersectorielles. L'objectif est de fournir une approche complémentaire aux modèles existants, offrant une vision dynamique des risques d'un assureur-crédit en réaction aux fluctuations économiques. La seconde partie applique ce modèle pour évaluer les pertes attendues d'un assureur-crédit fictif selon divers scénarios économiques.

Contexte et motivations

L'assurance-crédit, qui couvre le non-paiement des factures entre entreprises, est fortement influencée par les fluctuations économiques et politiques. Les assureurs-crédits doivent anticiper les défaillances de leurs clients, souvent liées à des facteurs de risque systémiques. Ces défaillances ont tendance à se produire simultanément, en raison de l'exposition commune des entreprises aux mêmes facteurs de risque économiques et des relations commerciales entre entreprises.

Les modèles actuels, basés sur des approches de type *Merton*, reposent sur des prévisions économiques à un instant donné et nécessitent des mises à jour régulières pour rester pertinents. Ils n'intègrent pas les fluctuations économiques en temps réel. Dans un contexte d'incertitude économique, comme celui engendré par la récente crise du COVID-19, la capacité à mesurer l'impact de la conjoncture sur la distribution des pertes est essentielle pour un assureur. Or, la nécessité de recalibrer ces modèles à chaque évolution de l'environnement peut s'avérer coûteuse et inefficace.

Face à ces défis, ce mémoire propose une approche innovante basée sur les processus de Hawkes pour modéliser la fréquence des défaillances par secteur d'activité. Ces processus permettent de capturer les effets de contagion aussi bien intra qu'intersectoriels, offrant une modélisation dynamique des risques en fonction de l'évolution économique.

Le choix de cette approche repose sur le constat que les défaillances d'entreprises au sein d'un même secteur sont corrélées en raison de leur exposition à des risques communs. Par ailleurs, des interdépendances existent entre secteurs, liées aux échanges commerciaux et aux interactions au sein des chaînes d'approvisionnement. Les processus de Hawkes permettent précisément de modéliser ces dynamiques. Leur intensité de base sera conditionnée par des indicateurs économiques, tandis que leurs propriétés d'excitation permettront de représenter les effets de contagion intra et intersectoriels, intégrant ainsi la dépendance entre défauts.

Processus de Hawkes

Un processus de Hawkes est un processus de comptage, au même titre que le processus de Poisson, dont la particularité est que son intensité conditionnelle est un processus stochastique qui possède des propriétés d'excitation conditionnées au passé du processus de Hawkes. Autrement dit, chaque événement survenu dans le passé peut augmenter le taux d'occurrence des événements futurs.

L'intensité conditionnelle d'un processus de Hawkes, qui représente le taux de survenance instantané d'un événement à chaque instant t , prend la forme suivante

$$\lambda^*(t) = \mu(t) + \int_0^t \phi(t-s) dN_s = \mu(t) + \sum_{T_n < t} \phi(t - T_n),$$

où $\mu : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ représente l'intensité de base du processus, $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ est la fonction d'excitation et $(T_n)_{n \geq 0}$ sont les temps de sauts du processus.

L'intensité conditionnelle d'un processus de Hawkes est donc constituée de deux composantes. Une première composante déterministe $\mu(\cdot)$ représentant le risque de base du processus et une seconde composante *auto-excitante* représentée par la somme sur les T_n . Cette dernière traduit l'influence du passé du processus sur l'intensité conditionnelle.

La définition se généralise à un cadre multivarié, en considérant d processus, l'intensité conditionnelle de chaque processus $k \in \{1, \dots, d\}$ est définie par

$$\lambda_k^*(t) = \mu_k(t) + \sum_{j=1}^d \int_0^t \phi_{k,j}(t-s) dN_s^{(j)} = \mu_k(t) + \sum_{j=1}^d \sum_{T_n^{(j)} < t} \phi_{k,j}(t - T_n^{(j)}),$$

où $\mu_k : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ et $\phi_{k,j} : \mathbb{R}^+ \rightarrow \mathbb{R}^+$.

Chaque événement passé, associé à l'un des processus constituant $(N_t)_{t \geq 0}$, a une influence sur l'intensité du processus $(N_t^{(k)})_{t \geq 0}$. Dans la double somme, $j = k$ caractérise le phénomène d'auto-excitation : les événements passés du processus lui-même impactent son intensité. En revanche, $j \neq k$ caractérise les effets d'inter-excitation : les événements passés des autres processus influencent l'intensité de $(N_t^{(k)})_{t \geq 0}$.

Application à l'assurance-crédit

Pour modéliser la fréquence des défauts d'un assureur-crédit par un processus de Hawkes multivarié, nous considérons que chaque processus représente le nombre de défauts d'un groupe homogène d'entreprises. N'ayant pas à disposition de données réelles d'assureur-crédit, nous avons basé notre analyse sur les données publiques de défaillances d'entreprises, publiées mensuellement par la Banque de France. Nous avons regroupé les entreprises en cinq secteurs d'activité (*Agriculture, Industrie, Construction, Enseignement, santé humaine et services aux ménages, et Services marchands*), en supposant que les entreprises d'un même secteur sont exposées aux mêmes facteurs de risque. Cependant, la méthodologie détaillée dans ce mémoire se veut généralisable à des regroupements plus fins, permettant une éventuelle application de cette approche à des données réelles d'assureur-crédit.

Modèle initial

Dans un premier temps, nous avons développé un modèle afin d'étudier la capacité des processus de Hawkes à modéliser les effets de contagion entre les défauts, avec l'hypothèse forte que l'environnement

économique serait implicitement capturé par le modèle à travers l'évolution du nombre de défaillances dans chaque secteur d'activité.

Le modèle repose sur un processus de Hawkes multivarié $N = (N^{(1)}, \dots, N^{(d)})$ dont l'intensité conditionnelle de chaque processus $N^{(k)}$ est définie par

$$\lambda_k^*(t) = \lambda_k + \sum_{j=1}^d \sum_{T_n^{(j)} < t} \alpha_{k,j} e^{-\beta_k(t-T_n^{(j)})}.$$

Chaque processus de comptage $(N_t^{(k)})_{t \geq 0}$ enregistre le nombre de défaillances dans un secteur d'activité, où $k = 1, \dots, d$ et d sont le nombre total de secteurs d'activité.

Le risque de base du processus λ_k dans chaque secteur k est constant. Les coefficients $\alpha_{k,j}$ quantifient l'impact d'un défaut dans le secteur j sur le taux de défaut du secteur k , modélisant ainsi la dépendance intersectorielle. Enfin, les paramètres β_k contrôlent la durée de cet effet.

Ce modèle s'est révélé insatisfaisant : il semble converger vers un niveau constant de défaillances mensuelles, indépendamment du point de départ des prédictions. Cette tendance est observée malgré le fait que l'historique du processus simulé ait été ajusté pour contenir l'historique de défaillances réellement observé avant la date de début de prédiction. Ce comportement semble lié à l'hypothèse d'une intensité de base constante λ_k , limitant ainsi la capacité du modèle à prédire en dessous des niveaux fixés par ces paramètres, un problème particulièrement notable sur la période post-COVID.

Néanmoins, les résultats des calibrages, tant en univarié qu'en multivarié, indiquent des paramètres $\alpha_{k,j}$ non nuls, soutenant l'hypothèse d'existence d'effets de contagion intra et inter-sectoriels. Cela nous a conduit à penser qu'en ajustant les paramètres d'intensité de base, le modèle pourrait fournir de meilleurs résultats.

Modèle final

Nous avons donc développé un second modèle, où l'intensité de base λ devient une fonction dynamique $\mu(\cdot)$ dépendant d'indicateurs économiques.

Le cadre reste inchangé, mais λ_k est remplacé par $\mu_k(\cdot)$, intégrant directement l'environnement économique dans l'évolution du risque de défaut sectoriel. Ainsi, la dynamique économique dicte la tendance générale du risque, tandis que la composante d'excitation du processus capture les effets d'accumulation et de contagion.

L'intensité conditionnelle est désormais définie par

$$\lambda_k^*(t) = \mu_k(t) + \sum_{j=1}^d \sum_{T_n^{(j)} < t} \alpha_{k,j} e^{-\beta_k(t-T_n^{(j)})}, \quad \forall k = 1, \dots, d.$$

Les paramètres $\alpha_{k,j}$ et β_k conservent leur rôle. La fonction $\mu_k(\cdot)$, qui représente l'intensité de base du processus $(N_t^{(k)})_{t \geq 0}$, est construite par régression linéaire sur différents indicateurs macroéconomiques et sa valeur fluctue en fonction de l'évolution de ces indicateurs.

Résultats

Pour évaluer la performance du modèle final, nous avons réalisé plusieurs simulations sur des périodes passées, en prolongeant l'historique des données. La moyenne empirique des trajectoires simulées est retenue comme prédiction.

La précision des prévisions est mesurée via la *Mean Absolute Percentage Error* (MAPE), une métrique adaptée aux variations importantes dans le nombre de défaillances au fil du temps et entre différents secteurs. En normalisant les erreurs par rapport aux valeurs observées, elle permet une comparaison des performances à travers différentes échelles. Le tableau ci-dessous présente les MAPE calculées pour différentes périodes de prédiction.

Période de prédiction	(A)	(I)	(C)	(E)	(S)	Moyenne
01/2022 - 06/2024	0.1042	0.1337	0.0780	0.1209	0.0675	0.1009
01/2011 - 12/2023	0.1279	0.1418	0.1368	0.1392	0.1097	0.1297
01/2021 - 12/2022	0.1320	0.2994	0.1723	0.3281	0.2321	0.2522
01/2015 - 06/2024	0.1317	0.1745	0.1386	0.1578	0.1121	0.1429

Bien que globalement satisfaisant, le modèle peine à capturer la forte baisse des défaillances observée durant la crise sanitaire et la période post-COVID, en grande partie due aux mesures de soutien gouvernementales. En janvier 2021, après une année marquée par un niveau exceptionnellement bas de défaillances, l'intensité conditionnelle du modèle reste surestimée. La décroissance de la composante d'excitation, bien que présente, demeure insuffisante pour compenser la surévaluation de l'intensité de base à cette période, limitant ainsi l'ajustement du modèle à la réalité observée.

Les performances se dégradent nettement entre 2020 et 2022, comme en témoigne l'augmentation des MAPE sur les périodes incluant la crise sanitaire. Toutefois, sur un horizon plus long, le modèle fournit des résultats cohérents, suggérant que la précision des prévisions dépend fortement de l'estimation de la fonction $\mu(\cdot)$. Intégrer des indicateurs sectoriels plus spécifiques pourrait améliorer cette estimation et renforcer la robustesse des prévisions à long terme.

Le tableau ci-dessous compare la performance du modèle final à plusieurs alternatives : le modèle initial, un processus de Poisson non homogène et un modèle « naïf » basé sur la moyenne empirique des données. L'objectif est d'évaluer les améliorations apportées par l'intégration des facteurs économiques dans l'intensité de base et également de mesurer l'apport de la composante d'excitation. Le modèle « naïf » sert de point de comparaison pour évaluer l'efficacité des autres modèles plus complexes.

Modèle	(A)	(I)	(C)	(E)	(S)	Moyenne
Hawkes (2.22)	0.1247	0.1319	0.1417	0.1376	0.0951	0.1262
Hawkes (2.18)	0.1259	0.1816	0.2143	0.1333	0.1441	0.1598
Poisson	0.1259	0.1546	0.1481	0.2107	0.1015	0.1476
Naïf	0.1573	0.2599	0.2556	0.3024	0.1665	0.2283

MAPE des différents modèles sur la période 2011-2024

Les résultats du tableau montrent que le modèle final de Hawkes offre la meilleure précision, avec une MAPE plus faible que les autres approches. L'intégration des facteurs économiques améliore nettement les performances par rapport au premier modèle de Hawkes.

Le processus de Poisson non-homogène fournit des résultats convenables, mais reste moins précis que le modèle final. La contribution de la composante d'excitation au sein de notre modèle se révèle donc bien réelle, bien qu'elle joue un rôle moins significatif que prévu.

Impact du contexte économique sur les pertes potentielles d'un assureur-crédit fictif

L'objectif est d'expliquer comment les travaux réalisés sur les processus de Hawkes peuvent être utilisés pour modéliser les pertes potentielles. Notre modélisation se concentre uniquement sur les pertes attendues associées aux contrats dénommés. Par ailleurs, compte tenu des données disponibles publiquement, nous avons uniquement modélisé les défauts résultant de l'insolvabilité totale des entreprises, ce qui exclut les défauts de type *Protracted*.

Modèle simplifié

La perte attendue sur la période $[0, t]$ est donnée par

$$L_t = \sum_{k=1}^d \tilde{N}_t^{(k)} \times \bar{E}_k \times \sigma$$

où :

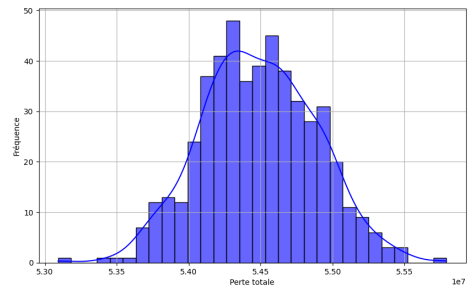
- $\tilde{N}_t^{(k)} = N_t^{(k)} \times p_k$ représente les défauts affectant le portefeuille de l'assureur dans le secteur k , avec $N_t^{(k)}$ modélisé par un processus de Hawkes ;
- \bar{E}_k est l'exposition moyenne garantie dans le secteur k ;
- $\sigma = 6.8\%$ correspond au taux de perte moyen des sinistres.

L'exposition sectorielle a été estimée à partir des rapports annuels de Coface et des bases de données INSEE. La cohérence du modèle a été validée par comparaison avec les pertes déclarées par Coface entre 2021 et 2023. Le tableau ci-dessous présente les données d'exposition construites pour l'année 2024.

Secteur d'activité	n_k	E_k	\bar{E}_k	p_k
(A)	71 521	11 564 500 000	161 693.76	0.1842
(I)	10 583	23 533 000 000	2 223 665.74	0.0297
(C)	19 574	7 423 500 000	379 256.67	0.0298
(S)	40 843	24 502 600 000	599 916.16	0.0117
(E)	33 116	2 171 500 000	65 573.46	0.0032
Total	175 637	69 195 100 000	393 967.08	0.0278

Prévisions à fin 2024 et allocation optimale

La figure à droite présente la distribution de la perte attendue pour notre assureur-crédit fictif pour l'année 2024. Il convient de rappeler que la seule composante stochastique du modèle est la fréquence des défaillances, laquelle est modélisée par un processus de Hawkes multivarié. Les résultats présentés sont issus de 500 simulations du modèle final. Selon les prévisions du modèle, la perte attendue moyenne pour l'année 2024 s'élève à 54.5 M€, avec un intervalle de confiance à 95% s'étendant de 53.7 M€ à 55.3 M€. Cette estimation représente une augmentation



de 17,2% par rapport à l'année 2023. Cet écart résulte exclusivement de l'évolution des défaillances, toutes choses étant égales par ailleurs.

En outre, si l'on considère le montant de la perte attendue dans un scénario où le niveau de défaillances serait resté proche de celui observé avant la pandémie de Covid-19, la perte moyenne aurait été de 49 840 180 €, ce qui correspond à un écart relatif de 9,3%.

Enfin, nous avons analysé la répartition sectorielle de l'exposition de notre assureur fictif et exploré des pistes d'optimisation pour minimiser les pertes attendues en 2024, en maintenant un volume de primes constant. L'objectif était de déterminer l'allocation optimale de cette exposition dans les différents secteurs dans le contexte actuel, avec pour objectif principal la minimisation des pertes attendues pour l'année 2024.

Le problème d'optimisation est le suivant

$$\begin{cases} \min_{\theta \in \mathbb{R}^d} \sum_{k=1}^d x_k \theta_k, \\ s.c : \sum_{k=1}^d \theta_k = 1, \quad |\theta_k - \theta_k^0| \leq s \text{ et } \theta_k \geq 0, \forall k, \end{cases}$$

où θ_k représente la part d'exposition optimisée pour le secteur k , θ_k^0 l'allocation initiale, et s définit la tolérance maximale autorisée pour la variation entre l'allocation initiale et l'allocation optimale.

Nous avons étudié deux scénarios pour 2024 : un scénario central basé sur la moyenne des simulations et un scénario pessimiste correspondant au quantile à 99,5%. Dans les deux cas, nous avons observé que la tendance des défaillances est à la hausse dans tous les secteurs. L'optimisation favorise donc les secteurs agricoles et des services non marchands, où les risques sont plus faibles, et réduit l'exposition dans l'industrie et la construction où le rapport entre l'exposition et le nombre de défauts est moins favorable pour l'année 2024.

Ainsi, bien qu'une redistribution de l'exposition puisse, théoriquement, constituer une stratégie intéressante pour piloter le risque d'un assureur-crédit, notre modélisation simplifiée conduit à une allocation maximale vers les secteurs présentant le rapport le plus favorable entre exposition moyenne et nombre de défauts et dépend des capacités réelles de l'assureur à ajuster sa présence sur certains marchés, en tenant compte des contraintes de diversification et des réalités du marché de l'assurance. En examinant l'allocation des secteurs de Coface au cours des trois dernières années, on constate qu'elle a varié de moins de 1% sur chaque secteur chaque année, ce qui souligne une inertie potentielle dans les ajustements des expositions.

Conclusion

Ce mémoire a exploré l'impact de la conjoncture économique sur l'activité d'un assureur-crédit fictif à travers une modélisation dynamique des défauts basée sur les processus de Hawkes. Ces derniers se sont révélés pertinents pour capturer les interactions entre secteurs, bien qu'une segmentation plus fine (exposition, notation, localisation) puisse améliorer leur précision.

Le modèle développé intègre l'environnement de risque propre à chaque secteur à travers les fonctions d'intensité de base et la dépendance entre défauts via les paramètres d'excitation, offrant une prédiction dynamique tenant compte d'évolutions économiques sans nécessité de réajustements. Toutefois, cette approche favorise une modélisation à court terme, bien qu'un couplage avec un générateur de scénarios économiques puisse permettre des prévisions à long terme. Le modèle a montré des limites pour modéliser les défauts durant la crise du Covid, principalement à cause de la difficulté à

intégrer les mesures de soutien de l'État. Cette surestimation a néanmoins permis d'estimer l'effet de rattrapage attendu.

Les processus de Hawkes présentent néanmoins des limites : la complexité croissante avec le nombre de variables et une forte sensibilité aux données, impactant l'estimation des paramètres et les prédictions. Malgré ces défis, le modèle de pertes a produit des résultats cohérents avec les pertes réelles sur trois ans. Enfin, l'optimisation de l'allocation sectorielle de l'assureur a montré des résultats intéressants mais une flexibilité limitée par les contraintes du marché, réduisant la capacité d'ajustement de l'assureur.

Synthesis note

The recent economic upheavals, marked by persistent inflation, rising interest rates, geopolitical tensions, and the consequences of the Covid-19 pandemic, have weakened the global economic fabric. These developments have highlighted the vulnerability of businesses to growing financial risks, directly impacting the activities of credit insurers. The latter must anticipate and manage risks in an uncertain environment, where the effects are amplified by the interdependence between economic sectors.

This paper aims to leverage Hawkes processes to model the frequency of business failures by industry sector, incorporating both intra- and inter-sectoral dependencies. The objective is to provide a complementary approach to existing models, offering a dynamic perspective on the risks faced by a credit insurer in response to economic fluctuations. The second part applies this model to assess the expected losses of a fictional credit insurer under various economic scenarios.

Context and motivations

Credit insurance, which covers the non-payment of invoices between businesses, is highly influenced by economic and political fluctuations. Credit insurers must anticipate the defaults of their clients, which are often linked to systemic risk factors. These defaults tend to occur simultaneously due to companies' shared exposure to the same economic risk factors and the commercial relationships between firms.

Current models, based on Merton-type approaches, rely on economic forecasts at a given point in time and require regular updates to remain relevant. However, they do not incorporate real-time economic fluctuations. In times of economic uncertainty, such as the recent COVID-19 crisis, the ability to measure the impact of economic conditions on the loss distribution is crucial for insurers. Yet, the need to recalibrate these models with every shift in the economic environment can be both costly and inefficient.

To address these challenges, this paper proposes an innovative approach based on Hawkes processes to model the frequency of business failures by industry sector. These processes allow for the capture of contagion effects both within and between sectors, providing a dynamic modeling framework that adjusts to economic developments.

The choice of this approach is based on the observation that business failures within the same sector are correlated due to their exposure to common risks. Additionally, interdependencies exist between sectors, driven by trade relationships and interactions within supply chains. Hawkes processes are particularly well-suited for modeling these dynamics. Their baseline intensity will be conditioned by economic indicators, while their self-exciting properties will allow the representation of intra- and inter-sector contagion effects, thereby integrating the dependence between defaults.

Hawkes Processes

A Hawkes process is a counting process, similar to the Poisson process, with the key distinction that its conditional intensity is a stochastic process exhibiting self-exciting properties influenced by its past occurrences. In other words, each past event can increase the likelihood of future events occurring.

The conditional intensity of a Hawkes process, representing the instantaneous occurrence rate of an event at any time t , is given by

$$\lambda^*(t) = \mu(t) + \int_0^t \phi(t-s) dN_s = \mu(t) + \sum_{T_n < t} \phi(t-T_n),$$

where $\mu : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ represents the basic risk of the process, $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is the excitation function and $(T_n)_{n \geq 0}$ are the jump times of the process.

The conditional intensity of a Hawkes process thus consists of two components. A deterministic component $\mu(\cdot)$ representing the basic risk and a self-exciting component, represented by the sum over T_n , which accounts for the influence of past events on the conditionnal intensity.

This definition extends to the multivariate setting, we consider d processes, the conditional intensity of each process $k \in \{1, \dots, d\}$ is defined as

$$\lambda_k^*(t) = \mu_k(t) + \sum_{j=1}^d \int_0^t \phi_{k,j}(t-s) dN_s^{(j)} = \mu_k(t) + \sum_{j=1}^d \sum_{T_n^{(j)} < t} \phi_{k,j}(t-T_n^{(j)}),$$

where $\mu_k : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ and $\phi_{k,j} : \mathbb{R}^+ \rightarrow \mathbb{R}^+$.

Each past event, associated with one of the processes in $(N_t)_{t \geq 0}$, influences the intensity of the process $(N_t^{(k)})_{t \geq 0}$. In the double sum, $j = k$ characterizes the self-excitation phenomenon, meaning that past events of the process itself impact its intensity. In contrast, $j \neq k$ characterizes cross-excitation effects, where past events from other processes influence the intensity of $(N_t^{(k)})_{t \geq 0}$.

Application to Credit Insurance

To model the default frequency of a credit insurer using a multivariate Hawkes process, we consider each process as representing the number of defaults within a homogeneous group of companies. Since we do not have access to real credit insurer data, our analysis is based on publicly available business failure data, published monthly by the Banque de France. We have grouped businesses into five industry sectors (Agriculture, Industry, Construction, Education, Human Health and Household Services, and Market Services), assuming that companies within the same sector are exposed to the same risk factors. However, the methodology outlined in this thesis is designed to be generalizable to more granular groupings, allowing for a potential application of this approach to real credit insurer data.

Initial model

Initially, we developed a model to study the ability of Hawkes processes to model contagion effects between defaults, with the strong assumption that the economic environment would be implicitly captured by the model through the evolution of the number of failures in each sector.

The model relies on a multivariate Hawkes process $N = (N^{(1)}, \dots, N^{(d)})$, where the conditional intensity of each process $N^{(k)}$ is defined by

$$\lambda_k^*(t) = \lambda_k + \sum_{j=1}^d \sum_{T_n^{(j)} < t} \alpha_{k,j} e^{-\beta_k(t-T_n^{(j)})}.$$

Each counting process $(N_t^{(k)})_{t \geq 0}$ records the number of defaults in a sector, where $k = 1, \dots, d$ and d is the total number of sectors.

The basic risk λ_k of the process in each sector k is constant. The coefficients $\alpha_{k,j}$ quantify the impact of a default in sector j on the default rate of sector k , thus modeling intersectoral dependence. Finally, the parameters β_k control the duration of this effect.

This model proved unsatisfactory: it seems to converge toward a constant level of monthly defaults, regardless of the starting point of the predictions. This trend is observed despite the fact that the simulated process history was adjusted to include the actual default history observed before the prediction start date. This behavior appears to be linked to the assumption of a constant basic intensity λ_k , limiting the model's ability to predict below the levels set by these parameters, a problem particularly notable in the post-COVID period.

Nevertheless, the calibration results, both univariate and multivariate, indicate non-zero $\alpha_{k,j}$ parameters, supporting the hypothesis of intra- and inter-sectoral contagion effects. This led us to believe that adjusting the baseline intensity parameters could improve the model's performance.

Final model

We therefore developed a second model, where the basic risk λ becomes a dynamic function $\mu(\cdot)$ dependent on economic indicators.

The framework remains unchanged, but λ_k is replaced by $\mu_k(\cdot)$, directly integrating the economic environment into the evolution of sectoral default risk. Thus, the economic dynamics dictate the general trend of risk, while the excitation component of the process captures accumulation and contagion effects.

The conditional intensity is now defined by

$$\lambda_k^*(t) = \mu_k(t) + \sum_{j=1}^d \sum_{T_n^{(j)} < t} \alpha_{k,j} e^{-\beta_k(t-T_n^{(j)})}, \quad \forall k = 1, \dots, d.$$

The parameters $\alpha_{k,j}$ and β_k retain their roles. The function $\mu_k(\cdot)$, which represents the basic intensity of the process $(N_t^{(k)})_{t \geq 0}$, is constructed by linear regression on various macroeconomic indicators, and its value fluctuates based on the evolution of these indicators.

Results

To evaluate the performance of the final model, we conducted several simulations on past periods, extending the historical data. The empirical average of the simulated trajectories is retained as the prediction.

The accuracy of the forecasts is measured using the *Mean Absolute Percentage Error* (MAPE), a metric suited to significant variations in the number of defaults over time and across different sectors. By

normalizing errors relative to observed values, it allows for performance comparisons across different scales. The table below presents the MAPE calculated for different prediction periods.

Prédiction Period	(A)	(I)	(C)	(E)	(S)	Mean
01/2022 - 06/2024	0.1042	0.1337	0.0780	0.1209	0.0675	0.1009
01/2011 - 12/2023	0.1279	0.1418	0.1368	0.1392	0.1097	0.1297
01/2021 - 12/2022	0.1320	0.2994	0.1723	0.3281	0.2321	0.2522
01/2015 - 06/2024	0.1317	0.1745	0.1386	0.1578	0.1121	0.1429

Although generally satisfactory, the model struggles to capture the sharp decline in defaults observed during the health crisis and the post-COVID period, largely due to government support measures. In January 2021, after a year marked by an exceptionally low level of defaults, the model's conditional intensity remains overestimated. The decay of the excitation component, although present, is insufficient to compensate for the overestimation of the baseline intensity during this period, limiting the model's ability to align with observed reality.

Performance deteriorates significantly between 2020 and 2022, as evidenced by the increase in MAPE for periods including the health crisis. However, over a longer horizon, the model provides consistent results, suggesting that forecast accuracy strongly depends on the estimation of the function $\mu(\cdot)$. Integrating more specific sectoral indicators could improve this estimation and enhance the robustness of long-term forecasts.

The table below compares the performance of the final model to several alternatives: the initial model, a non-homogeneous Poisson process, and a "naïve" model based on the empirical average of the data. The goal is to evaluate the improvements brought by integrating economic factors into the baseline intensity and to measure the contribution of the excitation component. The "naïve" model serves as a benchmark to assess the effectiveness of the more complex models.

Model	(A)	(I)	(C)	(E)	(S)	Mean
Hawkes (2.22)	0.1247	0.1319	0.1417	0.1376	0.0951	0.1262
Hawkes (2.18)	0.1259	0.1816	0.2143	0.1333	0.1441	0.1598
Poisson	0.1259	0.1546	0.1481	0.2107	0.1015	0.1476
Naïve	0.1573	0.2599	0.2556	0.3024	0.1665	0.2283

MAPE for the various models over the period 2011-2024

The results in the table show that the final Hawkes model offers the best accuracy, with a lower MAPE than the other approaches. The integration of economic factors significantly improves performance compared to the initial Hawkes model.

The non-homogeneous Poisson process provides decent results but remains less accurate than the final model. The contribution of the excitation component in our model is therefore real, although it plays a less significant role than expected.

Impact of the economic context on potential losses of a fictional Credit Insurer

The objective is to explain how the work on Hawkes processes can be used to model potential losses. Our modeling focuses solely on the expected losses associated with named contracts. Additionally,

given the publicly available data, we only modeled defaults resulting from the total insolvency of companies, excluding Protracted defaults.

Simplified Model

The expected loss over the period $[0, t]$ is given by

$$L_t = \sum_{k=1}^d \tilde{N}_t^{(k)} \times \bar{E}_k \times \sigma$$

where :

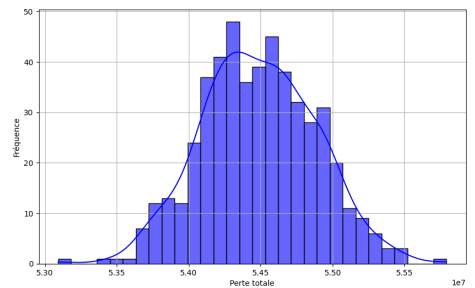
- $\tilde{N}_t^{(k)} = N_t^{(k)} \times p_k$ represents the defaults affecting the insurer's portfolio in sector k , with $N_t^{(k)}$ modeled by a Hawkes process;
- \bar{E}_k is the average guaranteed exposure in sector k ;
- $\sigma = 6.8\%$ corresponds to the average loss rate of claims.

Sectoral exposure was estimated from Coface's annual reports and INSEE databases. The model's consistency was validated by comparing it with the losses reported by Coface between 2021 and 2023. The table below presents the exposure data constructed for the year 2024.

Sector	n_k	E_k	\bar{E}_k	p_k
(A)	71 521	11 564 500 000	161 693.76	0.1842
(I)	10 583	23 533 000 000	2 223 665.74	0.0297
(C)	19 574	7 423 500 000	379 256.67	0.0298
(S)	40 843	24 502 600 000	599 916.16	0.0117
(E)	33 116	2 171 500 000	65 573.46	0.0032
Total	175 637	69 195 100 000	393 967.08	0.0278

Forecasts for end of 2024 and Optimal allocation

The figure on the right presents the distribution of the expected loss for our fictional credit insurer for the year 2024. It should be noted that the only stochastic component of the model is the frequency of defaults, which is modeled by a multivariate Hawkes process. The results presented are based on 500 simulations of the final model. According to the model's forecasts, the average expected loss for 2024 amounts to 54.5 M€, with a 95% confidence interval ranging from 53.7 M€ to 55.3 M€. This estimate represents a 17.2% increase compared to 2023. This gap results exclusively from the evolution of defaults, all else being equal.



Furthermore, if we consider the amount of expected loss in a scenario where the level of defaults would have remained close to that observed before the COVID-19 pandemic, the average loss would have been 49,840,180 €, corresponding to a relative gap of 9.3%.

Finally, we analyzed the sectoral distribution of our fictional insurer's exposure and explored optimization strategies to minimize expected losses in 2024, while maintaining a constant premium

volume. The goal was to determine the optimal allocation of this exposure across different sectors in the current context, with the primary objective of minimizing expected losses for 2024.

The optimization problem is as follows:

$$\begin{cases} \min_{\boldsymbol{\theta} \in \mathbb{R}^d} \sum_{k=1}^d x_k \theta_k, \\ s.c : \sum_{k=1}^d \theta_k = 1, \quad |\theta_k - \theta_k^0| \leq s \text{ et } \theta_k \geq 0, \forall k, \end{cases}$$

where θ_k represents the optimized exposure share for sector k , θ_k^0 the initial allocation, and s defines the maximum tolerance allowed for the variation between the initial and optimized allocation.

We studied two scenarios for 2024: a central scenario based on the average of the simulations and a pessimistic scenario corresponding to the 99.5% quantile. In both cases, we observed that the trend of defaults is upward in all sectors. The optimization therefore favors the agricultural and non-market services sectors, where risks are lower, and reduces exposure in the industry and construction sectors, where the ratio between exposure and the number of defaults is less favorable for 2024.

Thus, although a redistribution of exposure could, theoretically, constitute an interesting strategy for managing the risk of a credit insurer, our simplified modeling leads to a maximum allocation toward sectors with the most favorable ratio between average exposure and the number of defaults, and depends on the insurer's actual ability to adjust its presence in certain markets, taking into account diversification constraints and the realities of the insurance market. Examining Coface's sectoral allocation over the past three years, we observe that it varied by less than 1% in each sector annually, highlighting potential inertia in exposure adjustments.

Conclusion

This paper explored the impact of the economic environment on the activity of a fictional credit insurer through a dynamic modeling of defaults based on Hawkes processes. These processes proved relevant for capturing interactions between sectors, although a finer segmentation (exposure, rating, location) could improve their accuracy.

The developed model integrates the risk environment specific to each sector through basic risk functions and the dependence between defaults via excitation parameters, offering a dynamic prediction that accounts for economic developments without the need for readjustments. However, this approach favors short-term modeling, although coupling it with an economic scenario generator could enable long-term forecasts. The model showed limitations in modeling defaults during the Covid crisis, mainly due to the difficulty of integrating government support measures. This overestimation, however, allowed for an estimation of the expected catch-up effect.

Hawkes processes nevertheless have limitations: increasing complexity with the number of variables and high sensitivity to data, impacting parameter estimation and predictions. Despite these challenges, the loss model produced results consistent with actual losses over three years. Finally, the optimization of the insurer's sectoral allocation showed interesting results but limited flexibility due to market constraints, reducing the insurer's ability to adjust.

Remerciements

Je tiens tout d'abord à remercier mes encadrants en entreprise, Romain SPEISSER et Godfried GUENDEHOU, pour leur accompagnement précieux et leurs conseils avisés, qui ont été essentiels à la réussite de ce mémoire.

Je remercie tout particulièrement les associés du pôle Actuariat de Forvis Mazars, Alice THOU, Alexandre GUCHET et Grégory BOUTIER, pour m'avoir offert l'opportunité de réaliser ce mémoire au sein du cabinet, ainsi que toute l'équipe Actuariat pour son accueil chaleureux et sa bienveillance tout au long de mon stage.

Un grand merci à mes collègues stagiaires, Agathe WITTOECK et Liliana DOBRESKO, dont le soutien et la convivialité ont grandement contribué à faire de cette expérience un moment enrichissant et agréable.

Mes remerciements vont également à l'ensemble de l'équipe pédagogique du Master Actuariat de l'université Paris-Dauphine et plus particulièrement à Quentin GUIBERT, responsable du master, pour la qualité de l'enseignement dispensé et son engagement envers la formation, ainsi qu'à Christophe DUTANG, mon tuteur académique, pour son implication tout au long de mon stage et ses orientations précieuses qui m'ont été d'une grande aide pour la réalisation de cette étude.

Enfin, je tiens à adresser ma profonde gratitude à mes proches, et tout spécialement à Bérénice CHAUVEAU, dont les relectures minutieuses ont été d'un soutien inestimable pour finaliser ce travail.

Table des matières

Résumé	3
Abstract	4
Note de Synthèse	5
Synthesis note	13
Remerciements	19
Table des matières	21
Introduction	23
1 Cadre de l'étude	25
1.1 Présentation de l'assurance-crédit	25
1.2 Sources de risque pour l'assureur	30
1.3 Modélisation des pertes en assurance-crédit	34
2 Modélisation des défauts par les processus de Hawkes	45
2.1 Processus de Hawkes	45
2.2 Modélisation du nombre de défauts	58
2.3 Évaluation des performances du modèle	79
3 Impact du contexte économique	91
3.1 Modélisation des pertes attendues	91
3.2 Mesure de l'intensité de l'effet de rattrapage de la crise Covid	99
3.3 Analyse des pertes selon les expositions sectorielles	101
Conclusion	107
Bibliographie	109

A	Annexes	111
A.1	Exemple de grille de notation	111
A.2	Algorithme de <i>thinning</i> de Lewis	112
A.3	Méthode BFGS	113
A.4	Résultats calibrage	114

Introduction

Avant la crise du Covid-19, la France enregistrait en moyenne 59 342 défaillances d'entreprises par an (chiffre mesuré sur la période 2010-2019). De manière contre-intuitive, le nombre de défaillances a fortement reculé pendant la pandémie de Covid-19, en raison des mesures de soutien gouvernementales, et est resté très bas jusqu'à début 2022. A fin juin 2024, le nombre de défaillances cumulé sur les douze derniers mois atteint 61 091, un niveau légèrement supérieur à celui de la période précédant la crise Covid. Cependant, le contexte économique à cet instant est sensiblement différent de celui d'avant crise, marqué par une inflation toujours au-delà du seuil symbolique des 2%, des taux d'intérêts qui, après avoir été très bas pendant plusieurs années, sont à des niveaux inédits depuis plus de 10 ans, et des conflits géopolitiques, mettant ainsi à l'épreuve les assureurs-crédit.

Les modèles classiquement utilisés en assurance-crédit pour prédire la survenance des défauts reposent principalement sur des scores affectés aux entreprises, censés refléter la solidité financière de celles-ci. Ces modèles font intervenir des méthodes statistiques qui manipulent un grand nombre de variables, rendant parfois les résultats difficiles à interpréter. Ils estiment la dépendance entre défauts par une structure de corrélation entre différents facteurs de risques auxquels les entreprises sont exposées, représentant l'environnement économique associé à leur secteur d'activités et zone géographique. Cependant, ces modèles rencontrent des difficultés pour conditionner la fréquence des défauts à l'évolution de l'environnement économique. Des événements passés, comme la piraterie dans le canal de Suez fin 2023, responsable de retards de livraisons des fournitures et de coûts plus élevés entraînant des défaillances au niveau des entreprises opérant à l'international, soulignent l'importance d'une modélisation plus dynamique dans le cadre de la gestion des risques d'un assureur-crédit.

Dans cette perspective, les processus de Hawkes offrent une approche intéressante. Ces processus permettent de modéliser des effets d'excitation et de dépendance sans avoir à définir explicitement la structure de corrélation entre variables. Ce type de processus a déjà été utilisé dans un cadre de risque de crédit, comme le montrent les travaux d'ERRAIS et al. (2010). Récemment, des recherches menées dans un cadre assurantiel, notamment BESSY-ROLAND et al. (2021) et HILLAIRET et al. (2022), se sont concentrées sur le risque cyber. Cependant, à notre connaissance, aucune étude n'a exploré l'application de ces processus à l'assurance-crédit.

Ce mémoire est structuré en trois parties. La première partie introduit l'assurance-crédit, une branche assez singulière de l'assurance non-vie, en détaillant son fonctionnement, ses spécificités, ainsi que les principales sources de risques pour les assureurs. Les modèles actuellement utilisés par les assureurs-crédit seront également détaillés dans cette partie.

La seconde partie, qui constitue le cœur de ce travail, se concentre sur l'étude de la capacité des processus de Hawkes multivariés à modéliser les interdépendances entre défaillances sectorielles, tout en intégrant les fluctuations économiques. Le cadre théorique de ces processus sera d'abord introduit, suivi de la formalisation d'un modèle de fréquence pour l'assurance-crédit utilisant ces processus. Les solutions apportées face aux défis computationnels posés par les processus de Hawkes en grande dimension seront également présentées.

Enfin, un modèle visant à estimer les pertes attendues d'un assureur-crédit fictif en utilisant ces processus sera développé. Les données mobilisées ainsi que leur exploitation dans le cadre de ce modèle seront présentées. L'intensité et la durée de l'effet de rattrapage des défaillances lié à la crise du Covid-19, ainsi que son impact potentiel sur les pertes de cet assureur fictif, seront estimés. Une analyse de sensibilité sera ensuite réalisée sur la répartition de l'exposition de l'assureur entre différents secteurs. L'objectif sera d'examiner dans quelle mesure, à exposition constante, une répartition sectorielle optimisée selon les prévisions économiques peut réduire les pertes.

Chapitre 1

Cadre de l'étude

Dans ce premier chapitre, nous présentons les principaux éléments de contexte relatifs à notre étude. Nous commencerons par une description générale de l'assurance-crédit et des sources de risques pour l'assureur, notamment dans le contexte économique actuel. Ensuite, nous détaillerons la modélisation des pertes attendues en assurance-crédit, actuellement adoptée par les assureurs, ainsi que ses limites face aux défis actuels. Enfin, nous introduirons la problématique abordée dans ce mémoire et l'approche mise en oeuvre pour tenter d'y répondre.

1.1 Présentation de l'assurance-crédit

L'assurance-crédit est une branche de l'assurance non-vie opérant sur le marché des entreprises qui a pour rôle de couvrir le non-paiement des factures dans le cadre des relations commerciales entre entreprises. Le fonctionnement de cette branche d'assurance est particulier et diffère des produits d'assurance classiques. Il convient de commencer par une présentation générale de l'assurance-crédit et de ses spécificités.

1.1.1 Rôle de l'assurance crédit

Lorsqu'une entreprise fournit une prestation à une autre, qu'il s'agisse de biens ou de services, elle lui accorde généralement un délai de paiement, et donc lui consent un crédit. En France, ces crédits inter-entreprises (CIE) sont estimés à 810[†] milliards d'euros en 2022 et représentent la première source de financement à court terme des entreprises.

Cependant, accorder ces délais expose les entreprises au risque de ne jamais être payées pour les prestations livrées. Ceci dit, une entreprise qui choisirait de ne pas accorder ces délais de paiements à ses clients serait moins compétitive sur le marché face à des concurrents acceptant de les délivrer. Par ailleurs, en facilitant la gestion de trésorerie et en permettant aux entreprises de financer leur cycle d'exploitation sans recourir à des financements externes, les CIE contribuent de manière significative au développement des entreprises et jouent un rôle clé dans la croissance économique.

Néanmoins, les impayés représentent un facteur majeur de défaillances d'entreprises aujourd'hui. En France, 80% des entreprises font face à des défauts de paiement et 25% des dépôts de bilan sont attribuables à ces mêmes impayés[‡]. Les retards de paiement fréquents ou importants, ainsi que les

[†]Observatoire des délais de paiement - Rapport annuel 2023, Banque de France.

[‡]<https://www.coface.fr/a-propos/coface-en-france>

créances irrécouvrables, fragilisent en effet la trésorerie d'une entreprise et peuvent compromettre sa viabilité. Cet impact est particulièrement prononcé pour les entreprises à faible marge bénéficiaire, disposant de moins de marge de manœuvre pour absorber les perturbations financières. Ces retards peuvent aussi provoquer des réactions en chaîne sur d'autres entreprises, et potentiellement accentuer les problèmes de liquidité à travers différents secteurs de l'économie.

C'est là qu'intervient l'assurance-crédit. L'assureur accepte de prendre en charge le risque d'impayés de l'entreprise assurée et de l'indemniser en cas de défaut de son client, en contrepartie d'une prime d'assurance. De cette manière, l'assurance-crédit offre aux entreprises une garantie pour ce mode de financement et sécurise donc les transactions commerciales inter-entreprises.

Les risques généralement couverts par l'assurance-crédit sont :

- le risque commercial, qui inclut les défauts de type *Protracted*, c'est-à-dire les retards de paiement dus à des difficultés ponctuelles de trésorerie ou précédant une faillite, ainsi que les défauts de type *Insolvabilité*, qui correspondent aux créances irrécouvrables résultant de l'insolvabilité du client, déclaré en faillite par les tribunaux compétents ;
- le risque politique, c'est-à-dire le non-paiement causé par des événements extérieurs publics et civils défavorables comme par exemple le risque de déclenchement de guerre ou de restrictions de change. Ce risque concerne principalement les entreprises qui développent leur commerce à l'étranger.

En plus de fournir une protection contre ces risques, l'assurance-crédit renforce la crédibilité financière de l'entreprise assurée et améliore donc son image auprès des banques et autres investisseurs.

Malgré tout, il est légitime de se demander si une entreprise a plus intérêt à souscrire une assurance-crédit ou à disposer d'une équipe de gestion des risques en interne. Contrairement aux entités financières comme les banques et assurances, qui disposent de méthodes structurées pour évaluer le risque de crédit, les entreprises ne possèdent pas nécessairement les outils adéquats pour évaluer précisément ce risque.

Internaliser cette fonction nécessiterait des investissements importants dans une équipe spécialisée et des outils de gestion des risques, entraînant des coûts fixes significatifs. Par ailleurs, les informations sur la solvabilité et la solidité des clients ne sont pas toujours facilement accessibles, ce qui rend l'obtention et la maintenance de ces informations très coûteuses. En revanche, un assureur-crédit a pour activité principale l'évaluation et le pilotage des risques qu'il couvre. Il dispose ainsi des moyens nécessaires pour évaluer la solvabilité des entreprises sans coûts opérationnels supplémentaires. De plus, il peut mutualiser les coûts liés à l'acquisition de nouvelles informations sur les entreprises et à la mise à jour de ses bases de données sur l'ensemble des clients de ses assurés, rendant le suivi des risques plus rentable et efficace.

1.1.2 Spécificités de l'assurance-crédit

La principale particularité de l'assurance-crédit repose sur le fait que le risque ne provient pas de l'entreprise assurée mais de son client. Un contrat d'assurance-crédit peut donc être représenté par une relation en triangle, comme illustré sur la figure 1.1. Le contrat implique deux parties, l'entreprise souscriptrice du contrat et l'assureur, auxquelles s'ajoute une troisième partie : le client de l'assuré, que l'on nommera *acheteur*. Il est important de noter que, même si c'est sur l'acheteur que l'aléa repose, il n'existe pas de relation contractuelle entre l'acheteur et l'assureur.

Lorsqu'une entreprise souscrit un contrat d'assurance-crédit classique pour se prémunir contre le risque d'impayés de ses clients, elle doit effectuer une demande d'*agrément* pour chaque client qu'elle souhaite couvrir. Un agrément définit une limite de couverture spécifique à un assuré et un acheteur

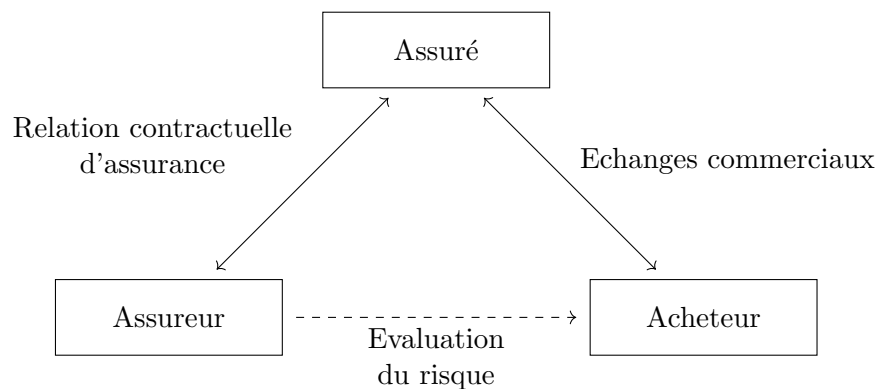


FIGURE 1.1 : Fonctionnement de l'assurance-crédit

donné. Si une entreprise souhaite assurer ses transactions avec plusieurs acheteurs, elle doit alors soumettre une demande d'agrément pour chacun d'entre eux.

Ensuite, pour chaque demande, la compagnie détermine une limite de décaissement correspondant au montant maximal qu'elle est prête à garantir pour cet acheteur. Cette limite est définie en fonction d'une note attribuée par l'assureur, représentant la solidité financière de l'acheteur.

En effet, l'assureur-crédit possède un service d'analystes crédits qui évaluent continuellement la santé financière des acheteurs. A l'image des grandes agences de notation (S&P, Moody's, Fitch, etc.), ils établissent une grille de notation sur les entreprises, basée sur une analyse des derniers bilans et comptes de résultats de celles-ci. Ces évaluations sont supposées refléter le risque de crédit lié à chaque entreprise. Un exemple de grille de notation est présenté en annexe A.1.

La gestion des agréments est assurée par le service d'*arbitrage*. C'est ce service qui délivre les agréments et détermine les limites de couverture pour chaque acheteur. En outre, les arbitres ajustent ces agréments en continu, tenant compte de l'évolution de la situation financière des acheteurs. En effet, une des caractéristiques essentielles de l'activité d'assurance-crédit est la capacité qu'a l'assureur de réduire ou d'annuler les garanties souscrites à tout moment, en réponse aux évolutions du risque de crédit de l'acheteur. En cas d'annulation ou réduction de la limite de couverture sur un acheteur, l'assureur notifie son assuré de l'évolution de son risque et l'agrément passé ne reste valable pour les produits et services déjà facturés ou en attente de règlement. Grâce à une surveillance régulière et à des analyses approfondies des bilans et comptes de résultats, le service d'arbitrage suit avec attention les entreprises susceptibles de se fragiliser et d'augmenter leur probabilité de défaut, et adapte en conséquence les limites de décaissement.

On peut noter qu'avec ce type de contrat, l'assureur-crédit bénéficie d'une borne maximale pour évaluer ses pertes potentielles, puisqu'il définit une limite d'engagement sur chaque acheteur.

Outre la couverture accordée par les agréments délivrés, le contrat d'assurance-crédit peut inclure une limite de crédit discrétionnaire. Cela permet à l'assuré d'être couvert pour des sinistres mineurs sans avoir besoin de demander un agrément préalable. Toutefois, bien que ces sinistres soient « petits », leur volume peut être significatif et doit être pris en compte par l'assureur. Pour ces acheteurs, appelés *non dénommés*, on parle de couverture à l'aveugle, car leur situation, étant inconnue, n'est pas suivie de manière spécifique par l'assureur.

En complément de l'indemnisation des impayés et de la prévention des risques d'impayés via le suivi continu de la solvabilité financière des clients de ses assurés, l'assurance-crédit intègre un service complémentaire essentiel : le *recouvrement* des factures. L'assureur s'occupe des démarches nécessaires

à la récupération des créances auprès des acheteurs en défaut et gère les éventuels contentieux.

Le processus d'indemnisation d'un sinistre en assurance-crédit se déroule en plusieurs étapes, illustrées sur la figure 1.2 :

- Émission d'une facture suite à la livraison d'une prestations ;
- Difficultés de paiement de l'acheteur et dépassement du délai contractuel ;
- Demande d'intervention contentieuse (DIC) : l'assuré déclare à l'assureur le dépassement du délai de paiement maximum contractuel par l'acheteur ;
- Mise en oeuvre de la procédure de recouvrement, d'abord à l'amiable, car dans la majorité des cas, l'impayé est dû à des problèmes ponctuels de trésorerie (environ 70% des impayés sont résolus à l'amiable) ;
- Si la procédure amiable réussit, l'assuré reçoit le montant récupéré sans attendre de délai de versement. En cas d'échec de la procédure, engagement des procédures judiciaires (litiges, procédures collectives, etc.) ;
- Passé un délai de carence, indemnisation des sinistres selon les termes du contrat. Les montants éventuellement récupérés post-indemnisation sont alloués au remboursement de la perte subie par l'assureur, incluant les indemnités versées et les frais éventuels de recouvrement, puis sont reversés à l'assuré, dans la limite de la perte subie.

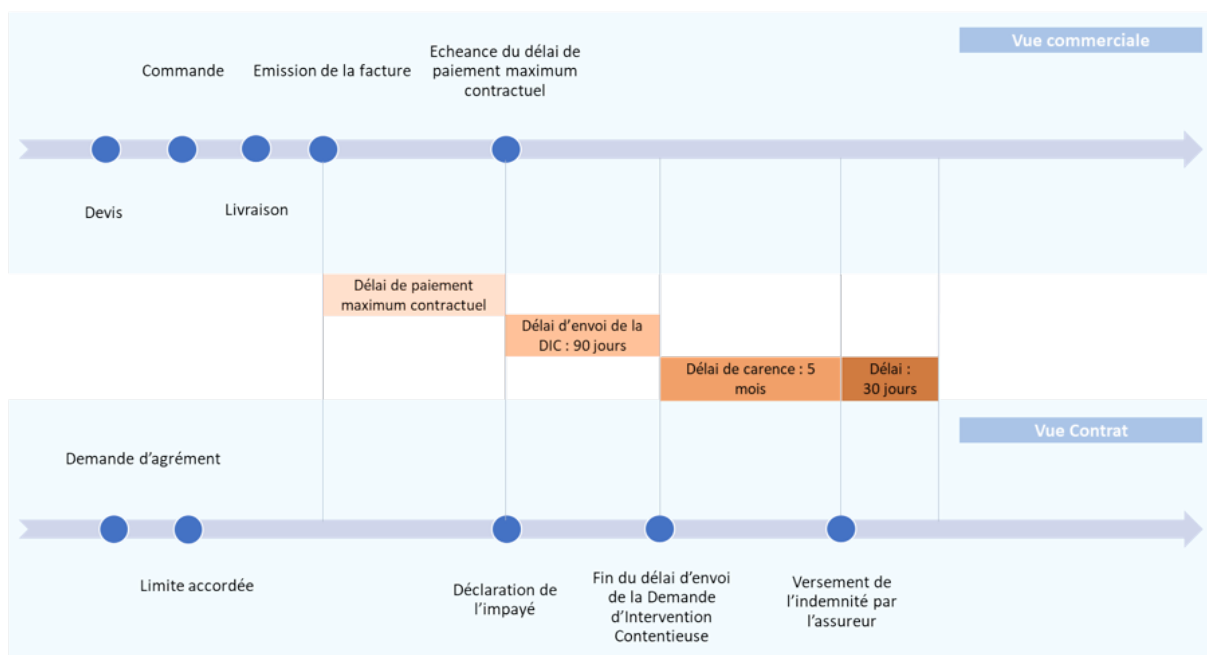


FIGURE 1.2 : Procédure d'indemnisation des sinistres en assurance-crédit

En résumé, l'assurance-crédit est une branche singulière de l'assurance non-vie, se positionnant comme un outil de gestion des risques répondant aux enjeux financiers des entreprises. Elle intègre trois services complémentaires : la prévention et la surveillance des risques clients, le recouvrement des créances impayées, et l'indemnisation des pertes garanties.

1.1.3 Marché de l'assurance-crédit

Le marché de l'assurance-crédit est un marché très concentré, avec trois grands acteurs dominant le secteur : Allianz Trade (Euler Hermès), Atradius et Coface, qui se partagent environ 70% des primes totales (voir Figure 1.3). La force de ces trois assureurs provient de la richesse de leurs bases de données, de leur implantation internationale leur permettant d'accompagner des entreprises sur des marchés étrangers qu'elles maîtrisent moins, et de leur offre globale dans la gestion du risque client.

En complément de ces acteurs majeurs, certains assureurs spécialisés ou « de niche » offrent des produits d'assurance spécifiques adaptés à des besoins particuliers. Par exemple, Axa Assurcrédit se concentre sur les PME et TPE, Groupama AC & Caution se spécialise dans le secteur agroalimentaire, tandis que d'autres assureurs ciblent des zones géographiques spécifiques. Ces acteurs de niche représentent environ 30% du marché global.

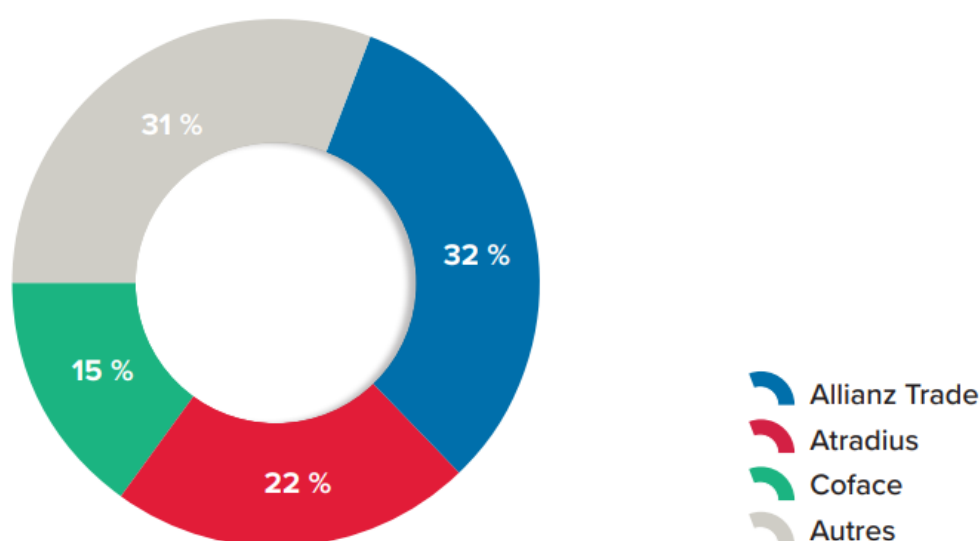


FIGURE 1.3 : Répartition du marché de l'assurance-crédit en 2023 - AU GROUP (2024)

L'assurance-crédit, bien qu'elle ne soit pas obligatoire, parvient à couvrir près de la moitié des CIE en France*. Ce taux de couverture est particulièrement élevé dans les secteurs caractérisés par une clientèle diversifiée et risquée, ainsi que dans les secteurs fortement exportateurs. Par exemple, le secteur de la construction, constitué de structures fragiles et présentant un taux de défaillance parmi les plus élevés, est l'un des plus demandeurs d'assurance-crédit.

C'est une branche courte, les sinistres sont généralement déclarés et indemnisés dans les 6 mois suivant leur survenance. Le ratio combiné des grands acteurs majeurs de l'assurance-crédit s'établit autour de 80%. Cependant, on peut observer sur la figure 1.4 que ce ratio a été très sensible à la crise des subprimes, et dans une moindre mesure à la pandémie de Covid-19.

La détérioration du ratio combiné en 2008 s'explique par la hausse soudaine du nombre de défaillances liée à la crise financière que les assureurs-crédit n'ont pas pu anticiper. En ce qui concerne la crise Covid, le pic observé peut surprendre étant donné que le nombre de défaillances a atteint un niveau extrêmement bas en 2020, mais s'explique par le partage des primes des assureurs avec les États dans le cadre des dispositifs de réassurance publique mis en place pour soutenir les entreprises.

*Mission d'information sur l'assurance-crédit, Assemblée nationale.

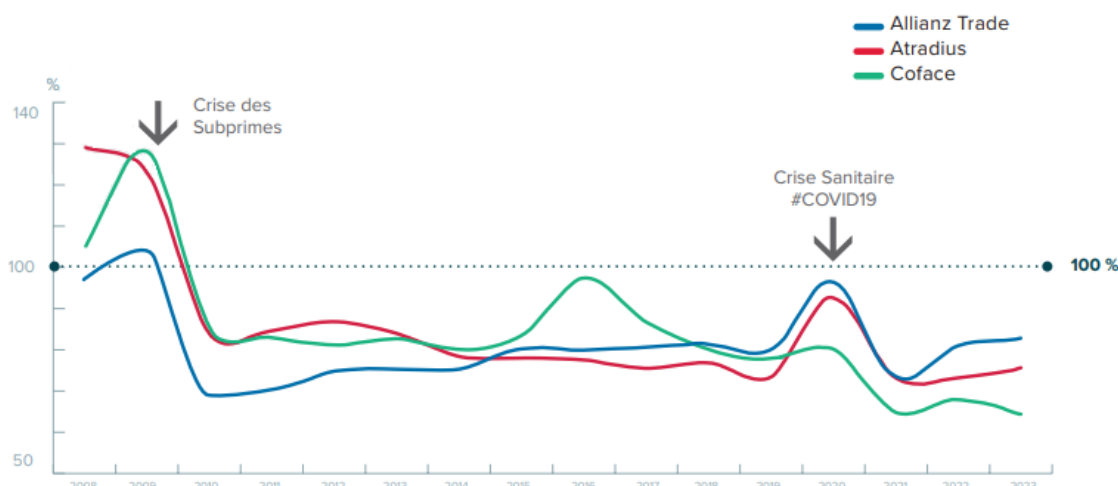


FIGURE 1.4 : Évolution du ratio combiné des trois acteurs majeurs -AU GROUP (2024)

En 2016, la dégradation du ratio combiné de l'assureur Coface résulte d'une sinistralité plus élevée qu'anticipée par l'assureur dans les pays émergents. De plus, le transfert de l'activité de réassurance publique à l'export (CAP, CAP+) de Coface vers la BPI a également pu contribuer à cette dégradation.

1.2 Sources de risque pour l'assureur

Un assureur, quel que soit son domaine d'activité, accepte par définition une certaine exposition au risque en prenant en charge les aléas auxquels ses assurés sont confrontés. Cela implique qu'il doit anticiper, évaluer et gérer les risques associés à ses engagements pour garantir la stabilité et la rentabilité de son activité. Dans le domaine spécifique de l'assurance-crédit, cette évaluation des risques est encore plus essentielle puisque l'assureur dispose d'une plus grande flexibilité pour gérer son exposition au risque par rapport à un assureur classique.

Les risques auxquels un assureur-crédit fait face sont étroitement liés aux fluctuations du contexte économique et politique. D'autre part, les défaillances des clients assurés peuvent présenter une certaine dépendance entre elles, que l'assureur doit anticiper pour piloter au mieux son risque. Cette section se propose d'examiner en détail les principales sources de risque pour un assureur-crédit.

1.2.1 Risque de souscription

Le risque de souscription est une source de préoccupation majeure dans toutes les branches d'assurance non-vie, mais il représente un véritable enjeu en assurance-crédit. Les trois grands assureurs-crédit ont d'ailleurs mis en place un Modèle Interne Partiel pour mieux évaluer ce risque. Il survient lorsque l'assureur évalue incorrectement les primes en fonction de la probabilité de sinistre et des pertes potentielles. En assurance-crédit, une particularité est que la tarification est partiellement assurée par des analystes crédits, elle dépend de l'accord de l'agrément ainsi que de la limite de couverture fixée en fonction de la note attribuée à l'acheteur.

De plus, la prime d'assurance évolue en fonction du chiffre d'affaires garanti, le risque étant supposé croître de manière linéaire avec ce chiffre d'affaires. Cela introduit une incertitude quant à son montant

final. On peut aussi noter que le risque de souscription est influencé par la concurrence entre assureurs, obligeant ces derniers à proposer des primes attractives.

Au sein de ce risque de souscription, on identifie un risque opérationnel significatif propre à l'assurance-crédit, lié à la gestion des limites de couverture, mécanisme fondamental pour limiter les conséquences d'une dégradation du contexte économique. Une mauvaise gestion de ces limites ou un manque d'anticipation peut entraîner des dommages financiers substantiels, comme l'ont démontré les pertes des trois acteurs majeurs lors de la crise de 2008.

1.2.2 Contexte économique et politique

La performance des entreprises est, par définition, étroitement liée à l'environnement dans lequel elles évoluent. En constante mutation, les conditions économiques et les décisions politiques influent directement sur leurs opérations, leur rentabilité et donc leur pérennité. La stabilité économique et politique fournit généralement un cadre favorable à la croissance et à l'innovation des entreprises. À l'inverse, les incertitudes et les crises diminuent la confiance des investisseurs, freinent les initiatives commerciales et peuvent ainsi fragiliser la trésorerie des entreprises. En tant que garant contre les défauts de paiement, l'assureur-crédit se doit d'avoir une compréhension approfondie de l'environnement dans lequel opèrent ses assurés. Cette section se propose de présenter certains facteurs conjoncturels pouvant avoir un impact sur l'activité des entreprises et, par extension, sur le niveau de risque auquel l'assureur crédit est exposé.

La crise du Covid-19, caractérisée par son ampleur et sa rapidité de diffusion sans précédent, a profondément perturbé l'économie mondiale. Les restrictions mises en place par les pouvoirs publics pour endiguer la pandémie, telles que le confinement, l'interruption de la production ou la fermeture des frontières, ont entraîné des perturbations majeures dans les chaînes d'approvisionnement des entreprises, et même une cessation totale des activités dans certains secteurs tels que la restauration et le tourisme. Le secteur des services non marchands, habituellement immunisé face aux crises, a subi pour la première fois un choc sur son activité. Malgré des rentrées quasi absentes durant la crise, les entreprises ont dû continuer à assumer leurs charges fixes telles que les loyers, salaires ou remboursements de leurs emprunts, ce qui a mis une pression financière supplémentaire sur les entreprises en cette période de crise.

Dans un contexte aussi tendu, on aurait pu s'attendre à une augmentation significative du nombre de défaillances d'entreprises. Cependant, grâce aux mesures de soutien gouvernementales, mises en place dans le cadre du fameux *quoi qu'il en coûte*, telles que le chômage partiel, le prêt garantis par l'État (PGE) et le fonds de solidarité pour les petites entreprises, ce scénario catastrophe a été évité. On a même observé une chute du nombre de défaillances d'entreprise jusqu'en 2022, comme le montre le graphique sur la figure 1.5.

Avant la crise sanitaire, la France enregistrait en moyenne 59 342 défaillances d'entreprises par an, chiffre mesuré sur la période 2010-2019. Durant et après la pandémie, le nombre de défaillances est resté très bas. Un effet de rattrapage du nombre de défaillances était attendu à la suite de l'arrêt progressif des aides de soutien de l'État. En effet, si l'on se réfère à la période 2010-2019, il manque environ 50 000 défaillances sur la période 2020-2023. À fin juin 2024, l'effet de rattrapage se confirme. Le nombre de défaillances cumulées sur les douze derniers mois atteint 61 081, un niveau légèrement supérieur à celui de la période d'avant crise (+2.9%) et globalement stable par rapport au mois précédent. Cependant, si l'on regarde les données à la maille trimestrielle, le T2-2024 est le trimestre observant le plus de défaillances d'entreprises depuis 1990. Cela peut s'expliquer, d'une part par l'effet de rattrapage mais également par une conjoncture économique globalement défavorable pour la croissance des entreprises depuis la crise sanitaire.

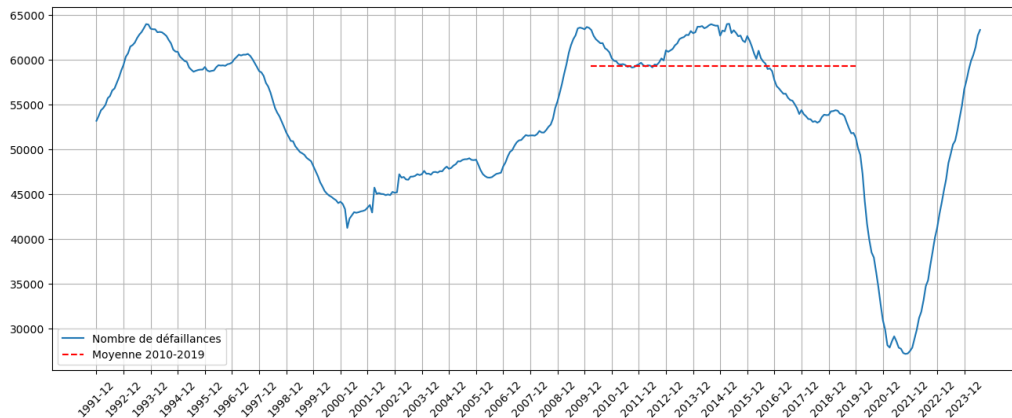


FIGURE 1.5 : Évolution du nombre de défaillances en France, cumul sur 12 mois glissants - BANQUE DE FRANCE (2024)

En effet, à la suite de la crise du Covid, l'économie a été sévèrement affectée par une crise énergétique, initiée par la forte reprise post-pandémie et amplifiée par la guerre en Ukraine, qui a largement contribué au choc inflationniste subi par l'économie à partir de 2021. L'augmentation des prix de l'énergie a entraîné une hausse des coûts opérationnels dans de nombreux secteurs d'activité, en particulier ceux qui dépendent fortement de l'énergie pour leurs opérations, comme les industries manufacturières ou le transport. Cette hausse a été particulièrement difficile pour les entreprises ayant moins de capacité à absorber ces frais supplémentaires, elles ont vu leurs marges se réduire, surtout si elles étaient incapables de répercuter ces coûts sur leurs clients par le biais d'une augmentation des prix.

Aujourd'hui, l'activité économique est contrainte par une inflation, bien qu'en baisse depuis un an, toujours au-delà du seuil symbolique de 2%. Ce niveau est considéré par la BCE et la FED comme un équilibre entre le maintien de la stabilité des prix et la création de conditions propices à la croissance économique. Une inflation élevée peut avoir des impacts majeurs sur le fonctionnement et la croissance des entreprises. Elle peut notamment entraîner une baisse de la demande en raison de la diminution du pouvoir d'achat des ménages, une incertitude quant aux coûts d'approvisionnement, ou encore une réduction des marges due à l'augmentation du prix des matières premières, de l'énergie ou des frais généraux.

Pour maîtriser l'inflation, la BCE a mis en place une politique monétaire restrictive à partir de juillet 2022, augmentant progressivement ses taux directeurs à un niveau inédit depuis 2008. C'est une situation sans précédent pour les entreprises qui ont bénéficié d'un environnement de taux bas pendant près d'une décennie. Cette hausse des taux d'intérêt a plusieurs conséquences sur les entreprises : elle augmente le coût de l'emprunt, rend les investissements (à l'exception des obligations) moins attractifs, diminue la demande de produits et services, et augmente le coût du financement des dettes existantes pour les entreprises ayant des emprunts à taux variables. Ces facteurs, combinés, contribuent à rendre le financement de leurs opérations plus difficile et peuvent entraîner une réduction des revenus et des bénéfices des entreprises, diminuant donc leurs marges.

Bien que la BCE ait engagé au mois de juin 2024 une baisse de ses taux, événement historique puisqu'elle précède pour la première fois la FED, l'effet sur les taux d'emprunt est encore imperceptible.

De plus, les tensions géopolitiques actuelles, comme la guerre en Ukraine et les conflits au Moyen-Orient, ajoutent une couche supplémentaire d'incertitude pour les entreprises. Ces tensions ont un impact direct sur les entreprises qui entretiennent des relations commerciales avec ces régions, rendant leurs opérations d'importation et d'exportation plus complexes et risquées. Cependant, ces conflits

peuvent avoir des répercussions au-delà de ces pays, comme par exemple la guerre entre l'Ukraine et la Russie qui a contribué à l'augmentation des prix des matières premières et de l'énergie, alimentant ainsi le contexte inflationniste actuel, qui a fragilisé l'ensemble de l'économie. A ces conflits s'ajoutent des incertitudes politiques comme la dissolution de l'Assemblée nationale en France, ou encore les élections présidentielles aux Etats-Unis.

En plus de ces défis, près de 680 000* entreprises en France sont actuellement confrontées à la nécessité de rembourser leur Prêt Garanti par l'État (PGE), malgré cette conjoncture économique défavorable. Cela ajoute une pression financière supplémentaire sur ces entreprises, et peut également expliquer en partie l'augmentation du nombre de défaillances au second trimestre 2024.

Dans ce contexte, l'évolution du nombre de défaillances d'entreprises dans les prochains mois reste incertaine. Une analyse des 12 derniers mois révèle des disparités assez importantes entre les secteurs d'activité, comme le montre le tableau 1.1. A l'exception du secteur de l'agriculture qui a montré une certaine résilience face aux défis de la conjoncture économique, on constate une augmentation significative du nombre de défaillances sur tous les secteurs.

Secteur d'activité	Moy. 2010-2019	mai-24	Mai 24/Mai 23	Mai 24/2010-2019
Agriculture, sylviculture et pêche (AZ)	1 359	1 349	9,90%	-0,70%
Industrie (BE)	4 442	4 081	15,90%	-8,10%
Construction (FZ)	14 684	13 291	35,30%	-9,50%
Commerce ; réparation automobile (G)	13070	12 956	21,30%	-0,90%
Transports et entreposage (H)	1 901	2 605	33,90%	37,00%
Hébergement et restauration (I)	7 374	8 161	23,60%	10,70%
Information et communication (JZ)	1 480	1 793	23,20%	21,10%
Activités financières et d'assurance (KZ)	1 150	1 537	39,50%	33,70%
Activités immobilières (LZ)	1 984	2 347	47,00%	18,30%
Soutien aux entreprises (MN)	6 380	7 140	29,70%	11,90%
Enseignement, santé, action sociale et service aux ménages (P à S)	5 311	5 695	22,80%	7,20%
Ensemble	59 342	61 036	26,70%	2,90%

TABLE 1.1 : Défaillances d'entreprise en France par secteur d'activité (cumul 12 derniers mois) - BANQUE DE FRANCE (2024)

Les secteurs du transport et de l'entreposage, des activités financières et d'assurance, ainsi que des activités immobilières ont enregistré une hausse considérable des défaillances par rapport à l'année précédente et la moyenne sur la période pré-Covid. En revanche, l'industrie, la construction et le commerce, bien qu'en hausse par rapport à l'année dernière, sont toujours à un niveau plus bas que la moyenne de la période pré-Covid. Les secteurs de l'hébergement et de la restauration, de l'information et de la communication, ainsi que les services aux entreprises connaissent une croissance modérée des défaillances.

Ces disparités s'expliquent en partie par la conjoncture économique mais également par les comportements de paiements des clients qui diffèrent d'un secteur à l'autre. Par exemple, le secteur de la restauration bénéficie de paiements comptants et subit donc moins de retards de paiements que le secteur de la construction où plus d'une entreprise sur deux est payée après le délai légal.

*Rembourser les prêts garantis par l'Etat, une priorité pour les entreprises, Le Monde (février 2024).

Dans cet environnement, l'exposition au risque d'un assureur-crédit dépend fortement de la composition de son portefeuille d'assurés. Les assureurs-crédit doivent comprendre en profondeur comment les différents secteurs réagissent aux variations économiques, en particulier en situation de crise. Par exemple, les entreprises du secteur de la construction, qui sont fortement dépendantes des prix des matières premières, pourraient être plus vulnérables en période d'inflation élevée. De même, les entreprises opérant dans des secteurs sensibles aux variations de la demande des consommateurs, comme le commerce de détail, pourraient être plus touchées par les variations du pouvoir d'achat des ménages. D'autre part, un assureur qui assure une grande part d'entreprises opérant à l'étranger sera plus sensible aux fluctuations économiques mondiales et aux instabilités politiques. Toutefois, on peut noter que même si un contexte difficile peut inciter les entreprises à demander davantage d'accords de crédit, le volume des transactions aura normalement tendance à diminuer, n'augmentant donc pas forcément l'exposition de l'assureur.

1.2.3 Dépendance entre défauts

Un assureur-crédit, à l'image d'un gestionnaire d'actifs, doit prêter une attention particulière à la diversification de son portefeuille d'assurés. De nombreuses études macroéconomiques, comme DAS et al. (2007), ont démontré que les entreprises ont une propension à faire défaut simultanément. Cette interdépendance entre les défauts peut s'expliquer par le fait que la performance d'une entreprise est intrinsèquement liée à l'environnement économique dans lequel elle opère. Ainsi, les entreprises qui opèrent dans le même pays, le même secteur, ou qui partagent des clients ou des fournisseurs similaires, sont exposées à des facteurs de risques communs. Par exemple, un épisode de sécheresse prolongé en France pourrait entraîner une série de défauts dans le secteur agricole.

Cette corrélation entre les défauts s'explique également par un effet de contagion dans les relations interentreprises. Lorsqu'une entreprise fait défaut, elle peut potentiellement fragiliser toutes les entreprises avec lesquelles elle a des échanges commerciaux et qu'elle ne peut plus rembourser.

Déterminer cette dépendance est une chose complexe mais cruciale pour l'assureur dans le cadre de la gestion des limites de couverture. En anticipant suffisamment tôt les risques de contagion dans son portefeuille, l'assureur a la capacité de se désengager et donc d'éviter de subir les conséquences d'une propagation des risques.

1.3 Modélisation des pertes en assurance-crédit

La modélisation adoptée par les assureurs-crédit ne repose généralement pas sur une approche assurantielle classique, qui considère la charge globale par une approche fréquence-sévérité, mais s'appuie plutôt sur les méthodes traditionnelles utilisées par les banques pour mesurer le risque de crédit. Dans cette section, nous détaillerons les différents concepts utilisés pour modéliser les pertes selon l'approche *risque de crédit*, avant de présenter les principes des modèles actuellement en vigueur chez les assureurs.

1.3.1 Mesure du risque

Pour évaluer efficacement le risque de crédit de son portefeuille, un assureur-crédit doit prendre en compte trois éléments clés :

- la probabilité de défaut ;
- l'exposition en cas de défaut ;
- la perte en cas de défaut.

Probabilité de défaut

La probabilité de défaut PD à la date t correspond à la probabilité qu'au moins un assuré déclare un sinistre sur un acheteur pendant une période donnée, en général dans les 12 mois suivants.

Elle peut être estimée empiriquement comme étant le nombre d'acheteurs ayant eu au moins un défaut rapporté au nombre total d'acheteurs

$$PD_t = \frac{d_t}{n_t},$$

où d_t correspond au nombre d'acheteurs ayant fait défaut durant les 12 mois suivant la date t et n_t est le nombre total d'acheteurs à la date t .

La PD peut être calculée selon différentes segmentations, telles que le pays, le marché (domestique ou export), la note de l'acheteur, le type de défaut (*Protracted* ou *Insolvability*), le secteur d'activité ou encore le niveau d'exposition. Le choix de la segmentation est généralement réalisé par un jugement d'expert.

Au-delà des probabilités de défaut individuelles, la corrélation entre les défauts est un facteur crucial à considérer dans l'évaluation du risque de crédit. En effet, un portefeuille composé d'entreprises présentant un faible risque de défaut individuel, mais étant fortement interdépendantes, expose l'assureur à un risque d'accumulation de sinistres.

Il convient de distinguer deux types de dépendances. D'une part, le risque de contagion, qui concerne les entreprises ayant des relations commerciales entre elles. D'autre part, le risque de dépendance associé à l'environnement économique, qui résulte de l'exposition des entreprises à des facteurs de risque communs.

La corrélation entre les défauts est généralement modélisée à travers la modélisation de la probabilité de défaut. En pratique, cette dernière est souvent modélisée par une approche de type *Merton*, que nous présenterons dans la section suivante.

Exposition en cas de défaut

Sauf exception, l'assureur ne prend pas en charge la totalité des encours de l'assuré. Il ne paie pas au-delà de la limite accordée et peut réduire la perte associée au défaut par des mécanismes de recouvrement. De plus, à l'instar de nombreux contrats d'assurance, des franchises peuvent également s'appliquer. Ainsi, pour déterminer le montant de perte en cas de défaut, c'est-à-dire la perte réelle de l'assureur associé à l'impayé, l'assureur doit d'abord déterminer son exposition en cas de défaut *EAD* (*Exposure At Default*).

Celle-ci correspond à l'encours garanti au moment du défaut et se calcule selon la formule suivante

$$EAD = L^{(0)} \times UGD,$$

où $L^{(0)}$ est la limite initialement accordée par l'assureur et le taux UGD (*Usage Given Default*) représente la part de l'encours total effectivement utilisée au moment du défaut.

Ce ratio permet de prendre en compte deux effets distincts. D'une part, il intègre les éventuelles réductions ou l'annulation de la limite accordée avant le moment de défaut. D'autre part, il tient compte du fait que l'assureur ne connaît les encours réels des assurés sur un acheteur donné qu'en cas de défaut, ces encours étant liés aux factures émises par les assurés à leur client. Par conséquent, ces encours peuvent être considérablement inférieurs à la limite accordée par l'assureur. A l'inverse, il est également possible que les assurés facturent au-delà de la limite.

Cependant, ces deux phénomènes peuvent évoluer ensemble. En effet, lorsque l'assureur réduit ou annule une limite de couverture, l'assuré est informé de la dégradation de la solvabilité de l'acheteur. Généralement, l'assuré diminue alors le volume des transactions avec son client, interprétant cette information comme une indication d'une augmentation de la probabilité de défaut de celui-ci. De plus, la réduction de sa couverture a entraîné une augmentation de son exposition au risque.

En pratique, ce sont les *UGD* qui sont modélisés par les assureurs. Ils peuvent être modélisés de façon stochastique à l'aide de distributions calibrées sur différents segments. La segmentation est effectuée à l'aide d'un modèle de classification se basant sur les caractéristiques des assurés et des acheteurs comme les forêts aléatoires.

Perte en cas de défaut

Une fois le taux d'utilisation de l'agrément déterminé, et par conséquent l'exposition en cas de défaut, on obtient la perte en cas de défaut *LGD* (*Loss Given Default*), en appliquant à l'*EAD* les différentes mesures de réduction de pertes pour l'assureur.

On a

$$LGD = EAD \times (1 - \tau_r),$$

où le taux de réduction de perte τ_r représente la part de l'exposition que l'assureur ne prend pas en charge grâce à divers mécanismes de limitation de ses pertes.

Certaines de ces mesures sont intégrées directement dans les clauses du contrat, telles que :

- les franchises ;
- les quotités garanties, qui sont des franchises proportionnelles applicables au montant de sinistre, l'assureur ne prend en charge qu'un pourcentage de l'indemnisation ;
- la garantie annuelle, qui est une autre forme de franchise qui s'applique à la somme des sinistres d'un assuré (assimilable à une clause *Annual Agregate Deductible* en réassurance) ;
- la limite de décaissement (ou clause de responsabilité maximale), qui définit le montant maximum que l'assureur paiera à l'assuré sur une période fixée.

Ces clauses de partage des pertes entre l'assureur et l'assuré sont généralement mises en place pour limiter l'anti-sélection. En complément, l'assureur peut réduire sa perte en récupérant tout ou partie des impayés via le recouvrement.

La figure 1.6 ci-dessous résume le passage de la limite accordée initialement par l'assureur à la perte effective en cas de défaut.

La courbe verte trace l'évolution de la limite de l'agrément, qui diminue chaque fois que l'assureur dégrade le client de l'assuré. En parallèle, la courbe jaune dépeint l'évolution de l'exposition en cas de défaut, qui fluctue en fonction de l'activité de l'assuré avec son client et est plafonnée au montant de la limite accordée. En cas de défaut, l'*UGD* correspond au rapport entre la garantie utilisée au moment du défaut et la limite initialement accordée. Par la suite, divers mécanismes de réduction de perte permettent de passer de l'encours de l'assuré au moment du défaut à la perte réelle pour l'assureur, cette évolution est représentée par la courbe bleue. Le rapport entre la perte subie et la limite utilisée définit le taux *LGD*.

Remarque 1.1. Le taux d'utilisation de l'agrément *UGD* peut servir, en quelque sorte, d'indicateur du degré d'anticipation de l'assureur par rapport à l'évolution de la solvabilité de l'acheteur. En effet, si l'assureur a suivi attentivement la situation financière de l'entreprise et a anticipé le défaut, ce ratio devrait être faible, voire nul dans le meilleur des cas. Par ailleurs, la *LGD* peut être perçue comme une mesure de l'efficacité de l'assureur dans la gestion de ses sinistres. Une *LGD* faible voire nulle indique que l'assureur a réussi à recouvrer efficacement les factures impayées.

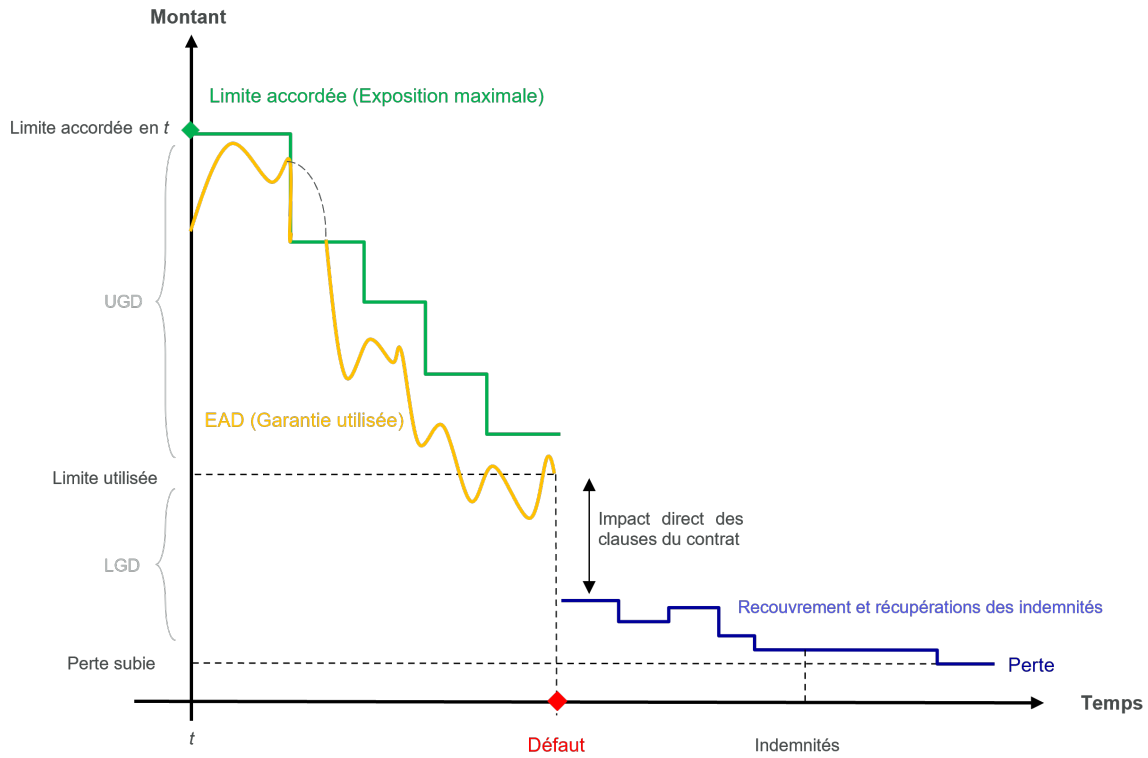


FIGURE 1.6 : Perte en cas de défaut

Remarque 1.2. La LGD est définie ici en montant, mais elle s'exprime souvent en pourcentage et représente la part de l'exposition qui n'est pas récupérée par l'assureur. Dans ce cas, le taux de perte LGD est défini par

$$LGD = (1 - \tau_r).$$

Sauf mention contraire, dans la suite du mémoire, nous considérerons la LGD comme un pourcentage de la perte en cas de défaut. La perte en cas de défaut l s'obtient alors par le calcul suivant

$$l = EAD \times LGD.$$

En pratique, les taux LGD sont modélisés de façon stochastique à l'aide de distributions calibrées sur différents segments. La méthodologie est similaire à celle utilisée pour les UGD .

Après le calcul de l'ensemble de ces paramètres, qui sont au cœur de la modélisation, il sera possible de déterminer les pertes attendues *dénomées*. Il s'agit des pertes associées aux agréments délivrés, où l'entreprise en défaut est connue et suivie par l'assureur, en opposition aux pertes non dénomées. Pour rappel, le non dénomé correspond à une couverture à l'aveugle de l'assureur sur les petits sinistres de l'assuré.

Nous n'aborderons pas le calcul des pertes attendues non dénomées dans ce mémoire.

Notons :

- $L^{(0)}$ la limite de couverture initialement accordée sur l'agrément i ;
- UGD_i le taux d'utilisation de l'agrément i ;
- LGD_i le taux de perte en cas de défaut de l'acheteur associé à l'agrément i ;

- PD_i la probabilité de défaut de l'entreprise couverte par l'agrément i .

La distribution des pertes d'un portefeuille composé de n agréments est définie par la variable aléatoire

$$L = \sum_{i=1}^n L_i^{(0)} \times UGD_i \times LGD_i \times \mathbf{1}_{D_i} \quad \text{où } D_i \sim \mathcal{B}(PD_i),$$

et la perte attendue dénommée EL (*Expected Loss*) s'obtient par la formule suivante

$$EL = \mathbb{E}[L] = \sum_{i=1}^n L_i^{(0)} \times UGD_i \times LGD_i \times PD_i, \quad (1.1)$$

Nous avons présenté les différentes variables utilisées dans la modélisation des pertes selon l'approche *risque de crédit*. Il est maintenant temps de se pencher plus en détail sur la manière dont ces variables sont modélisées par les assureurs-crédit.

1.3.2 Modèle multi-facteurs

Il existe très peu de littérature consacrée à l'analyse des risques en assurance-crédit. La plupart des recherches ont été menées dans l'optique d'évaluer le risque de crédit des institutions bancaires. C'est pourquoi les assureurs-crédit se sont appropriés les modèles classiques de crédit bancaire pour modéliser leur sinistralité. Ils semblent avoir opté pour un type de modèle de risque de crédit largement reconnu dans le secteur bancaire : le modèle de MERTON (1974).

Nous commencerons par une présentation du cadre simpliste du modèle de Merton. Ensuite, nous présenterons un modèle à facteurs gaussiens, couramment utilisé par les assureurs-crédits. Ce modèle reprend les mêmes concepts que le modèle de Merton et y intègre des facteurs systémiques permettant de tenir compte de l'environnement économique.

Modèle de Merton

Le modèle de Merton fait partie des modèles structurels de crédit, proposant de modéliser la survenance d'un défaut à partir de la *valeur* de l'entreprise. Ce modèle repose sur une représentation simplifiée du bilan d'une entreprise. La valeur des actifs de l'entreprise évolue selon un mouvement brownien géométrique et la dette de la compagnie est modélisée par une obligation zéro-coupon. Dans ce cadre, le modèle postule qu'une entreprise fait défaut lorsque le niveau de ses actifs tombe en dessous de la valeur de sa dette à un horizon de temps fixé. Ce dernier correspond à l'échéance de l'obligation qui symbolise la dette de l'entreprise.

Notons A_t la valeur de l'actif de l'entreprise à la date t , Merton suppose une dynamique Black-Scholes pour le processus $(A_t)_{t \geq 0}$ soit

$$\frac{dA_t}{A_t} = \mu dt + \sigma dW_t, \quad A_0 > 0, \quad (1.2)$$

où μ et σ représentent respectivement le rendement et la volatilité des actifs, et W_t est un mouvement brownien standard.

Le lemme d'Itô appliqué à l'EDS (1.2) permet d'obtenir une formule explicite pour la valeur de l'actif de l'entreprise à la date t

$$A_t = A_0 \times e^{((\mu - \frac{\sigma^2}{2})t + \sigma \sqrt{t}X)}, \quad X \sim \mathcal{N}(0, 1). \quad (1.3)$$

En se plaçant à l'échéance T , la valeur de la dette obligataire est égale au nominal D de l'obligation zéro-coupon et représente le seuil de défaut.

La probabilité de défaut de l'entreprise à la date T s'écrit comme suit

$$\mathbb{P}(A_T < D) = \mathbb{P}(X < d) = \Phi(d),$$

où $d = \frac{\ln(\frac{D}{A_0}) - (\mu - \frac{\sigma^2}{2})T}{\sigma\sqrt{T}}$ et $\Phi(\cdot)$ est la fonction de répartition de la loi normale centrée réduite.

Remarque 1.3. Dans ce modèle, l'évènement de défaut ne peut survenir qu'à la date T , correspondant à l'échéance de l'obligation.

La figure 1.7 résume l'approche du modèle de Merton. Le schéma représente deux trajectoires différentes pour la valeur des actifs de l'entreprise. La ligne en pointillés représente la valeur moyenne des actifs jusqu'à l'échéance, prédite par le modèle, tandis que la courbe verte dépeint la distribution de la valeur des actifs à l'échéance.

La courbe bleue présente un scénario relativement favorable pour l'entreprise, le niveau des actifs à l'échéance est légèrement supérieur à la moyenne modélisée par le modèle. La courbe orange montre quant à elle un scénario défavorable. La valeur des actifs tombe plusieurs fois en dessous du seuil de défaut et est en dessous de celui-ci à la date d'échéance. L'entreprise est donc considérée en défaut par le modèle.

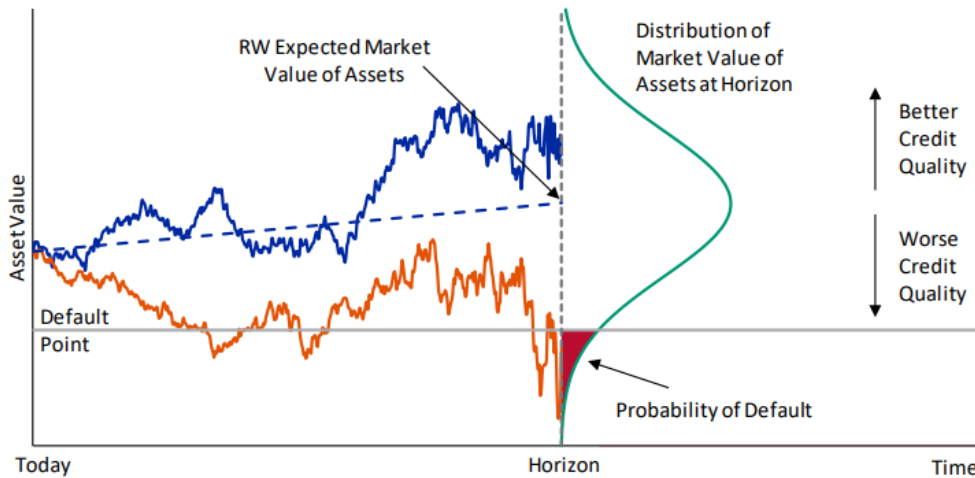


FIGURE 1.7 : Schéma modèle de Merton (THOMPSON et JESSOP, 2018)

Le modèle de Merton repose sur des hypothèses relativement simplificatrices, mais s'est imposé comme un modèle de référence pour la modélisation des probabilités de défaut. Il offre un premier cadre pour comprendre les liens entre le défaut d'une entreprise et son activité économique, ce qui constitue un point de départ dans la compréhension du risque de crédit.

Cependant, une des limites de ce modèle est qu'il n'intègre pas de corrélation entre les défauts, puisque les probabilités de défaut sont calculées indépendamment pour chaque entreprise. VASICEK (2002) propose une extension du modèle permettant d'introduire une structure de dépendance entre les défauts.

Considérons un portefeuille composé de n acheteurs susceptibles de faire défaut. Le modèle de Vasicek reprend le cadre du modèle de Merton. Ainsi, pour chaque entreprise k , la valeur des actifs de l'entreprise $A^{(k)}$ satisfait l'équation (1.2).

On a

$$A_t^{(k)} = A_0^{(k)} \times e^{((\mu_k - \frac{\sigma_k^2}{2})t + \sigma_k \sqrt{t} X_k)} \quad \text{où} \quad X_k \sim \mathcal{N}(0, 1), \quad (1.4)$$

et la corrélation entre les défauts est introduite en supposant que les X_k sont corrélées deux à deux selon un facteur ρ^2 , de telle sorte que chaque variable X_k peut être représentée de la manière suivante

$$X_k = \rho Z + \sqrt{1 - \rho^2} \epsilon_k, \quad (1.5)$$

où les variables aléatoires $Z, \epsilon_1, \dots, \epsilon_n$ sont indépendantes et identiquement distribuées (iid) de loi $\mathcal{N}(0, 1)$.

Les composantes de l'équation (1.5) peuvent s'interpréter comme suit :

- X_k incarne la performance (ou *capacité à payer*) de l'acheteur k ;
- Z correspond au risque systémique, facteur commun à toutes les entreprises du portefeuille représentant l'état de l'économie ;
- ρ représente la sensibilité de la performance de l'entreprise au risque systémique ;
- ϵ_k constitue le risque idiosyncratique, spécifique à l'acheteur k .

Comme pour le modèle de Merton, la probabilité de défaut de l'acheteur k s'écrit

$$p_k = \mathbb{P}(A_T^{(k)} < D_k) = \mathbb{P}(X_k < d_k),$$

et la probabilité de défaut de l'entreprise k conditionnellement au risque systémique Z s'obtient par le calcul suivant

$$\begin{aligned} p_k^*(z) &= \mathbb{P}(A_T^{(k)} < D_k | Z = z) \\ &= \mathbb{P}(X_k < d_k | Z = z) \\ &= \mathbb{P}(\rho Z + \sqrt{1 - \rho^2} \epsilon_k < d_k | Z = z) \\ &= \Phi\left(\frac{d_k - \rho z}{\sqrt{1 - \rho^2}}\right) \\ &= \Phi\left(\frac{\Phi^{-1}(p_k) - \rho z}{\sqrt{1 - \rho^2}}\right). \end{aligned}$$

Le modèle de Vasicek, en plus d'intégrer des corrélations entre les défauts, permet de calculer la probabilité de défaut des acheteurs conditionnellement à différents scénarios macroéconomiques, à travers la variable Z . Cependant, le fait de n'utiliser qu'un seul facteur de risque commun est un peu simpliste pour refléter l'environnement économique (et politique) des entreprises. Par ailleurs, le modèle tel que présenté par Vasicek dans son article suppose une corrélation constante entre les défauts des différentes entreprises, ce qui est contestable. Les modèles actuellement en place chez les assureurs-crédit généralisent l'approche proposée par Merton et Vasicek en intégrant plusieurs facteurs de risque systémiques.

Modèle multi-facteurs

Les modèles à facteurs, dont fait partie le modèle de Vasicek que nous venons de présenter, sont des modèles structurels de type Merton. Ils reposent sur l'idée que le défaut d'une entreprise survient lorsqu'une variable latente, représentant la santé financière de l'entreprise, franchit un certain seuil.

On considère un portefeuille d'assurance-crédit constitué de n entreprises dont nous surveillons le défaut. La valeur des actifs de chaque entreprise k est modélisée par un mouvement brownien géométrique $(A_k(t))_{t \geq 0}$ satisfaisant (1.4). Le défaut de l'entreprise k survient lorsque la variable X_k , qui est communément interprétée comme le log-rendement standardisé des actifs, tombe sous un certain seuil d_k .

Le modèle suppose une dépendance entre la variable latente X_k et N facteurs de risques gaussiens. Ces facteurs décrivent l'environnement dans lequel évoluent les entreprises. Ils peuvent représenter des indices macroéconomiques, géographiques, sectoriels ou tout autre indice pouvant avoir un impact sur les entreprises.

Le rendement standardisé des actifs X_k de l'entreprise k est donc modélisé comme suit

$$X_k = \rho_k R_k + \sqrt{1 - \rho_k^2} \epsilon_k, \quad (1.6)$$

où :

- $R_k = \sum_{l=1}^N w_{kl} Z_l$ est le facteur *composite*, combinaison linéaire des facteurs de risques systémiques $(Z_l)_{l=1, \dots, N}$. Les poids $(w_{kl})_{l=1, \dots, N}$ représentent la sensibilité de l'acheteur k à chaque facteur systémique ;
- ρ_k représente la corrélation entre X_k et le facteur composite. Plus ρ_k est grand, plus la corrélation avec les risques systémiques, et par extension avec les autres entreprises, est grande ;
- ϵ_k est le risque individuel de l'acheteur k .

Le vecteur $Z = (Z_1, \dots, Z_N)$ des facteurs systémiques suit une loi normale multivariée de matrice de covariance (non diagonale) Σ et les variables aléatoires $\epsilon_1, \dots, \epsilon_n$ sont iid de loi $\mathcal{N}(0, 1)$ et indépendantes des facteurs systémiques.

Les poids $(w_{kl})_{l=1, \dots, N}$ vérifient la relation $\sum_{l=1}^N w_{kl}^2 = 1$ afin de conserver l'hypothèse sur la distribution des rendements standardisés : $X_k \sim \mathcal{N}(0, 1)$.

La probabilité de défaut de l'acheteur k s'écrit toujours

$$p_k = \mathbb{P}(X_k < d_k),$$

et l'on peut calculer la probabilité de défaut de l'entreprise k conditionnellement aux facteurs de risques systémiques (Z_1, \dots, Z_N) par la formule suivante

$$\begin{aligned} p_k^*(z_1, \dots, z_N) &= \mathbb{P}(X_k < d_k | Z_1 = z_1, \dots, Z_N = z_N) \\ &= \Phi\left(\frac{\Phi^{-1}(p_k) - \rho_k \sum_{l=1}^N w_{kl} z_l}{\sqrt{1 - \rho_k^2}}\right). \end{aligned}$$

Dans ce cadre, on a donc un modèle à facteurs pour chaque entreprise susceptible de faire défaut. Pour chaque acheteur, le modèle décompose le risque de défaut en deux composantes : un risque individuel propre à chaque acheteur (et indépendant des autres risques) qui caractérise le risque

associé à la gestion de l'entreprise, et un risque global constitué de facteurs de risque communs à toutes les entreprises représentant le contexte économique. La corrélation entre les défauts est bien prise en compte dans le calcul de la probabilité de défaut et est déterminée à travers la corrélation entre les facteurs de risque communs.

En pratique, les assureurs-crédit segmentent leur portefeuille en groupes supposés homogènes. Par exemple, les entreprises peuvent être regroupées selon leur pays, leur secteur d'activité, leur classe de notation ou encore leur niveau d'encours assuré. La probabilité de défaut est alors déterminée pour chaque groupe homogène. Ainsi, deux entreprises appartenant au même groupe auront la même probabilité de défaut dans ce modèle.

L'avantage de cette segmentation est qu'elle permet de réduire le nombre de modèles à gérer, facilitant ainsi leur mise en œuvre. Il peut en effet être coûteux en temps de calcul d'avoir un modèle distinct pour chaque entreprise. De plus, les paramètres du modèle étant estimés sur la base des probabilités de défaut observées, le regroupement des entreprises permet d'obtenir des estimations plus robustes de ces paramètres, car on dispose alors de plus de données pour les calibrer.

Le modèle KMV, développé par Moody's (une présentation complète du modèle est consultable dans SAUNDERS et ALLEN (2010)), décompose le modèle multi-facteur en trois niveaux de risques. Cette décomposition, illustrée sur la figure 1.8, permet une compréhension détaillée des différents types de risques. Bien que la documentation des modèles internes utilisés par les assureurs-crédit ne soit pas accessible, des travaux tels que ceux de ULLMO (2019) et CAJA (2014) suggèrent que ces modèles sont assez proches du modèle KMV. C'est pourquoi nous avons choisi de présenter ce dernier dans le cadre de notre discussion.

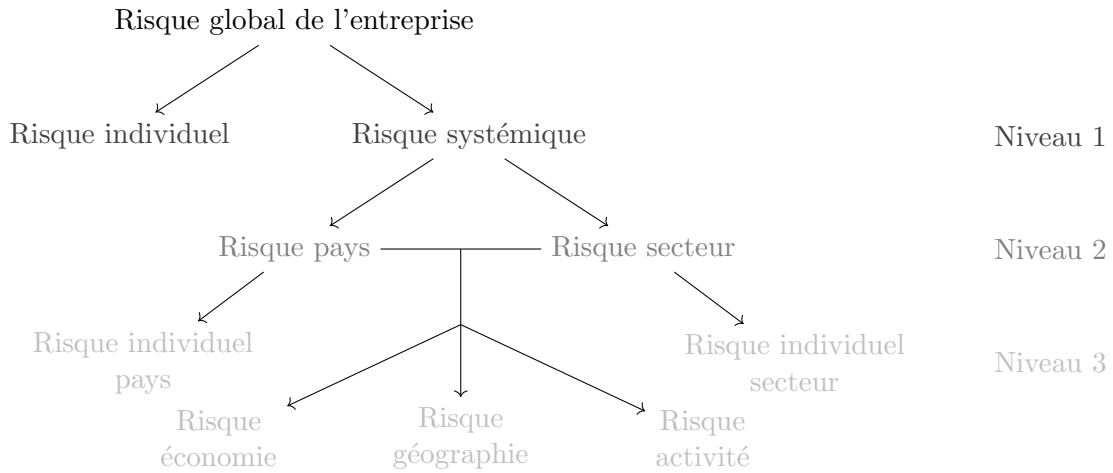


FIGURE 1.8 : Modèle multi-facteurs à trois niveaux

Au premier niveau, le modèle distingue le risque spécifique du risque systémique comme indiqué dans l'équation (1.6).

Ensuite, le second niveau sépare le risque systémique en facteurs de risque liés au pays de l'acheteur Z^P et ceux liés au secteur d'activité Z^S

$$R_k = \sum_{l=1}^{k^P} w_{kl}^P Z_l^P + \sum_{l=1}^{k^S} w_{kl}^S Z_l^S.$$

Enfin, le dernier niveau exprime les facteurs de risque pays et secteur comme une combinaison de

facteurs systémiques et d'un facteur idiosyncratique

$$Z_l^P = \sum_{m=1}^M w_{ml} Z_m + \epsilon_l^P \quad \text{et} \quad Z_l^S = \sum_{m=1}^M w_{ml} Z_m + \epsilon_l^S.$$

Les facteurs de risques communs Z_1, \dots, Z_M sont répartis en trois groupes. Un premier groupe capture les effets de l'économie mondiale, un second couvre les spécificités économiques régionales et un dernier englobe les risques liés à l'activité spécifique de l'entreprise, tel que l'impact de l'évolution du prix du pétrole dans le secteur du transport de marchandises.

La décomposition en trois niveaux du modèle KMV offre une analyse précise des risques auxquels peuvent être exposées les entreprises, intégrant l'impact global de l'économie sur les pays et les industries, tout en reconnaissant les spécificités inhérentes à chaque pays et secteur.

Une fois les paramètres calibrés, la distribution des pertes est obtenue à l'aide de simulations Monte-Carlo. Répéter N_{sim} fois la procédure suivante permet d'obtenir une distribution des pertes :

1. Simuler un scénario $z = (z_1, \dots, z_n)$ à partir de la loi $\mathcal{N}(0, \Sigma)$.
2. Pour chaque acheteur $k = 1, \dots, n$ du portefeuille :
 - (a) Simuler une réalisation e_k du risque spécifique ϵ_k de loi normale centrée réduite.
 - (b) Calculer le seuil de défaut $\xi_k = \frac{\Phi^{-1}(PD_k) - \rho_k \sum_{l=1}^N w_{kl} z_l}{\sqrt{1 - \rho_k^2}}$ où PD_k est la probabilité de défaut empirique calculée sur le segment dont fait partie l'acheteur k .
 - (c) Simuler les taux UGD_k et LGD_k
 - (d) Calculer la perte $l_k = L_k^{(0)} \times UGD_k \times LGD_k \times \mathbf{1}_{\{e_k < \xi_k\}}$.
3. Calculer la perte globale $L = \sum_{i=1}^n l_i$.

1.3.3 Limites

Bien que les modèles à facteurs aient prouvé leur efficacité pour modéliser les défauts en assurance-crédit, ils ne sont pas exempts de limites. En effet, leur calibrage repose sur des prévisions économiques à un instant donné et nécessite le jugement d'un expert. Pour rester pertinents et précis, ces modèles doivent être régulièrement mis à jour afin de refléter les conditions économiques du moment. Ils n'intègrent pas de façon intrinsèque les fluctuations de l'environnement économique.

Dans un contexte d'incertitude économique, comme celui engendré par la récente crise du COVID-19, la capacité à mesurer l'impact de la conjoncture sur la distribution des pertes de l'assureur est essentielle. Or, la nécessité de recalibrer le modèle à chaque fluctuation de l'environnement peut s'avérer coûteuse. De surcroît, les tensions politiques actuelles ajoutent une couche supplémentaire d'incertitude quant à l'évolution future de l'économie mondiale.

Face à ces défis, nous proposons dans la suite de ce mémoire une approche sensiblement différente pour étudier les pertes d'un assureur-crédit, tout en tenant compte en temps réel de l'évolution du contexte économique. Nous suggérons une modélisation de la fréquence des défauts par secteur d'activité en utilisant les processus de Hawkes. L'intensité de ces processus sera dépendante d'indicateurs économiques, et les propriétés d'excitation des processus permettront de modéliser des effets de contagion intra et intersectoriels, permettant de tenir compte de la dépendance entre défauts.

Notre objectif n'est pas de concurrencer le modèle à facteurs avec un modèle qui serait plus performant, mais plutôt de proposer une approche qui complète et pallie certaines de ses limites, en offrant une vision plus immédiate de l'évolution des pertes dans un contexte économique très incertain. Cette approche pourrait également se révéler précieuse pour les arbitres, en leur fournissant un outil supplémentaire dans leur gestion des limites.

Chapitre 2

Modélisation des défauts par les processus de Hawkes

Ce chapitre vise à présenter l'application des processus de Hawkes à la modélisation du nombre de défauts en assurance-crédit. Nous commencerons par une présentation théorique des processus de Hawkes, puis nous étudierons leur capacité à modéliser la fréquence des défauts dans un contexte économique fluctuant.

2.1 Processus de Hawkes

Les processus de HAWKES (1971), sont des processus de comptage pouvant s'apparenter à des processus de Poisson. Ils se distinguent de ces derniers par le fait que l'intensité d'un processus de Hawkes est un processus stochastique qui possède des propriétés d'excitation conditionnées au passé du processus de Hawkes. Autrement dit, chaque événement survenu dans le passé peut augmenter le taux d'occurrence des événements futurs.

Ces processus ont trouvé des applications dans divers domaines, comme par exemple en sismologie (OGATA, 1988), en finance (DUFFIE et al., 2003) et plus récemment en assurance pour l'étude des risques cyber (BESSY-ROLAND et al., 2021). Nous proposons dans ce mémoire une application à l'assurance-crédit.

Avant d'aborder l'utilisation des processus de Hawkes pour la modélisation du nombre de défauts, nous effectuons d'abord quelques rappels sur les processus de comptage afin d'établir les concepts et outils mathématiques dont nous aurons besoin pour définir formellement les processus de Hawkes. Nous procéderons ensuite à une présentation générale de ces processus, d'abord en une dimension, puis dans le cas multivarié.

2.1.1 Processus de comptage

Un processus de comptage est un processus stochastique qui opère en temps continu et prend ses valeurs dans l'ensemble des entiers naturels \mathbb{N} . Ce type de processus permet de suivre et de compter le nombre (aléatoire) d'événements d'un certain type sur une période de temps.

Définition 2.1. Un processus stochastique $N = (N_t)_{t \geq 0}$ est un processus de comptage s'il satisfait les propriétés suivantes :

- (i) $N_0 = 0$ et $N_t \in \mathbb{N} \cup \{+\infty\}$, $\forall t \geq 0$;

- (ii) N est croissant et càdlàg (continu à droite et admet une limite à gauche) ;
- (iii) les incréments sont de taille 1 : $\forall t \geq 0, N_{t+} - N_t \in \{0, 1\}$;
- (iv) $N_t \xrightarrow[t \rightarrow +\infty]{} +\infty$ *p.s.*

Remarque 2.1. La définition d'un processus de comptage peut varier selon les ouvrages. Certains auteurs n'intègrent pas les propriétés (iii) et (iv) dans leur définition. Nous avons choisi ici de les inclure car elles seront vérifiées dans le cadre classique des processus de Hawkes.

Un processus de comptage peut être vu comme le dénombrement d'événements survenus. Ainsi, à chaque processus de comptage, on peut associer une suite de variables aléatoires (v.a) positives $(T_n)_{n \geq 0}$ représentant les temps de sauts du processus, c'est-à-dire les dates de survenance des événements comptés par le processus. Ces temps de sauts aléatoires forment un processus ponctuel simple.

Définition 2.2. Une suite de v.a $(T_n)_{n \geq 0}$ est appelée processus ponctuel simple si elle satisfait les conditions suivantes : elle prend ses valeurs dans \mathbb{R}^+ , elle est croissante presque sûrement (*i.e* $0 \leq T_1 \leq T_2 \leq \dots$ *p.s*) et ne converge pas.

Un processus de comptage $(N_t)_{t \geq 0}$ peut être défini en fonction de son processus ponctuel associé $(T_n)_{n \geq 0}$ par la relation suivante

$$N_t = \sum_{n \geq 0} \mathbf{1}_{T_n \leq t},$$

et de façon équivalente, le processus ponctuel peut s'exprimer en fonction du processus de comptage par l'égalité suivante

$$T_n = \inf\{t \geq 0 : N_t \geq n\}.$$

Les termes processus de comptage et processus ponctuel sont souvent utilisés de manière interchangeable dans la littérature. Par exemple, le terme processus de Poisson, qui est le processus de comptage le plus simple et le plus répandu, peut se référer au processus de comptage ou au processus ponctuel associé selon le contexte.

Définition 2.3. Un processus de Poisson (simple) d'intensité $\lambda > 0$, est un processus de comptage pour lequel les temps d'intervalle entre deux sauts successifs, notés $E_i = T_i - T_{i-1}$, sont i.i.d selon une loi exponentielle de paramètre λ .

L'hypothèse $E_i \stackrel{\text{i.i.d}}{\sim} \mathcal{E}(\lambda)$ est équivalente à la vérification des propriétés suivantes pour le processus de comptage :

1. $(N_t)_{t \geq 0}$ est à accroissements stationnaires et indépendants ;
2. $\forall t > s \geq 0, N_t - N_s \sim \mathcal{P}(\lambda(t - s))$.

Le paramètre λ d'un processus de Poisson s'interprète comme le nombre moyen d'événements qui se produisent par unité de temps. Par ailleurs, son inverse $\frac{1}{\lambda}$ représente le temps moyen entre chaque saut.

La définition précédente du processus de Poisson suppose un paramètre d'intensité λ constant au cours du temps, ce qui implique une survenance homogène des événements au fil du temps. Cependant, dans certaines situations, le taux d'occurrence des événements peut varier avec le temps. Pour prendre en compte cette variabilité, il est possible de définir un processus de Poisson où le paramètre d'intensité λ est une fonction du temps, $\lambda(t)$. On parle alors de processus de Poisson *non-homogène* puisque la répartition des temps de sauts n'est plus équilibrée au cours du temps.

Définition 2.4. Soit $\lambda : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ une fonction localement intégrable et $\Lambda : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ définie par

$$\Lambda(t) = \int_0^t \lambda(s) ds, \quad \forall t \geq 0.$$

Un processus de comptage $(N_t)_{t \geq 0}$ est un processus de Poisson non-homogène d'intensité $\lambda(\cdot)$ si :

- (i) $(N_t)_{t \geq 0}$ est à accroissements indépendants ;
- (ii) $\forall t > s \geq 0, N_t - N_s \sim \mathcal{P}(\Lambda(t) - \Lambda(s))$.

La proposition suivante décrit comment construire un processus de Poisson non-homogène à partir d'un processus de Poisson simple, et inversement.

Proposition 2.1. Si $(N_t)_{t \geq 0}$ est un processus de Poisson simple d'intensité 1, alors le processus $(\tilde{N}_t)_{t \geq 0}$ défini par

$$\tilde{N}_t = N_{\Lambda(t)} \quad \text{où} \quad \Lambda(t) = \int_0^t \lambda(s) ds,$$

est un processus de Poisson non-homogène d'intensité $\lambda(\cdot)$.

Réciproquement, si $(\tilde{N}_t)_{t \geq 0}$ est un processus de Poisson non-homogène d'intensité $\lambda(\cdot)$, alors le processus $(N_t)_{t \geq 0}$ défini par

$$N_t = \tilde{N}_{\Lambda^{-1}(t)} \quad \text{où} \quad \Lambda^{-1}(t) = \inf\{x \in \mathbb{R}^+ : \Lambda(x) \geq t\},$$

est un processus de Poisson simple d'intensité 1.

Nous terminons cette section en introduisant un dernier élément fondamental pour la compréhension des sections suivantes : l'intensité conditionnelle.

Définition 2.5. Soit $(N_t)_{t \geq 0}$ un processus de comptage, le processus d'intensité conditionnelle associé $(\lambda^*(t))_{t \geq 0}$ est défini par

$$\lambda^*(t) = \lambda(t|\mathcal{F}_t) = \lim_{h \rightarrow 0} \mathbb{E} \left[\frac{N_{t+h} - N_t}{h} | \mathcal{F}_t \right] = \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}(N_{t+h} - N_t > 0 | \mathcal{F}_t), \quad t \geq 0,$$

où $\mathcal{F}_t = \sigma(\{N_s\}_{s \leq t})$ est la filtration caractérisant l'historique du processus de comptage jusqu'à la date t (non incluse).

Le processus d'intensité conditionnelle représente le taux de survenance instantané d'un événement à chaque date t . Il définit de façon unique la structure de probabilité du processus de comptage associé (voir DALEY et VERE-JONES, 2003, p.233). Il s'agit généralement d'un processus stochastique, le cas déterministe où $\lambda^*(t) = \lambda(t)$ renvoyant au processus de Poisson.

2.1.2 Processus de Hawkes univarié

Nous pouvons à présent introduire les processus de Hawkes, initialement appelés processus auto-excitants par HAWKES (1971). Nous commençons par le processus de Hawkes univarié.

Définition 2.6. Un processus de Hawkes est un processus de comptage $(N_t)_{t \geq 0}$ dont l'intensité conditionnelle est de la forme

$$\lambda^*(t) = \mu(t) + \int_0^t \phi(t-s) dN_s = \mu(t) + \sum_{T_n < t} \phi(t-T_n), \quad (2.1)$$

où $\mu : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ représente l'intensité de base du processus, $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ est la fonction d'excitation et $(T_n)_{n \geq 0}$ sont les temps de sauts du processus.

L'intensité conditionnelle d'un processus de Hawkes est donc constituée de deux composantes. Une première composante déterministe $\mu(\cdot)$ représentant le risque de base du processus et une seconde composante *auto-excitante* représentée par la somme dans la formule (2.1). Cette dernière traduit l'influence du passé du processus sur l'intensité conditionnelle.

La fonction d'excitation $\phi(\cdot)$ détermine de quelle manière un événement déjà survenu impacte le taux de survenance des événements futurs. En général, c'est une fonction décroissante de sorte que les événements récents auront plus d'impact sur l'intensité que les événements plus anciens. Un choix largement étudié dans la littérature pour définir la fonction d'excitation est le noyau exponentiel, défini par $\phi(u) = \alpha e^{-\beta u}$ où α et β sont des paramètres strictement positifs. Dans la suite de ce mémoire, nous nous concentrerons exclusivement sur cette forme de fonction d'excitation. Cette décision est motivée d'une part par la pertinence de l'hypothèse d'une fonction décroissante pour notre étude, et d'autre part par les simplifications qu'elle offre en termes de calculs.

Remarque 2.2. Si la fonction $\phi(\cdot)$ est nulle, le processus d'intensité défini en 2.1 devient déterministe et correspond alors à l'intensité d'un processus de Poisson.

Remarque 2.3. Dans la définition proposée par A.G. Hawkes, le risque de base du processus est défini comme un paramètre λ constant au fil du temps. Le choix d'une fonction déterministe dépendante du temps permet de prendre en compte des tendances dans le niveau de risque initial.

La figure 2.1 montre une simulation d'un processus de Hawkes avec noyau exponentiel ainsi que l'évolution de son intensité à chaque saut. Cette illustration met en évidence la propriété d'auto-excitation du processus.

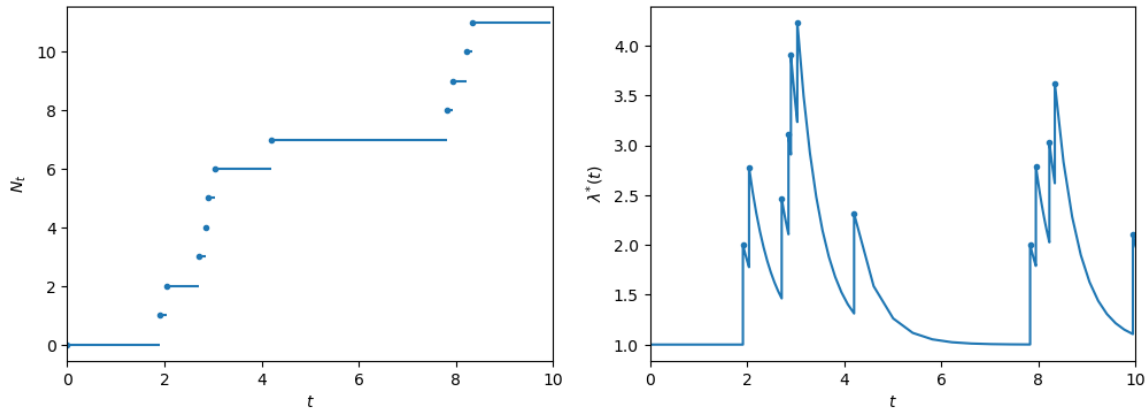


FIGURE 2.1 : Exemple de processus de Hawkes et son intensité

On distingue des périodes de forte activité, se manifestant par une accumulation de sauts sur une courte durée, comme le montre le graphique de gauche qui représente le processus de Hawkes. En parallèle, le graphique de droite, qui illustre l'évolution de l'intensité du processus, affiche des pics correspondant à ces mêmes périodes. On observe clairement que l'intensité conditionnelle augmente suite à chaque événement, puis décroît de manière exponentielle, revenant vers son niveau de base lors des périodes de faible activité.

Afin d'éviter une croissance non contrôlée du processus de comptage, qui pourrait résulter d'une accumulation d'événements augmentant de manière significative l'intensité couplée à une décroissance de l'intensité trop faible pour revenir à son niveau de base, il est nécessaire d'imposer la condition suivante

$$\mathbb{E}[\lambda^*(t)] < +\infty, \forall t \geq 0.$$

Pour satisfaire cette condition, il suffit de choisir une fonction $\mu(\cdot)$ bornée et de sélectionner judicieu-

sement les paramètres α et β de façon à ce que la propriété d'auto-excitation ne fasse pas exploser l'intensité conditionnelle. Pour cela, nous nous appuyons sur un résultat démontré dans LAUB et al. (2015). Il est établi qu'un processus de Hawkes dont l'intensité est de la forme

$$\tilde{\lambda}^*(t) = \lambda + \sum_{T_n < t} \alpha e^{-\beta(t-T_n)}$$

est stationnaire si $\alpha < \beta$. Pour un processus de Hawkes dont l'intensité conditionnelle $\lambda^*(.)$ est de la forme (2.1), nous avons

$$\lambda^*(t) = \tilde{\lambda}^*(t) - \lambda + \mu(t), \forall t \geq 0.$$

Par conséquent, en passant à l'espérance, si $\mu(.)$ est bornée et $\alpha < \beta$, nous obtenons bien que

$$\mathbb{E}[\lambda^*(t)] < +\infty, \forall t \geq 0.$$

Remarque 2.4. Un processus est stationnaire lorsque ses deux premiers moments (espérance et variance) sont finis et indépendants du temps.

Interprétation des paramètres

Comprendre l'influence des différents paramètres sur l'intensité conditionnelle est fondamental pour être en mesure d'interpréter nos résultats et vérifier leur cohérence.

Pour faciliter l'analyse, nous nous plaçons dans le cadre d'un processus de Hawkes dont la fonction d'excitation est représentée par un noyau exponentiel et dont le risque de base est constant en fonction du temps. Dans ce contexte, l'intensité conditionnelle prend la forme suivante

$$\lambda^*(t) = \lambda + \sum_{T_n < t} \alpha e^{-\beta(t-T_n)}, \quad (2.2)$$

où λ , α et β sont des paramètres strictement positifs.

Afin d'étudier le rôle de chaque paramètre, plusieurs simulations de processus de Hawkes ont été effectuées en faisant varier les paramètres individuellement.

Le paramètre λ symbolise le risque de base du processus et il est indépendant du passé du processus. Il définit d'une certaine manière le niveau de risque minimum du processus, la fonction d'excitation étant positive, le processus d'intensité est toujours supérieur à la valeur de λ . Ceci est illustré sur la figure 2.2, qui montre l'évolution de l'intensité conditionnelle d'un processus de Hawkes pour différentes valeurs du paramètre λ .

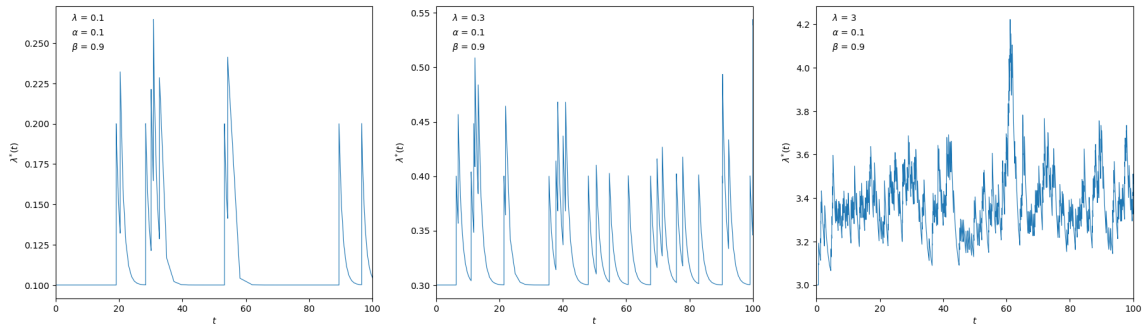


FIGURE 2.2 : Influence du paramètre λ

On remarque également que si λ est relativement élevé par rapport à α et β , le processus semble maintenir un niveau moyen supérieur à son risque de base, indiquant que le processus d'intensité est en permanence auto-excité.

Le paramètre β détermine la vitesse de décroissance de l'exponentielle vers son niveau de base. L'évolution de l'intensité en fonction de différentes valeurs de β est illustrée sur la figure 2.3.

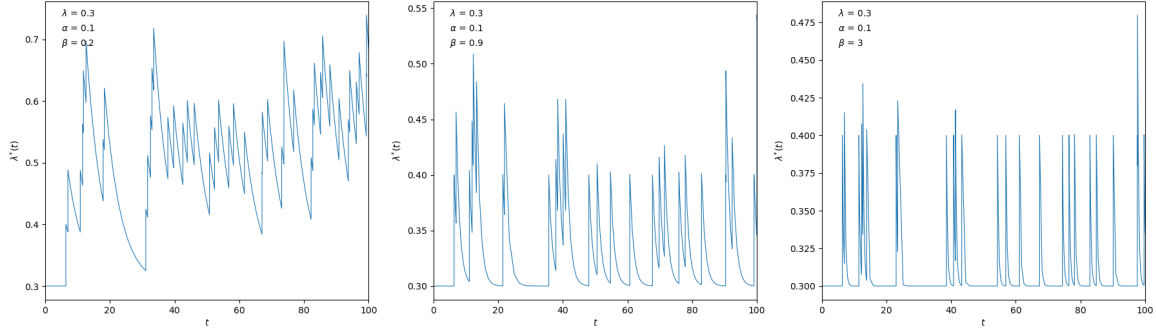


FIGURE 2.3 : Influence du paramètre β

Plus la valeur de β est élevée, plus la décroissance de l'intensité est rapide. Cependant, si β est trop élevé, l'intensité revient presque instantanément à son niveau initial, rendant la propriété d'excitation à peine perceptible. À l'inverse, si β est trop faible, l'intensité ne parvient pas à revenir à son niveau de base.

Le paramètre α est le paramètre d'excitation. La fonction d'excitation ϕ est décroissante et atteint son maximum α en 0. Par conséquent, l'excitation de l'intensité associée à un temps de saut T_n est maximale lorsque $t = T_n$, c'est-à-dire au moment de l'occurrence de l'événement. En d'autres termes, chaque fois qu'un événement survient, l'intensité augmente instantanément de α puis décroît progressivement.

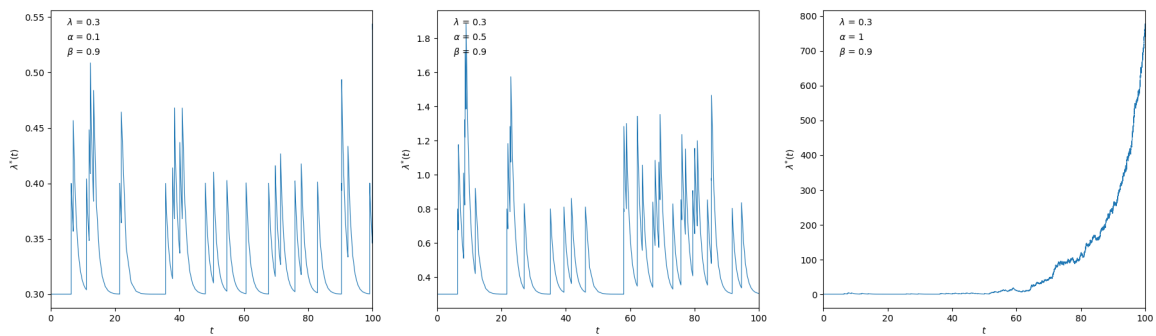


FIGURE 2.4 : Influence du paramètre α

Graphiquement, sur la figure 2.4, on peut observer des sauts de taille α à chaque occurrence d'un événement. Plus α est grand, plus les pics à chaque occurrence sont importants. Pour une même intensité de base, on observe que le niveau moyen de l'intensité conditionnelle augmente avec α . De plus, lorsque $\alpha > \beta$, c'est-à-dire lorsque le paramètre d'excitation est supérieur à la vitesse de décroissance de l'excitation, alors l'intensité conditionnelle semble diverger.

Simulation d'un processus de Hawkes

Pour simuler un processus de Hawkes, on peut recourir à l'algorithme de *thinning* modifié d'OGATA (1981). Cet algorithme est applicable à tout processus ponctuel dont l'intensité conditionnelle est connue, ce qui inclut les processus de Hawkes. L'algorithme opère en partant du fait que l'intensité conditionnelle est déterministe entre deux dates de survenance $[t_k, t_{k+1}[$. Y. Ogata a modifié l'algorithme proposé par LEWIS et SHEDLER (1979), conçu pour simuler un processus de Poisson non-homogène, afin de simuler des processus de comptage ayant une intensité conditionnelle stochastique. L'algorithme de Lewis est détaillé en annexe A.2.

Sachant $T_1 = t_1, \dots, T_k = t_k$, la génération du point t_{k+1} peut être vue comme la simulation du premier point d'un processus de Poisson non-homogène d'intensité déterministe $\lambda(t) = \lambda^*(t_k), t \geq t_k$. Après avoir généré le temps t_{k+1} , il suffit de mettre à jour l'intensité conditionnelle et l'intensité du processus de Poisson non-homogène pour générer un nouveau point.

Algorithme 1 : Simulation d'un processus de Hawkes sur $[0, T]$

Initialisation : $t \leftarrow 0, k \leftarrow 0$;

tant que $t < T$ **faire**

 Déterminer une borne λ_{sup} telle que $\lambda^*(u) \leq \lambda_{sup}, \forall t \leq u \leq T$;

 Simuler $\tau \sim \mathcal{E}(\lambda_{sup})$;

$t \leftarrow t + \tau$;

 Simuler $U \sim \mathcal{U}(0, 1)$;

si $U \leq \frac{\lambda^*(t)}{\lambda_{sup}}$ **et** $t < T$ **alors**

 On accepte le point candidat : $t_{k+1} \leftarrow t$;

$k \leftarrow k + 1$;

renvoyer $(t_i)_{i=1, \dots, k}$

L'algorithme 1 décrit la procédure proposée par Y. Ogata. Cet algorithme fonctionne pour toute fonction $\mu(\cdot)$ et $\phi(\cdot)$ tant qu'il est possible de trouver une borne pour la fonction d'intensité $\lambda^*(\cdot)$ entre chaque date d'occurrence. Le choix de cette borne joue un rôle important dans la vitesse d'exécution de l'algorithme. Plus cette borne est petite, plus le taux d'acceptation des temps simulés sera élevé, accélérant ainsi l'exécution de l'algorithme.

Dans le cas d'une intensité avec noyau exponentiel, il existe une propriété intéressante concernant l'intensité conditionnelle. En effet, cette dernière peut être calculée de manière récursive. Pour tout $s \in [t_k, t_{k+1}[$,

$$\begin{aligned} \lambda^*(s) &= \mu(s) + \sum_{t_i < s} \alpha e^{-\beta(s-t_i)} \\ &= \mu(s) + \sum_{t_i < s} \alpha e^{-\beta(t_k-t_i)} e^{-\beta(s-t_k)} \\ &= \mu(s) + (\lambda^*(t_k) + \alpha - \mu(t_k)) e^{-\beta(s-t_k)}. \end{aligned} \tag{2.3}$$

Cette propriété rend l'évaluation de la fonction $\lambda^*(\cdot)$ très efficace, ce qui contribue à améliorer la rapidité d'exécution de l'algorithme.

De plus, lorsque l'intensité conditionnelle adopte la forme (2.2), il est possible de définir à chaque étape $\lambda_{sup} = \lambda^*(t_k)$. En effet, étant donné que la fonction $\phi(\cdot)$ est décroissante et que la fonction $\mu(\cdot)$ est constante, $\lambda^*(\cdot)$ est également décroissante sur l'intervalle $[t_k, t_{k+1}[$ et atteint donc son maximum en t_k .

Calibrage des paramètres

Pour calibrer un processus de Hawkes, la méthode standard consiste à utiliser la méthode du maximum de vraisemblance. Nous allons nous appuyer sur le résultat suivant, qui fournit une formule pour calculer la vraisemblance d'un processus ponctuel. Le lecteur intéressé pourra consulter la démonstration de ce résultat dans LAUB et al. (2021).

Théorème 2.1. *Considérons $N = (N_t)_{t \geq 0}$ un processus de comptage d'intensité conditionnelle $\lambda^*(\cdot)$. Si l'on note t_1, \dots, t_n les temps de sauts du processus observés sur $[0, T]$, alors la fonction de vraisemblance \mathcal{L} de N est*

$$\mathcal{L} = \prod_{i=1}^n \lambda^*(t_i) e^{-\Lambda(T)} \quad \text{où} \quad \Lambda(T) = \int_0^T \lambda^*(u) du, \quad (2.4)$$

et la log-vraisemblance ℓ s'écrit

$$\ell = \sum_{i=1}^n \log(\lambda^*(t_i)) - \Lambda(T). \quad (2.5)$$

En appliquant ce résultat à un processus de Hawkes dont l'intensité est de la forme (2.1), on obtient la fonction de vraisemblance suivante

$$\ell(\mu, \phi) = - \int_0^T \mu(u) du - \int_0^T \sum_{t_k < u} \phi(u - t_k) du + \sum_{i=1}^n \log \left(\mu(t_i) + \sum_{k=1}^{i-1} \phi(t_i - t_k) \right). \quad (2.6)$$

En remarquant que

$$\begin{aligned} \int_0^T \sum_{t_k < u} \phi(u - t_k) du &= \sum_{i=0}^n \int_{t_i}^{t_{i+1}} \sum_{t_k < u} \phi(u - t_k) du \quad \text{où} \quad t_{n+1} := T \\ &= \sum_{i=0}^n \sum_{k=0}^i \int_{t_i}^{t_{i+1}} \phi(u - t_k) du \\ &= \sum_{i=0}^n \sum_{k=0}^i \int_{t_i - t_k}^{t_{i+1} - t_k} \phi(u) du \\ &= \sum_{i=1}^n \int_0^{T - t_i} \phi(u) du, \end{aligned}$$

on obtient

$$\ell(\mu, \phi) = - \int_0^T \mu(u) du - \sum_{i=0}^n \int_0^{T - t_i} \phi(u) du + \sum_{i=1}^n \log \left(\mu(t_i) + \sum_{k=1}^{i-1} \phi(t_i - t_k) \right). \quad (2.7)$$

Lorsque la fonction d'excitation $\phi(\cdot)$ est le noyau exponentiel, $\phi(t) = \alpha e^{-\beta t}$, la fonction de log-vraisemblance devient

$$\ell(\mu, \alpha, \beta) = - \int_0^T \mu(u) du - \frac{\alpha}{\beta} \sum_{i=0}^n (1 - e^{-\beta(T - t_i)}) + \sum_{i=1}^n \log \left(\mu(t_i) + \alpha \sum_{k=1}^{i-1} e^{-\beta(t_i - t_k)} \right).$$

On définit $A(i) := \sum_{k=1}^{i-1} e^{-\beta(t_i - t_k)}$, $i = 2, \dots, n$, et on pose $A_0 = 0$. Ce terme peut être calculé récursivement, en effet

$$\begin{aligned} A(i) &= \sum_{k=1}^{i-1} e^{-\beta(t_i - t_k)} \\ &= e^{-\beta(t_i - t_{i-1})} \sum_{k=1}^{i-1} e^{-\beta(t_{i-1} - t_k)} \\ &= e^{-\beta(t_i - t_{i-1})} \left(\sum_{k=1}^{i-2} e^{-\beta(t_{i-1} - t_k)} + 1 \right) \\ &= e^{-\beta(t_i - t_{i-1})} (A(i-1) + 1). \end{aligned}$$

La fonction de log-vraisemblance se réécrit alors

$$\ell(\mu, \alpha, \beta) = - \int_0^T \mu(u) du - \frac{\alpha}{\beta} \sum_{i=0}^n (1 - e^{-\beta(T - t_i)}) + \sum_{i=1}^n \log(\mu(t_i) + \alpha A(i)). \quad (2.8)$$

L'intérêt de cette dernière écriture réside dans le fait qu'il n'est pas possible de calculer analytiquement les paramètres optimaux pour cette log-vraisemblance. Il est nécessaire de recourir à des méthodes numériques pour estimer ces paramètres. Ces méthodes nécessitent d'évaluer la log-vraisemblance à chaque itération et le fait de réécrire le terme $A(i)$ récursivement permet un gain de temps dans l'évaluation de ℓ , qui peut s'avérer considérable lorsque le nombre de temps de sauts est grand.

2.1.3 Processus de Hawkes multivarié

La définition d'un processus de Hawkes peut être étendue à un cadre multivarié, offrant alors la possibilité de modéliser des effets d'auto-excitation, mais aussi des effets d'inter-excitation entre différents processus de comptage.

Définition 2.7. Considérons $d \in \mathbb{N}^*$ processus de comptage, notés $(N_t^{(1)})_{t \geq 0}, \dots, (N_t^{(d)})_{t \geq 0}$. Pour chaque $k \in \{1, \dots, d\}$, nous notons $(T_n^{(k)})_{n \geq 0}$ le processus ponctuel associé au processus $(N_t^{(k)})_{t \geq 0}$.

Le processus de comptage multivarié associé est défini par $N_t = (N_t^{(1)}, \dots, N_t^{(d)})$. Ce processus est un processus de Hawkes multivarié si chaque processus $(N_t^{(k)})_{t \geq 0}, k \in \{1, \dots, d\}$, a une intensité conditionnelle de la forme

$$\lambda_k^*(t) = \mu_k(t) + \sum_{j=1}^d \int_0^t \phi_{k,j}(t-s) dN_s^{(j)} = \mu_k(t) + \sum_{j=1}^d \sum_{T_n^{(j)} < t} \phi_{k,j}(t - T_n^{(j)}), \quad (2.9)$$

où $\mu_k : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ et $\phi_{k,j} : \mathbb{R}^+ \rightarrow \mathbb{R}^+$.

Comme pour le processus de Hawkes univarié, l'intensité conditionnelle de chaque processus de comptage $(N_t^{(k)})_{t \geq 0}$ est constituée d'une composante déterministe $\mu_k(\cdot)$ représentant l'intensité de base du processus, indépendante des événements passés, et d'une seconde composante qui caractérise l'influence du passé des processus sur l'intensité du processus $(N_t^{(k)})_{t \geq 0}$. Cette dernière est représentée par la double somme dans la formule (2.9).

Chaque événement passé, associé à l'un des processus constituant $(N_t)_{t \geq 0}$, a une influence sur l'intensité du processus $(N_t^{(k)})_{t \geq 0}$. Dans la double somme, $j = k$ caractérise le phénomène d'auto-excitation : les événements passés du processus lui-même impactent son intensité. En revanche, $j \neq k$ caractérise les effets d'inter-excitation : les événements passés des autres processus influencent l'intensité de $(N_t^{(k)})_{t \geq 0}$.

La fonction d'excitation $\phi_{k,j}(\cdot)$ détermine comment un événement passé du processus $(N_t^{(j)})_{t \geq 0}$ influence le taux de survenance instantané des événements futurs du processus $(N_t^{(k)})_{t \geq 0}$. Tout comme dans le cas univarié, il est préférable que cette fonction soit décroissante, de sorte que les événements les plus récents aient un impact plus important que les événements plus anciens. Nous nous intéresserons uniquement au noyau exponentiel $\phi_{k,j}(u) = \alpha_{k,j} e^{-\beta_k u}$, où $\alpha_{k,j}$ et β_k sont des paramètres strictement positifs, pour les raisons évoquées dans le cas univarié.

Pour $k \in \{1, \dots, d\}$, l'intensité conditionnelle du processus $(N_t^{(k)})_{t \geq 0}$ prend alors la forme suivante

$$\lambda_k^*(t) = \mu_k(t) + \sum_{j=1}^d \sum_{T_n^{(j)} < t} \alpha_{k,j} e^{-\beta_k(t-T_n^{(j)})}, \quad (2.10)$$

où $\mu_k : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, $\beta_k > 0$ et les $(\alpha_{k,j})_{j=1, \dots, d}$ sont des paramètres positifs.

Remarque 2.5. Il était possible d'envisager des paramètres $\beta_{k,j}$ distincts pour chaque temps de saut $T_n^{(j)}$, ce qui aurait permis d'illustrer un impact variable sur la décroissance de l'intensité en fonction du processus associé à l'événement survenu. Néanmoins, nous avons décidé d'adopter $\beta_{k,j} = \beta_k$ pour tout j . Ce choix, en plus de simplifier la simulation et le calibrage du processus, facilite également l'interprétation des résultats. Nous estimons que cette hypothèse n'entrave pas la cohérence de notre modèle.

Interprétation des paramètres

L'interprétation des paramètres dans le cadre multivarié suit le même principe que dans le cas univarié. Plus précisément :

- le paramètre $\alpha_{k,j}$ correspond à l'excitation du processus $(N_t^{(k)})_{t \geq 0}$ lors d'un saut du processus $(N_t^{(j)})_{t \geq 0}$. A cet instant, l'intensité conditionnelle $\lambda_k^*(\cdot)$ augmente immédiatement de $\alpha_{k,j}$;
- le paramètre β_k représente la vitesse de décroissance de l'impact d'un saut sur l'intensité conditionnelle $\lambda_k^*(\cdot)$.

Afin d'illustrer le phénomène d'inter-excitation propre au processus de Hawkes multivarié, nous proposons un exemple de processus de Hawkes bivarié ($d = 2$) sur la figure 2.5. Cette figure montre également l'évolution de son intensité. Les paramètres du processus sont les suivants :

$$\mu(t) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ pour tout } t \geq 0; \alpha = \begin{pmatrix} 0.1 & 0.2 \\ 0.4 & 0.3 \end{pmatrix} \text{ et } \beta = \begin{pmatrix} 1.5 \\ 1 \end{pmatrix}.$$

Sur les graphiques, le processus $N^{(1)}$ est représenté en bleu et le processus $N^{(2)}$ en orange. Dans cet exemple, tous les paramètres $\alpha_{j,k}$ sont non nuls, ce qui signifie que chaque événement influence l'intensité conditionnelle des deux processus.

Le processus $N^{(2)}$ présente une intensité de base nulle. Ainsi, à l'instant $t = 0$, la probabilité qu'un événement associé au processus $N^{(2)}$ survienne est nulle et le restera tant qu'aucun événement du processus $N^{(1)}$ ne se produit. En raison de l'inter-excitation entre les deux processus, symbolisée par

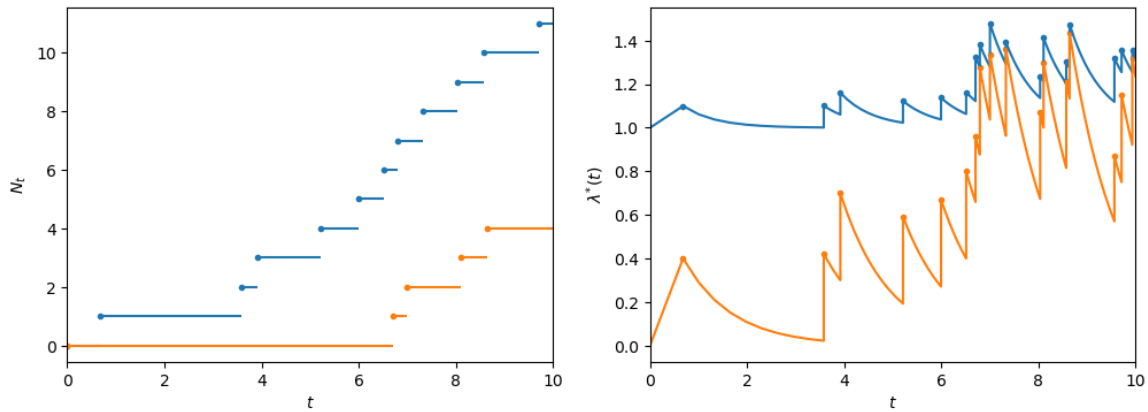


FIGURE 2.5 : Exemple processus de Hawkes bivarié et son intensité

$\alpha_{2,1}$ ici, l'émergence d'un événement dans le processus $N^{(1)}$ provoque une augmentation du taux de survénance du processus $N^{(2)}$, le faisant passer à une valeur positive. Ensuite, grâce à la propriété d'auto-excitation, la probabilité de survénance d'un événement du processus $N^{(2)}$ peut augmenter davantage. Cependant, si aucun événement du processus $N^{(1)}$ ou $N^{(2)}$ ne survient pendant un certain temps, l'intensité du processus $N^{(2)}$ peut retomber à zéro. Dans ce cas, il faudra attendre la survénance d'un nouvel événement du processus $N^{(1)}$ pour que le taux de survénance du processus $N^{(2)}$ redevienne positif. C'est ce que nous observons jusqu'à $t = 6$, l'inter-excitation de $N^{(1)}$ vers $N^{(2)}$, combinée à l'auto-excitation de $N^{(2)}$, permet de « démarrer » progressivement le processus $N^{(2)}$. Par la suite, nous observons une accélération dans les comptages, qui est due aux effets de l'inter-excitation et de l'auto-excitation des processus.

Simulation d'un processus de Hawkes multivarié

Pour la simulation d'un processus de Hawkes multivarié, un algorithme est également proposé dans OGATA (1981). Il s'agit d'une extension de l'algorithme de *thinning* que nous avons présenté dans la section précédente. L'algorithme s'appuie sur la proposition qui suit.

Proposition 2.2. *Considérons $N_t = (N_t^{(1)}, \dots, N_t^{(d)})$ un processus de Hawkes multivarié sur $[0, T]$ tel que défini en (2.9). Supposons qu'il existe un processus unidimensionnel F_t^N -prévisible, noté λ_t^* , satisfaisant*

$$\sum_{j=1}^d \lambda_j^*(t) \leq \bar{\lambda}_t^*, \quad 0 < t \leq T,$$

et définissons

$$\lambda_0^*(t) = \bar{\lambda}_t^* - \sum_{j=1}^d \lambda_j^*(t).$$

Soit $\bar{t}_1, \dots, \bar{t}_n$ les temps de sauts observés sur $[0, T]$ du processus de Poisson \bar{N}_t d'intensité $\bar{\lambda}_t^*$. Pour chaque point \bar{t}_k , $k = 1, \dots, n$, on associe une étiquette $D \in \{0, 1, \dots, d\}$ avec probabilité $\frac{\lambda_D^*(t)}{\lambda_t^*}$ respectivement.

Alors les points associés aux étiquettes $D = 1, \dots, d$ forment un processus de Hawkes multivarié d'intensité $\lambda_D^*(t)$, $D = 1, \dots, d$.

La proposition 2.2 assure que la simulation d'un processus de Hawkes multivarié peut s'effectuer à partir d'un processus de Poisson non-homogène comme dans le cas univarié. Générer un processus

de Hawkes multivarié consiste en fait à simuler un processus de Hawkes univarié dont l'intensité est définie par $\lambda^*(t) = \sum_{j=1}^d \lambda_j^*(t)$. Une étape supplémentaire est ajoutée à l'algorithme pour déterminer à quel processus $N^{(k)}$, $k \in \{1, \dots, d\}$, le point accepté est associé.

La procédure générale pour simuler un processus de Hawkes multivarié est résumée dans l'algorithme 2 ci-dessous.

Algorithme 2 : Simulation d'un processus de Hawkes multivarié de dimension d sur $[0, T]$

Initialisation : $t \leftarrow 0$, $k \leftarrow 0$;

tant que $t < T$ **faire**

 Déterminer λ_{sup} tel que $\sum_{j=1}^d \lambda_j^*(u) \leq \lambda_{sup}$ pour $t < u \leq T$;

 Simuler $\tau \sim \mathcal{E}(\lambda_{sup})$; $t \leftarrow t + \tau$;

 Simuler $U \sim \mathcal{U}(0, 1)$;

si $U \leq \frac{\sum_{j=1}^d \lambda_j^*(u)}{\lambda_{sup}}$ **et** $t < T$ **alors**

 On accepte le point candidat : $t_{k+1} \leftarrow t$;

 On détermine l'étiquette D_{k+1} associé au point t_{k+1} comme le plus petit entier m tel que $U \leq \frac{\sum_{j=0}^m \lambda_j^*(u)}{\lambda_{sup}}$;

$k \leftarrow k + 1$

renvoyer $\{ (t_i, D_i)_{i=1, \dots, k} : D_i \in \{1, \dots, d\} \}$

Lorsque l'intensité conditionnelle de chaque processus $(N_t^{(k)})_{t \geq 0}$ est de la forme

$$\lambda_k^*(t) = \lambda_k + \sum_{j=1}^d \sum_{T_n^{(j)} < t} \alpha_{k,j} e^{-\beta_k(t - T_n^{(j)})}, \quad (2.11)$$

on peut définir pour $t \in [t_k, t_{k+1}[$, $\lambda_{sup} = \sum_{j=1}^d \lambda_j^*(t_k)$. Cela est possible car la fonction $\lambda^*(.)$ est décroissante sur l'intervalle $[t_k, t_{k+1}[$ et atteint son maximum en t_k . Dans ce cas $\lambda_0^*(t) = 0$ et si un point candidat t est accepté, la distribution de l'étiquette D associée à ce point est la suivante

$$\mathbb{P}(D = i) = \frac{\lambda_i^*(t)}{\sum_{j=1}^d \lambda_j^*(t)}, \quad i = 1, \dots, d.$$

Comme dans le cas univarié, cette forme permet une évaluation très efficace de $\lambda^*(.)$. En effet, il est possible de calculer les fonctions $\lambda_j^*(.)$ de manière récursive, puis de les sommer pour obtenir $\lambda^*(.)$.

Fonction de vraisemblance

Pour obtenir l'expression de la vraisemblance d'un processus de Hawkes multivarié, nous pouvons nous baser sur la vraisemblance d'un processus ponctuel multivarié, qui est donnée dans le théorème suivant.

Théorème 2.2. *Considérons $N = (N^{(1)}, \dots, N^{(d)})$ un processus ponctuel multivarié d'intensité conditionnelle $\lambda^*(.) = (\lambda_1^*(.), \dots, \lambda_d^*(.))$. Si nous notons $\{(t_1, d_1), \dots, (t_n, d_n)\}$, où $n = \sum_{j=1}^d N_T^{(j)}$, les couples (temps de saut, étiquette) du processus observés sur $[0, T]$, alors la fonction de vraisemblance \mathcal{L} de N est*

$$\mathcal{L} = \left[\prod_{i=1}^n \lambda_{d_i}^*(t_i) \right] e^{-\sum_{j=1}^d \Lambda_j(T)} \quad \text{où} \quad \Lambda_j(T) = \int_0^T \lambda_j^*(u) du, \quad (2.12)$$

et la log-vraisemblance ℓ s'écrit

$$\ell = \sum_{i=1}^n \log(\lambda_{d_i}^*(t_i)) - \sum_{j=1}^d \Lambda_j(T). \quad (2.13)$$

Remarque 2.6. Ce théorème est présenté dans LAUB et al. (2021). Si on note, pour chaque $j \in \{1, \dots, d\}$, $(t_n^j)_{n=1, \dots, m_j}$ les temps de sauts du processus $N^{(j)}$ observés sur $[0, T]$, alors la formule donnée ci-dessus est équivalente à l'expression suivante

$$\ell = \sum_{j=1}^d \log \mathcal{L}^{(j)},$$

où

$$\log \mathcal{L}^{(j)} = \sum_{n=1}^{m_j} \log(\lambda_j^*(t_n^j)) - \Lambda_j(T), \quad 1 \leq j \leq d,$$

dont la démonstration est consultable dans CHEN (2016a).

Pour un processus de Hawkes multivarié avec des fonctions d'excitation exponentielles, tel que défini en (2.10), il est possible de calculer les intensités conditionnelles de manière récursive. Pour $s \in [t_k, t_{k+1}[$, la formule (2.3) que nous avons dans le cas univarié devient

$$\lambda_j^*(s) = \mu_j(s) + (\lambda_j^*(t_k) + \alpha_{k,d_k} - \mu_j(t_k)) e^{-\beta_j(s-t_k)}, \quad (2.14)$$

qui peut être réécrite en notation vectorielle comme suit

$$\boldsymbol{\lambda}^*(s) = \boldsymbol{\mu}(s) + (\boldsymbol{\lambda}^*(t_k) + \boldsymbol{\alpha}_{d_k} - \boldsymbol{\mu}(t_k)) e^{-\boldsymbol{\beta}(s-t_k)}, \quad (2.15)$$

$$\text{où } \boldsymbol{\lambda}^*(.) = \begin{pmatrix} \lambda_1^*(.) \\ \vdots \\ \lambda_d^*(.) \end{pmatrix}, \boldsymbol{\mu}(.) = \begin{pmatrix} \mu_1(.) \\ \vdots \\ \mu_d(.) \end{pmatrix}, \boldsymbol{\alpha}_{d_k} = \begin{pmatrix} \alpha_{1,d_k} \\ \vdots \\ \alpha_{d,d_k} \end{pmatrix} \text{ et } \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_d \end{pmatrix}.$$

D'autre part, on peut calculer pour $j = 1, \dots, d$

$$\Lambda_j(T) = \int_0^T \lambda_j^*(u) du = \int_0^T \mu_j(u) du + \sum_{(t_i, d_i): t_i < T} \frac{\alpha_{j,d_i}}{\beta_j} (1 - e^{-\beta_j(T-t_i)}), \quad (2.16)$$

qui se réécrit

$$\boldsymbol{\Lambda}(T) = \begin{pmatrix} \Lambda_1(T) \\ \vdots \\ \Lambda_d(T) \end{pmatrix} = M(T) + \sum_{(t_i, d_i): t_i < T} (\boldsymbol{\alpha}_{d_i} \circ \frac{\mathbf{1}}{\boldsymbol{\beta}}) (1 - e^{-\boldsymbol{\beta}(T-t_i)}), \quad (2.17)$$

$$\text{où } M(T) = \begin{pmatrix} \int_0^T \mu_1(u) du \\ \vdots \\ \int_0^T \mu_d(u) du \end{pmatrix}, \frac{\mathbf{1}}{\boldsymbol{\beta}} = \begin{pmatrix} \frac{1}{\beta_1} \\ \vdots \\ \frac{1}{\beta_d} \end{pmatrix} \text{ et } \circ \text{ représente le produit de Hadamard.}$$

Ces formules permettent de simplifier le calcul de la vraisemblance (2.13). Le calcul de la première composante est optimisé par la formule (2.15) permettant de calculer $\boldsymbol{\lambda}^*(.)$ récursivement. Et la formule (2.17) donne une expression explicite (à condition de pouvoir calculer $M(T)$) pour évaluer la seconde composante.

2.2 Modélisation du nombre de défauts

Dans la section précédente, nous avons introduit le cadre théorique des processus de Hawkes. Dans le premier chapitre de ce mémoire, nous avons exploré les diverses sources de risque en assurance-crédit, notamment la nécessité de prendre en compte les fluctuations du contexte économique et l'existence d'une structure de dépendance entre les défauts. Nous avons vu que les modèles en place représentent l'apparition d'un défaut à travers une probabilité de défaut sur des groupes homogènes d'entreprises.

Dans cette section, nous proposons une approche alternative : modéliser la fréquence des défauts d'un assureur-crédit par un processus de Hawkes multivarié. Chaque composante de ce processus représente le nombre de défauts d'un groupe homogène d'entreprises. N'ayant pas à disposition de données réelles d'assureur-crédit, nous baserons notre analyse sur les données publiques de défaillances d'entreprises, publiées mensuellement par la Banque de France. Nous regrouperons les entreprises par secteur d'activité, en supposant que les entreprises d'un même secteur sont exposées aux mêmes facteurs de risque. Cependant, la méthodologie détaillée dans cette section se veut généralisable à des regroupements plus fins, permettant une éventuelle application de cette approche à des données réelles d'assureur-crédit.

2.2.1 Description des données

Avant de passer à la présentation de notre modèle, il est nécessaire de commencer par une brève description des données utilisées pour cette étude. Nous expliquons également les quelques retraitements effectués.

Nous utilisons les données de défaillances d'entreprises fournies par la Banque de France. Selon la définition de la Banque de France, le terme *défaillance* désigne « l'ouverture d'une procédure collective donnant lieu au dépôt d'une déclaration de cessation de paiement ». Ainsi, les défaillances répertoriées dans cette base de données englobent les redressements et liquidations judiciaires, mais n'incluent pas les procédures de sauvegarde. Cela correspond donc à des défauts de type *Insolvability*, généralement irrécouvrables. Dans cette étude, les défauts de type *Protracted* ne seront donc pas modélisés ; il faudra tenir compte de ce facteur lors de l'évaluation des pertes. Les défaillances sont comptabilisées en termes d'unités légales, c'est-à-dire les entités identifiées par un numéro SIREN. La notion d'unité légale n'est pas équivalente à celle d'entreprise, cette dernière pouvant être une combinaison de plusieurs unités légales.

La base de données recense un total de 1 814 902 défaillances d'entreprises sur le territoire français depuis 1990. Ces informations sont disponibles sur une base mensuelle et par secteurs d'activités. Les secteurs d'activités considérés sont les suivants :

- Agriculture, sylviculture et pêche ;
- Industrie ;
- Construction ;
- Commerce et réparation automobile ;
- Transports et entreposage ;
- Hébergement et restauration ;
- Information et communication ;
- Activités financières et d'assurance ;
- Activités immobilières ;
- Soutien aux entreprises ;
- Enseignement, santé humaine, action sociale et services aux ménages.

La figure 2.6 dépeint l'évolution du nombre de défaillances d'entreprises en France depuis 1990. De façon très globale, les secteurs semblent suivre des tendances similaires, conséquence des fluctuations économiques. On observe ainsi des pics de défaillances lors de la récession de 1992-1993 en France, lors de la crise des subprimes en 2008 et, à l'inverse, une chute drastique lors de la crise du Covid-19. Ce graphique, en plus de re-souligner l'importance de prendre en compte le contexte économique, nous permet d'envisager l'hypothèse d'un effet de contagion entre les différents secteurs d'activités. En effet, on peut remarquer que les courbes de certains secteurs semblent suivre des tendances similaires au fil du temps. Par exemple, les secteurs de la construction et du commerce présentent des pics similaires, bien que leur intensité puisse varier.

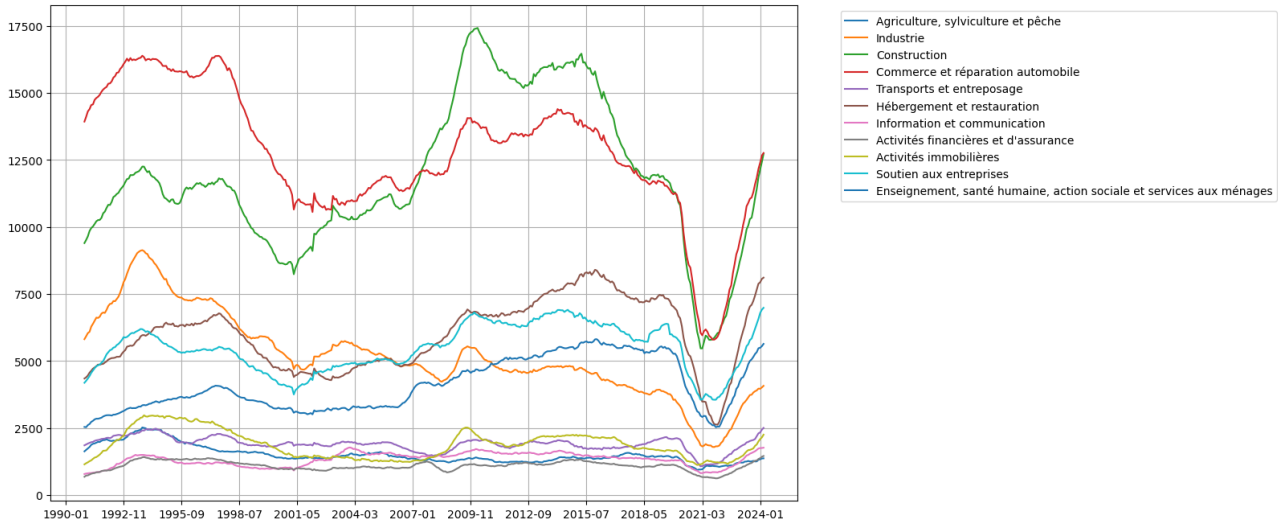


FIGURE 2.6 : Évolution du nombre de défaillances par secteur d'activité sur 12 mois glissants

Les données fournies sont de qualité, exemptes d'erreurs et ne comportent pas de valeurs manquantes. Notre travail de préparation des données se restreint donc à adapter ces dernières au format requis pour le modèle de Hawkes. En effet, pour calibrer notre modèle, il est nécessaire d'avoir les dates de survenance précises des défauts, car comme vu dans la section précédente, l'estimation des paramètres s'effectue par maximum de vraisemblance, cette dernière étant définie en fonction des temps de sauts observés. Nous disposons uniquement de données mensuelles, nous avons donc le total des défauts survenus chaque mois mais n'avons pas d'informations sur la répartition de ces défaillances au cours du mois. Afin de pallier ce problème, nous étudierons trois solutions en sections suivantes :

- répartition de tous les défauts en fin de mois : nous supposons que tous les défauts surviennent à la fin du mois avec un écart de ϵ entre chaque temps de sauts ;
- répartition uniforme des défaillances au sein de chaque mois ;
- répartition *hawkes* : pour chaque mois, nous simulons un processus de Hawkes univarié dont les paramètres sont choisis arbitrairement, jusqu'à obtenir le nombre total de défaillances dans le mois. Nous réajustons ensuite les dates d'occurrences obtenues pour qu'elles correspondent à une survenance dans le mois que nous simulons.

Bien que ces trois approches puissent introduire un certain biais dans le calibrage de notre modèle, nous considérons que ce biais n'est pas significatif dans la mesure où la répartition obtenue respecte le nombre total de défaillances réellement observées chaque mois.

La figure 2.7 ci-dessous détaille l'impact du choix de la répartition sur l'évolution du processus et de son intensité. On peut notamment remarquer que l'évolution de l'intensité est très sensible à la méthode de répartition sélectionnée.

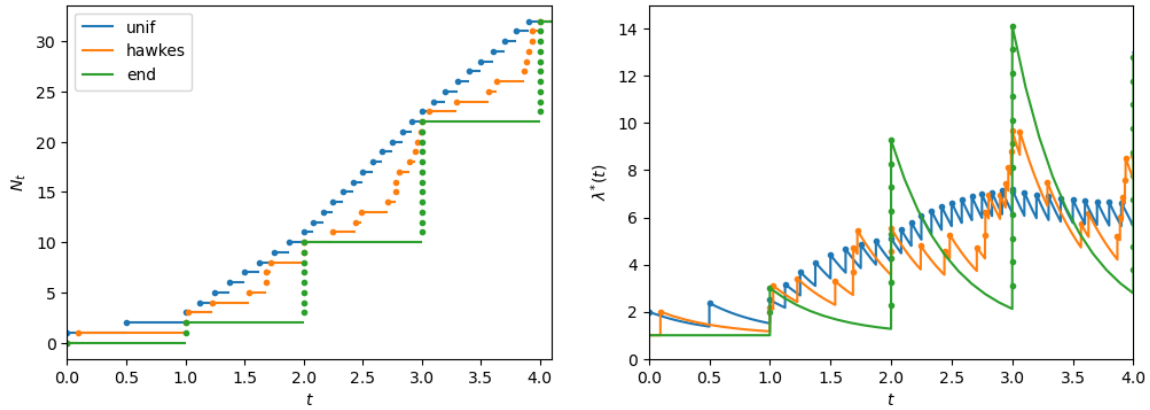


FIGURE 2.7 : Évolution d'un processus de Hawkes et de son intensité en fonction de la répartition des temps de sauts

Nous avons également envisagé d'exploiter les dates de jugement des procédures collectives disponibles dans la base de données BODACC. Cette démarche aurait permis d'obtenir une structure de répartition des temps de sauts plus fidèle à la réalité. Cependant, cette solution n'a pas été retenue en raison de l'absence d'informations uniformes sur les secteurs d'activités dans les données. En effet, la base de données recense 600 121 secteurs d'activités différents, une diversité due à l'orthographe et au vocabulaire utilisés par la personne ayant saisi l'information. L'idée d'implémenter un modèle de *Natural Language Processing* pour faire correspondre les secteurs renseignés aux 11 secteurs d'activités « standards » a été envisagée. Cependant, cette solution ne nous a pas semblé pertinente dans le sens où le développement d'un tel modèle est assez chronophage et qu'il s'avère inutile lorsqu'on travaille avec des données réelles d'assureur-crédit.

2.2.2 Modélisation de la dépendance entre secteurs d'activité

Dans cette étude, nous postulons que le risque de défaut peut être conceptualisé comme un risque de contagion. En effet, nous avons expliqué dans le premier chapitre de ce mémoire que les défauts ont tendance à se produire simultanément, en raison de l'exposition des entreprises à des facteurs de risque communs et de leur interdépendance commerciale. Nous émettons l'hypothèse que les entreprises d'un même secteur d'activité constituent un groupe homogène en termes d'exposition au risque de défaut.

D'une part, on peut avoir un effet de contagion intra-sectoriel, qui se caractérise par le fait qu'un défaut dans un secteur peut être lié à une évolution économique susceptible d'affecter toutes les entreprises de ce secteur. Ce phénomène peut être modélisé par la propriété d'auto-excitation d'un processus de Hawkes, chaque défaut survenu dans un secteur augmentant temporairement la probabilité qu'un autre défaut survienne dans ce secteur, reflétant l'environnement économique peu favorable dans ce secteur durant un certain temps. D'autre part, il existe des effets de contagion inter-sectoriels. L'idée est ici qu'un événement économique peut potentiellement impacter un secteur d'activité immédiatement, puis un autre après un certain délai. Cela peut se produire, par exemple, en raison d'un effet de contagion lié aux interactions commerciales entre les entreprises du premier secteur touché et celles du second secteur. Nous cherchons à capturer ces phénomènes à travers un modèle de Hawkes multivarié, avec une variable pour chaque secteur d'activité. En étudiant les liens entre différents secteurs d'activité, nous pourrions anticiper les défauts qui résultent de l'interdépendance entre les secteurs.

Formellement, nous considérons un processus de Hawkes multivarié $N = (N^{(1)}, \dots, N^{(d)})$ dont

l'intensité conditionnelle de chaque processus $N^{(k)}$ est définie par

$$\lambda_k^*(t) = \lambda_k + \sum_{j=1}^d \sum_{T_n^{(j)} < t} \alpha_{k,j} e^{-\beta_k(t-T_n^{(j)})}, \quad (2.18)$$

où les $(T_n^{(k)})_{n \geq 0}$ sont les temps de survenance des défauts du processus $(N_t^{(k)})_{t \geq 0}$. Pour ces moments, nous adoptons une échelle de temps où chaque unité correspond à un mois.

Chaque processus de comptage $(N_t^{(k)})_{t \geq 0}$ enregistre le nombre de défaillances dans un secteur d'activité, où $k = 1, \dots, d$ et d est le nombre total de secteurs d'activité.

Nous avons choisi de définir le risque de base du processus $N^{(k)}$ par un paramètre λ_k constant, qui donne une estimation du nombre minimal de défauts que l'on peut attendre chaque mois dans le secteur k . Ce paramètre représente le risque de défaut *structurel* du secteur k , indépendamment des événements passés et correspond à la valeur asymptotique vers laquelle le processus d'intensité conditionnelle converge en l'absence d'événements.

Le paramètre $\alpha_{k,j}$ représente ici l'effet d'un défaut survenant dans le secteur j sur le taux de défaut instantané du secteur k . C'est à travers les paramètres $(\alpha_{j,k})$ que la dépendance entre secteurs d'activité sera modélisée.

Le paramètre β_k détermine la durée de l'impact d'un défaut sur l'intensité du processus $N^{(k)}$.

Dans ce premier modèle, nous n'intégrons pas directement de facteurs représentant l'environnement économique. Notre objectif est d'étudier dans un premier temps l'aptitude des processus de Hawkes à modéliser le risque de contagion entre les défauts. Nous faisons l'hypothèse que l'impact de l'environnement économique sera implicitement intégré dans le modèle à travers l'évolution du nombre de défauts dans chaque secteur.

Méthode de calibrage

Pour calibrer les paramètres du modèle, la méthode du maximum de vraisemblance est équivalente à la résolution du problème de minimisation suivant

$$\min_{\lambda \in \mathbb{R}^d, \alpha \in \mathbb{R}^{d \times d}, \beta \in \mathbb{R}^d} -\ell(\lambda, \alpha, \beta), \quad (2.19)$$

où la fonction ℓ , définie en (2.13), est la log-vraisemblance du modèle (2.18).

La fonction de log-vraisemblance du modèle n'ayant pas une forme affine, des algorithmes d'optimisation non linéaire sont utilisés pour estimer les paramètres maximisant la vraisemblance. Une exploration de la littérature nous a orienté vers l'algorithme de NELDER et MEAD (1965) pour résoudre le problème (2.19). Il s'agit d'un algorithme d'optimisation non linéaire de type *gradient-free*, c'est-à-dire qu'il ne requiert pas l'évaluation ou l'approximation du gradient. Il commence par sélectionner plusieurs points candidats formant un simplexe dans l'espace des paramètres inconnus, puis remplace progressivement le moins bon candidat à chaque itération. Pour un problème de minimisation de la forme $\min_{\theta \in \mathbb{R}^n} f(\theta)$, l'algorithme peut être résumé comme suit :

1. On part d'un simplexe de l'espace des paramètres à estimer : $\{\theta_1^{(0)}, \dots, \theta_{n+1}^{(0)}\}$. Les valeurs initiales fournies à l'algorithme forment le premier sommet du simplexe. Les autres sommets sont générés en ajoutant une petite perturbation à chaque coordonnée du point de départ ;
2. Les points sont numérotés de façon à avoir : $f(\theta_1^{(0)}) \leq \dots \leq f(\theta_{n+1}^{(0)})$;

3. A chaque itération k , on effectue les opérations suivantes :

- Remplacer le point le moins performant $\theta_{n+1}^{(k)}$ par un meilleur point $\hat{\theta}$ tel que $f(\hat{\theta}) < f(\theta_{n+1}^{(k)})$;
- Mise à jour du simplexe :

$$\{\theta_1^{(k+1)}, \dots, \theta_n^{(k+1)}, \theta_{n+1}^{(k+1)}\} = \{\theta_1^{(k)}, \dots, \theta_n^{(k)}, \hat{\theta}\};$$

- Renommer les points du nouveau simplexe de sorte que

$$f(\theta_1^{(k+1)}) \leq \dots \leq f(\theta_{n+1}^{(k+1)});$$

4. L'algorithme se termine lorsque le critère d'arrêt est satisfait. L'optimum renvoyé par l'algorithme est $\theta_1^{(k+1)}$.

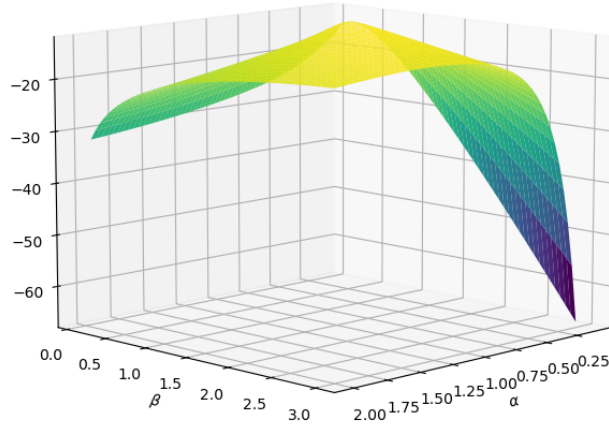
Le critère d'arrêt de l'algorithme de Nelder-Mead est atteint lorsque les points du simplexe sont suffisamment proches les uns des autres, selon un seuil de tolérance prédéfini, ou lorsque qu'un nombre maximum d'itérations est atteint.

Cet algorithme est applicable à toute fonction continue et est relativement simple à implémenter puisqu'il ne nécessite pas le calcul de dérivée. Cependant, sa sensibilité aux conditions initiales peut conduire à une convergence vers des minima locaux plutôt que vers le minimum global. Par conséquent, il est crucial de choisir judicieusement les valeurs initiales pour maximiser les chances de convergence vers l'optimum global. Ce choix a également une incidence sur la vitesse de convergence de l'algorithme. Par ailleurs, la performance de l'algorithme tend à diminuer à mesure que le nombre de paramètres augmente, limitant son utilisation en grande dimension. En effet, plus le nombre de paramètres est élevé, plus l'espace de recherche devient complexe, ce qui augmente le risque de convergence vers des optima locaux, et rend la recherche de l'optimum global plus difficile et coûteuse en temps de calcul. Enfin, lorsque l'évaluation de la fonction à minimiser demande un temps de calcul significatif, l'algorithme peut devenir particulièrement lent. En effet, à chaque itération, la fonction est évaluée pour chaque point constituant le simplexe, ce qui peut considérablement allonger le processus d'optimisation et rendre le calibrage des paramètres très long.

Optima locaux La fonction de log-vraisemblance (2.13) à maximiser n'est pas une fonction globalement concave et peut présenter une vallée plate, similaire à la fonction de Rosenbrock. Cette vallée, visible en jaune sur la figure 2.8, est facilement atteignable par un algorithme d'optimisation. Cependant, une fois dans cette vallée, la structure de la fonction rend la progression vers le maximum global difficile par une méthode numérique. La forme représentée sur la figure 2.8 correspond à la log-vraisemblance d'un processus de Hawkes univarié, calculée à partir de données simulées. Or, dans le cas de données réelles, en présence de bruit et de possibles valeurs aberrantes, et compte tenu du fait que le processus de Hawkes ne reflète pas parfaitement la survenance des événements modélisés, la log-vraisemblance est généralement plus complexe que celle illustrée. Il est alors courant de rencontrer plusieurs optima locaux.

Pour maximiser les chances de convergence vers l'optimum global, une stratégie fréquemment adoptée consiste à tester différentes valeurs pour les paramètres initiaux. L'utilisation de plusieurs méthodes d'optimisation pour estimer les paramètres est également une pratique courante. Il convient de noter que ces solutions permettent de limiter les risques de convergence vers un optimum local, mais ne permettent pas de garantir l'obtention du maximum global.

Pour ce faire, nous réalisons un pré-calibrage en ajustant d'abord un processus de Hawkes univarié pour chaque secteur d'activité. En variant les valeurs initiales, nous augmentons les chances d'atteindre un minimum global. Le calibrage d'un processus de Hawkes univarié est nettement moins coûteux que

FIGURE 2.8 : Log-vraisemblance d'un processus de Hawkes univarié de la forme (2.2) - λ fixé

celui d'un processus multivarié en termes de temps de calcul. En effet, le nombre de paramètres à estimer est réduit à trois par processus à calibrer, et l'historique est réduit puisqu'il se restreint aux défaillances observées sur le secteur modélisé. L'une des raisons motivant le choix de commencer par un calibrage en univarié est que les paramètres du modèle sont plus facilement interprétables. Cela permet une meilleure compréhension du comportement du modèle et un ajustement plus précis des valeurs initiales pour le calibrage du processus multivarié.

Pour calibrer, l'algorithme de Nelder-Mead est utilisé, et mis en concurrence avec l'algorithme L-BFGS-B afin de limiter au maximum le risque de minima locaux. La procédure détaillée de l'algorithme L-BFGS-B est décrite dans BYRD et al. (1995) et ZHU et al. (1997). Cet algorithme est une extension de l'algorithme Broyden-Fletcher-Goldfarb-Shanno (BFGS), présenté en annexe A.3, avec limite de mémoire (approximation et stockage d'une matrice Hessienne de taille réduite) et prise en compte de bornes sur les paramètres.

Ce pré-calibrage permet de tester un grand nombre de valeurs initiales, y compris différentes plages temporelles, afin d'obtenir une estimation robuste des paramètres du modèle univarié. Ces estimations servent ensuite de point de départ pour le calibrage du processus multivarié afin d'éviter une convergence trop longue vers l'optimum ou de se retrouver coincé dans un optimum local. Elles nous permettent ainsi de déterminer les valeurs initiales des paramètres $(\lambda_k)_{k=1,\dots,d}$, $(\beta_k)_{k=1,\dots,d}$ et $(\alpha_{k,k})_{k=1,\dots,d}$. Une étude des interactions entre les secteurs d'activité permettra d'estimer les valeurs initiales des paramètres $(\alpha_{j,k})_{j \neq k}$ restants.

Vitesse de convergence Un des défis majeurs de l'estimation par maximum de vraisemblance d'un processus de Hawkes réside dans le coût potentiellement élevé de l'évaluation de la vraisemblance, et en particulier de l'intensité conditionnelle. La complexité algorithmique provient de la double somme dans la première composante de la log-vraisemblance

$$\sum_{i=1}^n \log(\lambda_{d_i}^*(t_i)) = \sum_{i=1}^n \log \left(\lambda_{d_i} + \sum_{(t_k, d_k): t_k < t_i} \alpha_{d_i, d_k} e^{-\beta_{d_i}(t_i - t_k)} \right).$$

Cependant, grâce aux propriétés du noyau exponentiel, nous avons pu réécrire l'intensité conditionnelle de façon récursive (cf. équation (2.15)) afin de limiter les calculs en boucle. Cela a permis de passer d'un coût d'ordre $\mathcal{O}(n^2)$ à $\mathcal{O}(n)$ pour l'évaluation de la log-vraisemblance.

La fonction de vraisemblance du modèle (2.18) étant exprimée en fonction des temps de saut du processus, son coût d'évaluation dépend directement du nombre n de défaillances observées sur la période considérée. Plus l'historique utilisé pour le calibrage contient de défauts, plus le processus de calibrage sera long. Notre base de données compte plus d'un million de défaillances d'entreprises, ce qui rend le calibrage sur l'ensemble de cet historique particulièrement coûteux en termes de temps de calcul. De plus, il n'est pas toujours pertinent d'utiliser toute la période disponible, surtout lorsque le risque a évolué au fil du temps. Pour optimiser le calibrage, il a été choisi de restreindre l'horizon temporel aux défaillances survenues entre 2010 et 2019. La période 2020-2023 a été volontairement exclue, car nous considérons qu'il s'agit d'une période exceptionnelle qui n'est pas représentative du risque structurel que nous cherchons à modéliser. Par ailleurs, les évolutions technologiques rapides depuis les années 1990 rendent peu pertinente l'hypothèse que les tendances des décennies antérieures reflètent celles des défaillances actuelles. En choisissant de débuter en 2010, l'historique est suffisamment riche pour capturer au moins un « pic » de défaillances dans chaque secteur d'activité, tout en garantissant un temps de calibrage raisonnable. Le choix de la période 2010-2019 s'aligne également avec la plage temporelle utilisée comme référence par la Banque de France pour ses analyses mensuelles sur l'évolution des défaillances d'entreprises.

Un autre facteur pouvant ralentir la convergence de l'algorithme est le nombre de paramètres à calibrer. Dans notre modélisation, il est nécessaire d'estimer un total de $(d + 2) \times d$ paramètres. Avec $d = 11$ secteurs d'activités, cela représente 143 paramètres à calibrer. Pour réduire ce nombre et accélérer le temps de calibrage, nous avons opté pour un regroupement de certains secteurs d'activité. Nous avons suivi la segmentation utilisée par l'INSEE pour certains de ses indicateurs. Ainsi, les secteurs suivants : *Commerce et réparation automobile*, *Transports et entreposage*, *Hébergement et restauration*, *Information et communication*, *Activités financières et d'assurance*, *Activités immobilières* et *Soutien aux entreprises* sont regroupés sous la catégorie *Services marchands*. Cette agrégation nous fait passer de 11 à 5 secteurs d'activités, ce qui nous permet de diminuer considérablement le nombre de paramètres à estimer, passant de 143 à 35.

Enfin, bien définir le point de départ de l'algorithme est essentiel, non seulement pour assurer la convergence vers un optimum global comme expliqué précédemment, mais aussi pour limiter le nombre d'itérations de l'algorithme. En effet, plus l'algorithme démarre proche de la solution, plus la convergence sera rapide. Le temps de calibrage d'un processus univarié peut atteindre jusqu'à 3 minutes lorsque le point de départ est éloigné de la solution. Il est crucial de s'en approcher le plus possible en multivarié où l'espace de recherche est de 36 dimensions contre 4 dimensions en univarié et où l'historique, qui comprend les défaillances sur l'ensemble des secteurs, est plus conséquent.

Effet de bord La log-vraisemblance du modèle est définie sur un intervalle de temps $[0, T]$ qui correspond à l'historique utilisé pour calibrer les paramètres sur les données réelles. La date $t = 0$ marque le début du processus, en supposant qu'aucun événement n'est survenu avant cette date. Cependant, cette hypothèse paraît irréaliste pour modéliser le nombre de défauts. En effet, l'hypothèse que les défaillances d'entreprises ont commencé à exister en 2010 semble peu plausible. Des défauts non observés, survenus avant la date $t = 0$, pourraient avoir influencé la période d'observation. Comme ces défauts ne sont pas inclus dans nos données, leur contribution à l'intensité conditionnelle du processus n'est pas enregistrée. A l'instar de RASMUSSEN (2013), nous supposons que pour un jeu de données assez conséquent, ce qui est le cas avec l'historique des défaillances entre 2010 et 2019, l'effet de bord est négligeable sur l'estimation des paramètres.

En revanche, l'effet de bord sera pris en compte pour la simulation du processus. Pour cela, l'initialisation de l'algorithme 2 a été ajustée afin de simuler le processus de Hawkes comme le prolongement de l'historique réellement observé. Contrairement à une simulation standard qui démarre à la date $t = 0$, les simulations débutent à une date $t > 0$, correspondant au temps écoulé entre le début de l'historique et le début de la période de simulation. Ainsi, l'historique des temps de saut à cette première date de simulation n'est pas vide : il contient l'ensemble des défauts observés entre $t = 0$ et la date de départ de la simulation. De cette manière, l'intensité conditionnelle du processus à cette première date de simulation prend en compte l'information disponible pour prédire les défauts à venir.

Résultats du calibrage

Avant de procéder au pré-calibrage, les trois méthodes de répartition des temps de sauts décrites dans la section précédente sont évaluées pour chaque processus univarié. Deux périodes de calibrage relativement courtes ont été choisies pour assurer une convergence rapide de l'algorithme. La première période, 2008-2009, coïncide avec un pic de défaillances pour l'ensemble des secteurs, tandis que la seconde, 2011-2012, correspond à une période de stabilité relative des défaillances pour la plupart des secteurs. Cette démarche a pour objectif d'examiner l'impact de chaque méthode de répartition des temps de sauts sur l'estimation des paramètres du modèle et ainsi déterminer celle qui est la plus adaptée.

La première approche, consistant à supposer que les défauts surviennent tous en fin de mois, a été écartée assez rapidement. En effet, les paramètres estimés s'avèrent très sensibles au choix du pas de temps ϵ entre chaque saut. Le tableau 2.1 illustre cette sensibilité en présentant les résultats obtenus, pour un secteur d'activité donné, en ne faisant varier que le paramètre ϵ . Pour ces tests, les paramètres initiaux ont été fixés arbitrairement à $\alpha_0 = 100$ et $\beta_0 = 102$, et λ_0 correspond à la moyenne empirique observée sur la période considérée. On observe que lorsque l'écart entre les temps de sauts devient très petit, les valeurs des paramètres α et β explosent.

Période	ϵ	λ	α	β
2008-2009	10^{-3}	1.15	166.21	168.00
	10^{-6}	10^{-5}	166 555	32 365
	10^{-9}	10^{-5}	166 555 129	168 103 878
	0	108.54	3.58×10^{25}	6.58×10^{25}
2011-2012	10^{-3}	1.15	170.84	172.77
	10^{-6}	10^{-5}	171 187	172 875
	10^{-9}	10^{-5}	171 187 725	172 875 009
	0	102.46	3.38×10^{25}	6.21×10^{25}

TABLE 2.1 : Résultat des calibrages sur les deux périodes en supposant une répartition des temps de sauts en fin de mois

Cette approche pourrait convenir pour modéliser un risque rare, où l'intensité explose ponctuellement, ce qui rend le moment précis de survenance des événements (début, milieu ou fin de mois) moins important. Dans notre cas, elle n'est pas appropriée puisque nous cherchons à modéliser non seulement les phénomènes d'accumulation de défauts à travers les propriétés des processus de Hawkes, mais

également un flux régulier de défaillances chaque mois. En supposant que l'ensemble des défaillances surviennent en fin de mois, le modèle tend à sous-estimer le risque sur le reste de la période, le considérant comme nul. Il semble plus approprié de répartir la survenance des événements sur l'ensemble du mois plutôt que de les concentrer à un seul instant.

Les deux méthodes restantes ont été évaluées en testant plusieurs paramètres initiaux afin d'examiner la stabilité des résultats. Les résultats obtenus sont présentés en annexe A.4.1. L'approche retenue est de répartir uniformément les défauts au sein de chaque mois, car c'est la seule méthode qui assure des résultats suffisamment stables, quelles que soient les valeurs initiales testées. Bien qu'une répartition uniforme puisse ne pas sembler idéale pour modéliser un risque « évolutif » dont le taux de survenance augmente ponctuellement à chaque fois qu'un événement se produit, le fait de disposer du nombre réel de défauts mensuels permet néanmoins d'identifier les phénomènes d'excitation à un niveau de granularité mensuel.

Au-delà de nous permettre d'identifier une méthode de répartition des temps de sauts, ces premiers résultats confirment une première intuition que nous avons sur notre modèle : sur une période où l'évolution du nombre de défaillances est assez stable, les valeurs estimées pour les paramètres α ont tendance à être proches de zéro. Autrement dit, en période de stabilité économique, un modèle de Poisson suffit pour modéliser les défaillances.

Après avoir identifié la méthode de répartition des défauts la plus pertinente, nous avons procédé au pré-calibrage. Pour rappel, cette étape consiste à calibrer, sur la période 2010-2019, chaque processus univarié représentant un secteur d'activité. En testant différentes valeurs initiales et différentes méthodes d'optimisation, différents résultats peuvent être obtenus en raison de la non-convexité de la fonction minimisée. La valeur de la log-vraisemblance pour chaque résultat est alors comparée, et ceux pour lesquels elle est maximale sont retenus. Le tableau 2.2 ci-dessous présente les résultats conservés pour la suite.

Secteur d'activité	λ	α	β
Agriculture, sylviculture et pêche (A)	101.67	0.0024	0.0074
Industrie (I)	143.52	1.2085	1.9920
Construction (C)	223.70	3.2765	4.020
Enseignement, santé humaine, action sociale et services aux ménages (E)	381.52	0.0127	0.0701
Services marchands (S)	1018.24	3.0842	4.8764

TABLE 2.2 : Résultat du pré-calibrage sur chaque secteur

Il est à noter que les secteurs *Agriculture, sylviculture et pêche* et *Enseignement, santé humaine, action sociale et services aux ménages*, précédemment identifiés comme résilients à la conjoncture économique, ont les paramètres d'excitation les plus faibles.

Nous avons également effectué des calibrages sur des processus de Hawkes bivariés représentant des couples de secteurs d'activité. Le passage en deux dimensions a considérablement ralenti le temps d'exécution de l'algorithme, passant de moins d'une minute à environ trente minutes. De ce fait, nous avons restreint notre analyse aux couples de secteurs présentant des corrélations élevées. L'ensemble des résultats du pré-calibrage est disponible en annexe A.4.2.

Enfin, pour le calibrage des paramètres du modèle (2.18), nous définissons les valeurs initiales des

paramètres $(\lambda_k)_{k=1,\dots,d}$ et $(\beta_k)_{k=1,\dots,d}$ de la façon suivante, pour $k = 1, \dots, d$:

- le paramètre λ_k prend la valeur estimée par le pré-calibrage. Si plusieurs estimations sont disponibles, par exemple lorsqu'un calibrage univarié puis bivarié a été effectué pour le même secteur, nous retenons la valeur la plus faible. L'intuition est que ce paramètre doit diminuer lorsque des paramètres d'excitation supplémentaires sont introduits ;
- le paramètre β_k prend également la valeur issue par le pré-calibrage. En cas de choix multiple, nous conservons la valeur la plus élevée. Cela garantit que la décroissance de l'excitation est suffisamment rapide pour éviter la divergence du processus.

Les paramètres $(\alpha_{k,j})_{j=1,\dots,d}$ proviennent également du pré-calibrage lorsqu'ils ont été estimés. Pour les couples de secteurs d'activité pour lesquels aucun calibrage bivarié n'a été effectué, certaines valeurs sont manquantes. Dans ces cas, nous estimons les paramètres manquants à l'aide de la relation

$$\hat{\alpha}_{k,j} = \hat{\alpha}_k \times \rho_{k,j}, \quad (2.20)$$

où $\rho_{k,j}$ est un coefficient qui représente la dépendance entre les secteurs k et j et $\hat{\alpha}_k$ est estimé par le pré-calibrage.

Les paramètres $(\alpha_{k,j})_{1 \leq k,j \leq d, k \neq j}$ incarnent la structure de dépendance entre les secteurs d'activité. Afin de partir du meilleur point de départ possible pour ces paramètres, nous avons mené une étude sur la dépendance entre les secteurs d'activité. Pour estimer cette dépendance, nous avons utilisé deux approches : une méthode statistique basée sur les données de défaillances et une méthode économique s'appuyant sur les échanges commerciaux entre les secteurs d'activité.

Approche statistique Cette méthode repose sur l'utilisation de la corrélation de Pearson pour définir le paramètre $\rho_{k,j}$. Plus précisément, $\rho_{k,j}$ est déterminé comme la valeur absolue de la corrélation de Pearson entre les défaillances des secteurs k et j . La matrice ρ_1 ci-dessous présente les coefficients calculés.

$$\rho_1 = \begin{pmatrix} 1 & 0.802067 & 0.027828 & 0.205061 & 0.468785 \\ 0.802067 & 1 & 0.179752 & 0.245635 & 0.581652 \\ 0.027828 & 0.179752 & 1 & 0.718789 & 0.772031 \\ 0.205061 & 0.245635 & 0.718789 & 1 & 0.576607 \\ 0.468785 & 0.581652 & 0.772031 & 0.576607 & 1 \end{pmatrix}$$

Remarque 2.7. Les valeurs obtenues pour la corrélation de Pearson sont presque identiques à celles obtenues avec la corrélation de Spearman. Cela suggère que la relation entre les défaillances des différents secteurs d'activité est bien capturée par une corrélation linéaire et que l'utilisation de la corrélation de Pearson est adaptée.

Approche économique Pour analyser les relations entre branches d'activité, un outil couramment utilisé en économie est le tableau des entrées-sorties (TES). Le tableau 2.3 représente le TES de l'année 2021, dans lequel les branches d'activité ont été regroupées pour correspondre aux cinq secteurs d'activités du modèle. Dans ce tableau, les colonnes correspondent aux branches et les lignes aux produits. Ainsi, si nous considérons la colonne *Industrie*, nous avons la répartition de la consommation de la branche *Industrie* dans les autres branches d'activité. Par exemple, 45.28 est le montant de biens ou services consommés par la branche *Industrie* dans le secteur *Agriculture, sylviculture et*

Produits / Branches	(A)	(I)	(C)	(E)	(S)
(A)	18.88	45.28	0.33	0.27	2.25
(I)	31.03	531.59	101.24	85.11	208.47
(C)	0.45	4.02	61.85	9.79	12.85
(E)	0.25	8.83	1.72	18.33	25.20
(S)	5.81	121.42	48.27	101.40	823.06

TABLE 2.3 : Tableau des entrées intermédiaires 2021 (en milliards d'euros) - INSEE (2022)

pêche. Le TES permet d'analyser les interdépendances entre les différents secteurs de l'économie et de comprendre comment les changements dans un secteur peuvent affecter les autres.

Pour construire une matrice d'interactions entre les secteurs d'activités, le tableau 2.3 est normalisé en colonne. L'idée sous-jacente est que si une entreprise d'un certain secteur fait défaut, cela affectera les entreprises qui lui ont vendu des biens et services, car elles ne seront pas remboursées. Par conséquent, cela aura un impact sur les secteurs « consommés » par le secteur de l'entreprise défaillante. Après la normalisation du tableau, nous réajustons les coefficients de manière à ce que la diagonale ne contienne que des 1. Cette condition est imposée pour garder une cohérence entre la formule (2.20) et l'initialisation des paramètres $\alpha_{k,k}$.

Si nous notons $X = (x_{i,j})_{1 \leq i,j \leq d}$, la matrice contenant les valeurs du tableau 2.3, les étapes décrites pour construire la matrice d'interactions reviennent à multiplier X par une matrice diagonale D de taille $d \times d$, définie par

$$D = \text{diag} \left(\frac{1}{x_{1,1}}, \dots, \frac{1}{x_{d,d}} \right).$$

La matrice ρ_2 ci-dessous présente les coefficients calculés avec l'approche économique.

$$\rho_2 = \begin{pmatrix} 1 & 0,085171 & 0,005409 & 0,014968 & 0,002733 \\ 1,643450 & 1 & 1,636913 & 4,644134 & 0,253288 \\ 0,023722 & 0,007555 & 1 & 0,534014 & 0,015613 \\ 0,013114 & 0,016615 & 0,027850 & 1 & 0,030615 \\ 0,307883 & 0,228415 & 0,780475 & 5,533250 & 1 \end{pmatrix}$$

On peut noter qu'ici, contrairement à l'approche statistique, il est possible d'avoir $\rho_{k,j} > 1$. Cela signifie que l'inter-excitation entre le processus $N^{(k)}$ et le processus $N^{(j)}$ peut surpasser l'auto-excitation du processus $N^{(k)}$. Par exemple, pour le secteur de l'agriculture, on remarque que le coefficient $\rho_2(I, A)$ vaut environ 1.64. Cela signifie qu'un défaut dans le secteur de l'agriculture a un impact 1.64 fois plus fort sur le secteur de l'industrie que sur lui-même. Cette observation semble réaliste : il est plausible que des défaillances au sein du secteur agricole aient un effet plus marqué sur les distributeurs de leurs produits, plutôt que d'être simplement le reflet d'un environnement économique défavorable affectant les autres entreprises du secteur, surtout sachant que l'agriculture est généralement moins affectée par les fluctuations conjoncturelles.

D'autre part, la matrice ρ_2 , à la différence de la matrice ρ_1 , n'est pas symétrique. Cela signifie que la dépendance entre un secteur k et un secteur j varie selon la direction. Cette asymétrie nous semble pertinente, car les échanges commerciaux entre deux secteurs ne sont pas équilibrés. Par exemple, le tableau 2.3 montre que le secteur de la construction est un grand consommateur des produits de

l'industrie, ce qui paraît logique puisque les entreprises doivent se fournir en matériaux et équipements. En revanche, les produits et services issus du secteur de la construction ne sont pas aussi largement utilisés dans les processus industriels. Cela explique la dynamique unidirectionnelle observée à travers les paramètres $\rho_2(I, C)$ et $\rho_2(C, I)$.

Nous avons initié deux calibrages, partant des paramètres initiaux dérivés des deux approches présentées. Chaque calibrage a nécessité un temps d'exécution relativement long, environ 15 heures, ne nous permettant pas de tester un large échantillon de paramètres initiaux comme cela a été fait en univarié. Les paramètres finalement retenus sont ceux issus de l'approche statistique, car ils conduisent à une vraisemblance plus élevée qu'avec l'approche économique. Cela peut s'expliquer par le fait qu'étant basée sur les corrélations entre les défaillances des secteurs, l'approche statistique capture plus directement la dépendance entre les défaillances des secteurs que l'approche économique. Bien qu'elle donne une bonne indication des relations commerciales entre les secteurs, cette dernière peut ne pas refléter aussi précisément la manière dont une défaillance se propage à travers ces liens.

Les paramètres conservés pour la suite des travaux, arrondi à 10^{-4} pour une meilleure lisibilité, sont les suivants

$$\lambda = \begin{pmatrix} 111.17 \\ 135.29 \\ 195.09 \\ 283.21 \\ 1274.82 \end{pmatrix}, \quad \alpha^T = \begin{pmatrix} 0.0000 & 0.0000 & 0.0758 & 0.1021 & 0.0671 \\ 0.0000 & 0.6499 & 0.6587 & 0.0732 & 0.4600 \\ 0.0000 & 0.2998 & 2.8806 & 0.0225 & 2.5195 \\ 0.0000 & 0.1847 & 0.0000 & 0.3701 & 0.5561 \\ 0.0000 & 0.0000 & 0.0000 & 0.0140 & 1.2355 \end{pmatrix} \quad \text{et} \quad \beta = \begin{pmatrix} 0.00 \\ 2.97 \\ 3.68 \\ 1.65 \\ 4.61 \end{pmatrix}. \quad (2.21)$$

Les paramètres obtenus révèlent plusieurs éléments intéressants concernant les interdépendances entre secteurs capturées par le modèle. Tout d'abord, le secteur de l'agriculture (première colonne de la matrice α^T) n'est pas ou très peu influencé par les défaillances des autres secteurs, ni par ses propres défaillances, ce qui indique l'absence d'effet d'auto-excitation ou d'inter-excitation sur ce secteur. Toutefois, la survenance de défauts dans le secteur de l'agriculture semble tout de même impacter les autres secteurs, comme le montre sa contribution aux processus de certains secteurs, tels que les services marchands et la construction.

Il est surprenant de constater que le secteur industriel n'est pas affecté par les défaillances de l'agriculture, malgré des corrélations initiales élevées et des échanges commerciaux importants entre ces deux secteurs. Cela suggère que les corrélations observées sont le résultat d'une dépendance commune à un autre facteur et que, malgré ces liens économiques, les processus de défaillances ne se propagent pas de l'agriculture vers l'industrie.

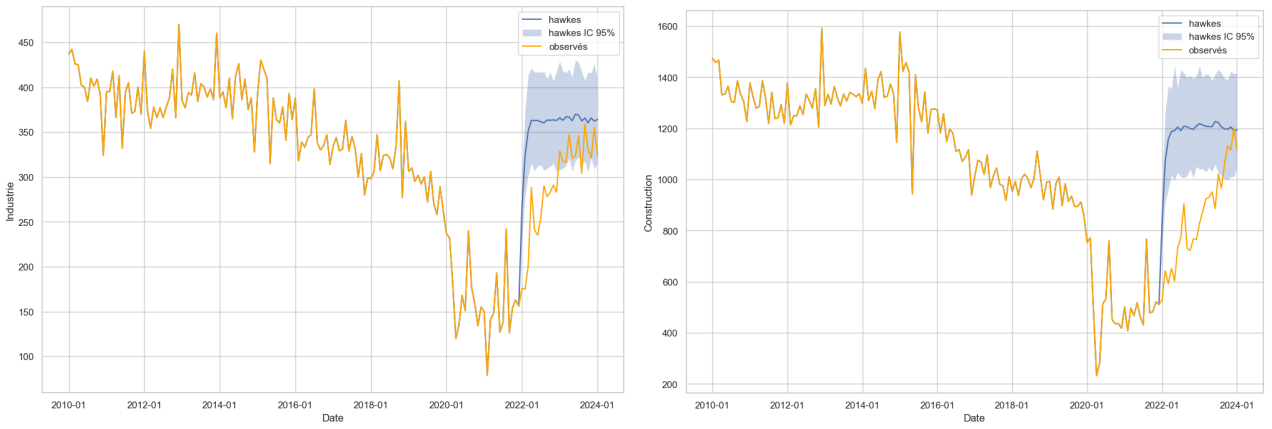
Pour les secteurs de la construction et des services marchands, qui présentent également une corrélation élevée, il apparaît qu'il existe une dépendance à sens unique. En effet, une défaillance dans les services marchands augmente immédiatement l'intensité conditionnelle des défaillances dans le secteur de la construction de 2.5195. En revanche, une défaillance dans la construction n'a pas d'impact sur l'intensité conditionnelle des services marchands, ce qui révèle une relation asymétrique entre ces deux secteurs.

Enfin, le secteur de la construction se distingue par une forte auto-excitation et inter-excitation, ce qui est cohérent avec son historique marqué par un grand nombre de défaillances et des fluctuations importantes. Cette dynamique reflète la nature volatile de ce secteur.

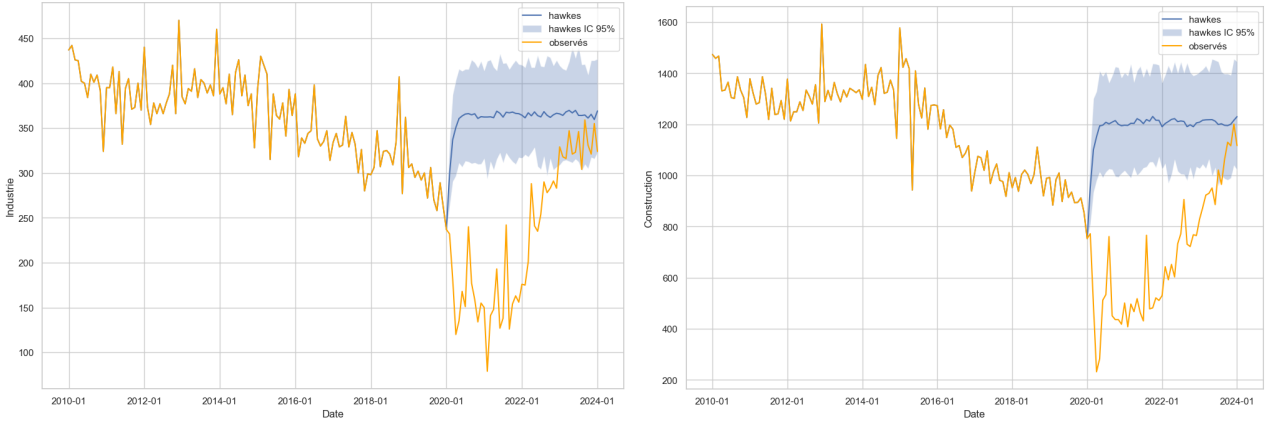
Premiers résultats

Une fois les paramètres calibrés, plusieurs simulations du processus sont effectuées sur diverses périodes passées pour évaluer la précision des prédictions. Pour tenir compte de l'influence du passé du processus à la première date de simulation, ces simulations sont générées comme le prolongement de l'historique des données. La moyenne empirique des trajectoires simulées est ensuite retenue comme prédiction. La figure 2.9 illustre les résultats obtenus pour deux périodes de prédictions sur deux secteurs d'activité.

Plus précisément, sur la figure 2.9b par exemple, nous avons effectué des simulations à partir de février 2020, en prenant en compte toutes les défaillances réellement survenues depuis 2010. Ensuite, à chaque date entre février 2020 et décembre 2023, le modèle intègre l'information passée, réelle puis simulée, jusqu'à la date de fin de prédiction, qui est décembre 2023.



(a) de janvier 2022 à décembre 2023



(b) de février 2020 à décembre 2023

FIGURE 2.9 : Prédications du modèle (2.18) avec les paramètres (2.21)

Les résultats obtenus avec ce premier modèle ne sont pas satisfaisants. En effet, le modèle semble converger vers un niveau constant de défaillances mensuelles, indépendamment du point de départ des prédictions. Cette tendance est observée malgré le fait que l'historique du processus simulé ait été ajusté pour contenir l'historique de défaillances réellement observées avant la date de début de prédiction. Ce comportement semble être attribuable à l'utilisation de paramètres fixes λ_k pour représenter le risque de base dans chaque secteur d'activité. L'hypothèse d'une intensité de base

constante ne correspond pas à la dynamique observée dans nos données, limitant ainsi la capacité du modèle à prédire en dessous des niveaux fixés par ces paramètres, un problème particulièrement notable sur les prédictions de la période post-COVID.

Néanmoins, les résultats des calibrages, tant en univarié qu'en multivarié, indiquent des paramètres $\alpha_{k,j}$ non nuls, soutenant l'hypothèse d'existence d'effets de contagion intra et intersectoriels. Cela nous laisse penser qu'en ajustant les paramètres d'intensité de base, le modèle pourrait fournir de meilleurs résultats.

En outre, une comparaison des prédictions du modèle avec celles obtenues à partir d'un processus de Poisson révèle que le modèle semble prendre en compte le passé récent dans une certaine mesure. Le processus de Hawkes semble intégrer l'information passée à très court terme. Cette comparaison est illustrée sur la figure 2.10.

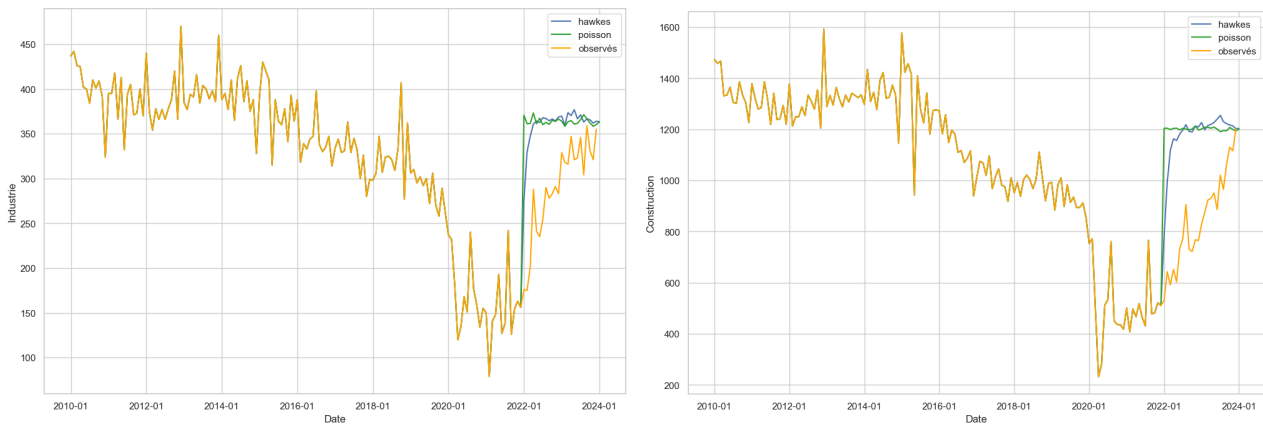


FIGURE 2.10 : Comparaison Hawkes et Poisson

2.2.3 Intégration de facteurs économiques

Les résultats du modèle précédent ont mis en évidence la nécessité de prendre en compte des facteurs externes pour modéliser le nombre de défaillances par secteur d'activité. L'hypothèse selon laquelle l'historique des défaillances intègre implicitement les fluctuations économiques n'est pas vérifiée, ou du moins les résultats obtenus sur les données à la maille mensuelle ne permettent pas de l'affirmer. Nous faisons désormais l'hypothèse que l'intensité de base λ du modèle doit être représentée par une fonction $\mu(\cdot)$ évoluant dans le temps en réponse à l'environnement économique. En effet, supposer que le risque de base est constant dans le temps et attribuer les tendances observées à la hausse ou à la baisse uniquement à des phénomènes de contagion ne semble pas réaliste.

Le cadre du modèle (2.22) est conservé, seul le paramètre λ_k est remplacé par une fonction $\mu_k(\cdot)$ qui dépend d'indicateurs économiques. Dans le premier modèle, l'objectif était de capturer les effets de l'environnement économique sur le risque de défaillance via les interactions entre secteurs, supposant que l'impact de l'économie se refléterait à travers celles-ci. L'environnement économique est désormais intégré directement dans le modèle, avec l'hypothèse qu'il dicte la tendance générale du risque de défaut sur chaque secteur. La composante associée à l'excitation du processus permettra de modéliser les cas d'accumulation de défauts dans un secteur spécifique et les effets de contagion entre secteurs.

Formellement, pour un processus de Hawkes multivarié $N = (N^{(1)}, \dots, N^{(d)})$, le modèle s'écrit

comme suit

$$\lambda_k^*(t) = \mu_k(t) + \sum_{j=1}^d \sum_{T_n^{(j)} < t} \alpha_{k,j} e^{-\beta_k(t-T_n^{(j)})}, \quad \forall k = 1, \dots, d, \quad (2.22)$$

où $\lambda_k^*(t)$ représente l'intensité conditionnelle du processus $(N_t^{(k)})_{t \geq 0}$ et les $(T_n^{(k)})_{n \geq 0}$ sont les temps de survenance des défauts du processus $(N_t^{(k)})_{t \geq 0}$.

Le paramètre $\alpha_{k,j}$ représente toujours l'effet d'un défaut survenant dans le secteur j sur le taux de défaut instantané du secteur k et β_k la durée de l'impact d'un défaut sur l'intensité du processus $N^{(k)}$.

La fonction $\mu_k(\cdot)$ représente l'intensité de base du processus $(N_t^{(k)})_{t \geq 0}$. Elle est construite à partir d'indicateurs macroéconomiques et sa valeur fluctue en fonction de l'évolution de ces indicateurs. Une faible valeur de cette fonction correspond à des conditions économiques favorables au développement des entreprises. Dans ce cas, l'intensité conditionnelle, et donc le nombre de défauts observé est réduit. Par conséquent, la probabilité de déclencher des phénomènes d'accumulation de défauts via la propriété d'excitation du processus est réduite. En revanche, une valeur élevée de cette fonction, qui indique des conditions économiques moins favorables, augmente le taux de survenance des défauts et peut conduire à des « pics » de défauts en raison de l'excitation du processus.

La figure 2.11 illustre l'évolution de l'intensité conditionnelle d'un processus de Hawkes dont l'intensité conditionnelle est définie comme en (2.22). La fonction de base $\mu_k(\cdot)$ est constante par morceaux, avec des « sauts » à chaque unité de temps, pouvant être à la hausse ou à la baisse. Notamment, au temps $t = 124$, un saut significatif de μ vers le haut entraîne une accélération immédiate des défaillances. Chaque point sur la figure représente un événement de défaillance, et l'augmentation de leur fréquence est clairement visible après ce saut de l'intensité. Par la suite, plusieurs « sauts » successifs de μ_k à la baisse entraînent une diminution progressive de l'intensité conditionnelle, ce qui se traduit par une réduction de la fréquence des défaillances.

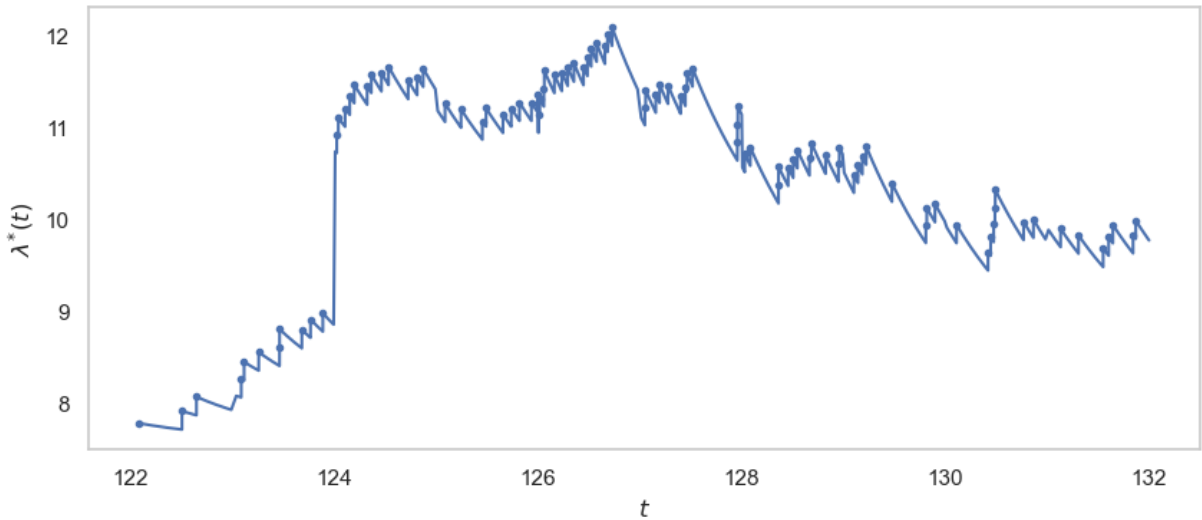


FIGURE 2.11 : Évolution intensité conditionnelle de la forme (2.22)

Indicateurs macroéconomiques

Une sélection d'indicateurs macroéconomiques, jugés pertinents pour refléter le contexte économique de chaque branche d'activité, est présentée ci-après. Cette sélection s'appuie sur les analyses des risques

sectoriels publiées par [Coface](#) et [Allianz Trade](#). Ces études sont réalisées chaque trimestre par des experts et à partir de données internes de ces assureurs sur les entreprises. Leur objectif est d'évaluer le risque de non-paiement des entreprises dans chaque secteur d'activité.

Pour l'ensemble des secteurs d'activité, nous avons retenu un certain nombre d'indicateurs permettant d'appréhender l'environnement économique général, lesquels impactent l'ensemble des secteurs d'activité. Ces indicateurs comprennent :

- le taux de croissance trimestriel du Produit Intérieur Brut (PIB) français, qui reflète l'évolution de l'état global de l'économie française ;
- les taux d'intérêts, déterminés par la Banque Centrale Européenne (BCE), qui représentent le coût de l'emprunt et les rendements des investissements. Ils ont une influence directe sur les décisions de consommation et d'investissement des entreprises ;
- l'indice des prix à la consommation (inflation), qui mesure la variation des prix des biens et services. L'inflation affecte les marges des entreprises et donc leur rentabilité ;
- le *Global Supply Chain Pressure Index (GSCPI)*, qui mesure les pressions exercées sur la chaîne d'approvisionnement mondiale. Il prend en compte des facteurs tels que les coûts de transport, les délais de livraison, les risques politiques et les fluctuations des prix des matières premières.

Pour certains secteurs spécifiques, nous avons sélectionné des indicateurs supplémentaires permettant d'affiner l'état de l'environnement économique associé à chacun de ces secteurs. Par exemple, pour le secteur de l'agriculture, nous examinons également l'évolution des prix agricoles et des matières premières grâce à l'indice FAO (Organisation des Nations Unies pour l'alimentation et l'agriculture). Pour les secteurs de la construction et de l'industrie, l'inflation de l'énergie fait l'objet d'une attention particulière. Enfin, nous intégrons également pour chaque secteur d'activité l'indice du climat des affaires fourni par l'INSEE, ce qui nous permet d'obtenir une vision plus complète de l'environnement économique dans lequel évoluent les entreprises.

Les conditions macroéconomiques ont un effet indirect sur les entreprises en impactant leurs revenus, leurs coûts de financement, les prix des matières premières, ou encore les conditions de la chaîne d'approvisionnement, ce qui peut ultimement affecter leur rentabilité et leur capacité à honorer leurs obligations financières. Nous supposons qu'un délai de quelques mois existe entre l'évolution de ces indicateurs et l'impact sur la solvabilité des entreprises, en raison du temps nécessaire pour que les changements économiques se propagent à travers l'économie et affectent les résultats financiers des entreprises.

Nous analyserons donc ces indicateurs avec un décalage temporel, ce qui nous permettra de les utiliser pour prédire à court terme (moins de 6 mois) les défaillances sans avoir à recourir à des prévisions pour ces indicateurs.

Construction de $\mu(\cdot)$

Pour construire la fonction $\mu(\cdot)$, nous allons exploiter les propriétés des processus de Poisson non-homogène. Si nous considérons que le processus $(N_t^{(k)})_{t \geq 0}$ est un processus de Poisson non-homogène d'intensité $m_k(\cdot)$, alors on a

$$\forall t > s \geq 0, \quad N_t^{(k)} - N_s^{(k)} \sim \mathcal{P}(M_k(t) - M_k(s)).$$

Ainsi, on obtient

$$\mathbb{E}[N_t^{(k)} - N_s^{(k)}] = M_k(t) - M_k(s),$$

où $M_k(t) = \int_0^t m_k(u) du$.

Étant donné que nous disposons de l'évolution des indicateurs par mois, nous allons construire la fonction $m_k(\cdot)$ comme une fonction constante par morceaux, sur chaque unité de temps.

En observant que pour deux dates entières successives t_{i+1} et t_i , on a

$$\begin{aligned} M_k(t_{i+1}) - M_k(t_i) &= \int_0^{t_{i+1}} m_k(u) du - \int_0^{t_i} m_k(u) du \\ &= \int_{t_i}^{t_{i+1}} m_k(s) ds. \end{aligned}$$

Puisque $m_k(u)$ est constante sur l'intervalle $[t_i, t_{i+1}]$ et que $t_{i+1} - t_i = 1$, l'expression se simplifie en

$$M_k(t_{i+1}) - M_k(t_i) = (t_{i+1} - t_i)m_k(t_{i+1}) = m_k(t_{i+1}).$$

La valeur de $m_k(\cdot)$ à la date $t \in [t_i, t_{i+1}]$ correspond donc au nombre moyen de défaillances comptabilisées par le processus $N^{(k)}$ sur l'intervalle $[t_i, t_{i+1}]$.

Pour déterminer la valeur de $m_k(\cdot)$, nous avons opté pour une régression linéaire afin d'expliquer le nombre de défaillances observées chaque mois à partir des indicateurs macroéconomiques sélectionnés pour chaque secteur d'activité. Nous avons choisi d'utiliser une régression linéaire car il s'agit d'un modèle d'apprentissage statistique parmi les plus simples à mettre en oeuvre et à interpréter. Ce modèle nous permettra d'avoir une approximation raisonnable du lien entre les défaillances et l'environnement économique. En outre, la régression linéaire est un outil couramment utilisé en risque de crédit pour projeter la variable latente Z du modèle de Merton (1.5) en fonction des projections des variables macroéconomiques.

Notons :

- $N^{(k)}$ le vecteur de taille n contenant le nombre de défaillances observées chaque mois dans le secteur k sur la période étudiée ;
- X_k la matrice de taille $n \times (n_k + 1)$, comprenant les n_k indicateurs conservés en tant que variable explicative pour le secteur k , ainsi qu'un vecteur unitaire correspondant à l'intercept du modèle.

Le modèle de régression linéaire prend la forme suivante

$$N^{(k)} = X\theta^{(k)} + \xi^{(k)}, \quad (2.23)$$

où $\theta^{(k)} = \begin{pmatrix} \theta_0^{(k)} \\ \vdots \\ \hat{\theta}_{n_k}^{(k)} \end{pmatrix}$ est un vecteur de taille $n_k + 1$ contenant les coefficients du modèle et $\xi^{(k)} \sim \mathcal{N}(0, I_n)$

est un vecteur de taille n représentant le bruit du modèle.

Les indicateurs macroéconomiques utilisés sont ceux présentés dans la sous-section précédente, analysés avec des décalages temporels de 1, 3, 6, 9 et 12 mois et lissés par moyennes mobiles sur 3, 6 et 12 mois afin d'éliminer les fluctuations transitoires de ces indicateurs.

L'intégration de multiples indicateurs avec différents décalages dans la matrice X peut cependant poser des problèmes de multicollinéarité. D'une part, deux variables représentant le même indicateur avec des décalages distincts peuvent être fortement corrélées entre elles. D'autre part, certains indicateurs peuvent eux-mêmes être corrélés, comme c'est le cas entre l'inflation et les taux d'intérêt, ces derniers étant ajustés par la BCE pour contrôler l'inflation.

Pour remédier à ce problème de multicollinéarité, nous procédons à une première identification « manuelle » des variables explicatives pour chaque secteur d'activité. Cette sélection est basée sur les

études sectorielles des assureurs-crédit ainsi que sur une analyse exploratoire des données (*Exploratory Data Analysis*). Nous confrontons ensuite cette première sélection aux résultats obtenus par différentes méthodes de sélection de variables (*Backward, Forward, Step-wise*). Afin de préserver une interprétabilité suffisante du modèle, nous limitons le nombre de variables explicatives retenues à un maximum de cinq.

Les coefficients sont estimés à l'aide de la méthode des moindres carrés ordinaires, en résolvant le problème de minimisation suivant

$$\arg \min_{\theta^{(k)}} \| N^{(k)} - X_k \theta^{(k)} \|^2,$$

dont la solution analytique est donnée par $\hat{\theta}^{(k)} = (X_k^\top X_k)^{-1} X_k^\top N^{(k)}$, (CORNILLON et al., 2023).

La fonction $m_k(\cdot)$ est alors obtenue à partir des prédictions du modèle de régression linéaire, sa valeur à la date t est calculée comme suit

$$m_k(t) = X_k(\tau) \hat{\theta}^{(k)},$$

où $\tau = \lfloor t \rfloor$ est la partie entière de t et $X_k(\tau)$ est un vecteur ligne contenant la valeur des variables explicatives à la date τ .

Les variables explicatives retenues pour chaque secteur d'activité sont présentées dans le tableau 2.4 ci-après. Pour chaque indicateur, nous adopterons la notation *indicateur*_L, où L représente le décalage, en mois, de ce dernier. On peut noter que certains indicateurs comme les taux d'intérêts et le GSCPI sont conservés pour la plupart des secteurs. De plus, on remarque que pour la plupart des secteurs, différents indicateurs sont sélectionnés avec des décalages variés, permettant ainsi de saisir les effets économiques à la fois à court et à moyen terme.

Secteur d'activité	Variables explicatives
(A)	Inflation_12, BCI_6, Taux_Interets_6, FAO_3
(I)	Inflation_9, PIB_6, GSCPI_3, BCI_Industrie_3, Taux_Interets_12
(C)	BCI_Construction_3, Taux_Interets_12, GSCPI_3
(S)	FAO_9, Taux_Interets_12, GSCPI_3

TABLE 2.4 : Variables explicatives conservées pour chaque secteur d'activité

Pour le secteur *Enseignement, santé humaine, action sociale et services aux ménages*, c'est-à-dire les services non marchands, nous n'avons pas réussi à créer une fonction $m_k(\cdot)$ pertinente à partir des indicateurs économiques mentionnés. Par conséquent, nous avons choisi de fixer la fonction $m_4(t) = cste, \forall t$. Nous espérons que les effets de contagion avec les autres secteurs permettront de révéler des évolutions dans la tendance générale du risque sur ce secteur.

La fonction $m_k(\cdot)$ que nous avons construite ne correspond pas exactement à la fonction $\mu_k(\cdot)$ que nous souhaitons utiliser dans le cadre du modèle (2.22). En effet, $m_k(\cdot)$ décrit l'évolution du risque de défaillance en fonction des fluctuations économiques, mais elle néglige l'effet d'excitation des événements passés. Autrement dit, $m_k(\cdot)$ représente ce que l'on obtiendrait si les paramètres d'excitation α et β étaient nuls.

Pour ajuster correctement la fonction $\mu_k(\cdot)$, qui doit modéliser l'intensité de base du risque en tenant compte de ces paramètres, nous proposons de la définir comme suit

$$\mu_k(t) = m_k(t) * \gamma_k, \quad (2.24)$$

où γ_k est un facteur d'échelle qui sera estimé par maximum de vraisemblance.

L'idée ici est de réajuster $m_k(\cdot)$, qui est issu d'une régression linéaire classique ayant pour objectif d'expliquer le nombre de défaillances par mois et tend par conséquent à être surévalué.

Pour le secteur des services non marchands, pour lequel nous avons choisi de fixer la fonction $m_4(\cdot)$ à une valeur constante, faute de meilleures alternatives, nous déterminons γ_4 de sorte que $\mu_4(t) = \hat{\lambda}_4, \forall t$, où $\hat{\lambda}_4$ est le paramètre d'intensité de base estimé après calibrage du modèle initial. Nous espérons que les effets de contagion avec les autres secteurs permettront de révéler des évolutions dans la tendance générale du risque sur ce secteur.

La figure 2.12 présente l'évolution du nombre de défaillances dans un secteur d'activité donné ainsi que celle de la fonction $\mu_k(\cdot)$ associée. Le paramètre γ_k estimé pour ce secteur est de 0.19, indiquant que plus de 80% des défaillances sont attribuables à la composante d'excitation du processus de Hawkes.

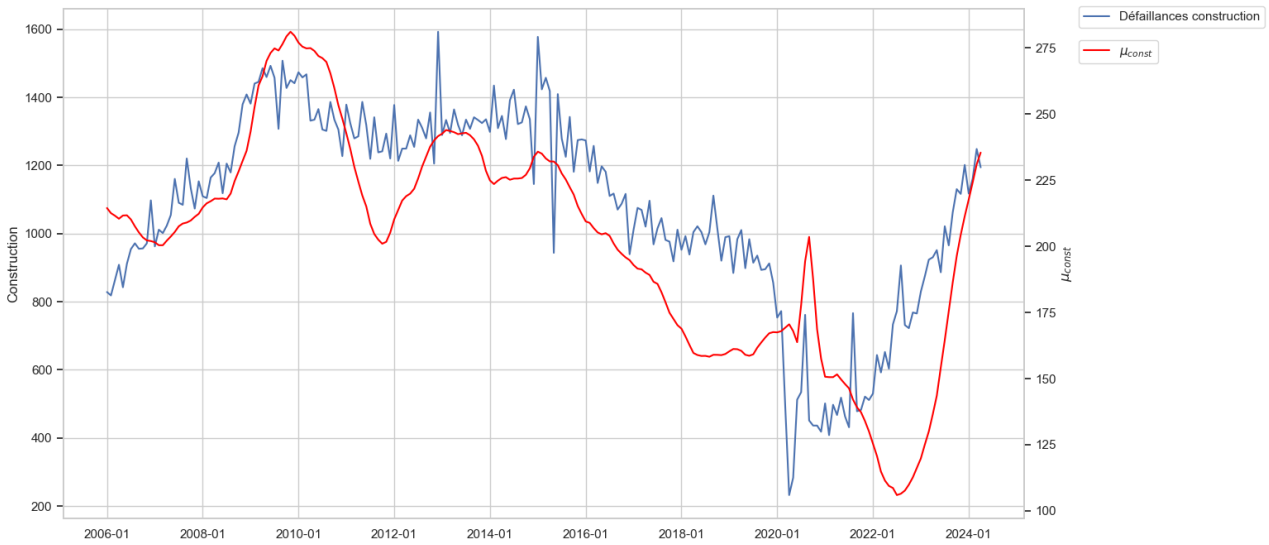


FIGURE 2.12 : Évolution du nombre de défaillances et de la fonction $\mu(\cdot)$ sur un secteur d'activité

De manière générale, on constate que $\mu_k(\cdot)$ suit une tendance similaire à celle du nombre de défaillances sur ce secteur. Cependant, durant la période de la pandémie de COVID-19, la fonction $\mu_k(\cdot)$ peine à capturer fidèlement la dynamique observée. Cette période est en effet marquée par d'importantes mesures de soutien gouvernementales, qui ne sont pas reflétées dans les indicateurs économiques utilisés pour la construction de $\mu(\cdot)$. Cela met en lumière une limite du modèle dans sa capacité à saisir des événements exogènes exceptionnels, non pris en compte par les données macroéconomiques conventionnelles.

La fonction $\mu(\cdot)$, telle que définie, repose sur des indicateurs macroéconomiques à des horizons temporels spécifiques (avec des décalages de 1, 3, 6, 9 et 12 mois). Cela signifie que le modèle (2.22) ne pourra fournir de prédictions qu'à court terme, jusqu'à l'horizon correspondant au plus petit décalage des indicateurs utilisés pour définir $\mu(\cdot)$.

Pour aller au-delà de cet horizon, plusieurs options peuvent être envisagées. La première consiste à exploiter les anticipations économiques à court terme publiées régulièrement par des institutions telles que les banques centrales, l'OCDE ou des instituts de conjoncture. Une autre possibilité, en l'absence de telles prévisions ou pour explorer des hypothèses particulières, serait de définir manuellement des scénarios macroéconomiques. Cela permettrait de simuler différentes trajectoires de risque à moyen ou long terme. Toutefois, cette approche nécessiterait une réflexion approfondie sur les hypothèses de

base pour assurer la cohérence des scénarios envisagés.

Néanmoins, projeter à court terme s'aligne bien avec le risque que nous modélisons, car l'assurance-crédit est une branche courte. Selon le rapport annuel 2023 de Coface, plus de 95% de l'exposition totale du Groupe concerne des risques à court terme, avec des durées de crédit rarement supérieures à 180 jours.

Estimation des paramètres

Le calibrage du modèle (2.22) suit la même méthode que celle appliquée pour le premier modèle. Nous cherchons à résoudre le problème de minimisation suivant

$$\min_{\gamma \in \mathbb{R}^d, \alpha \in \mathbb{R}^{d \times d}, \beta \in \mathbb{R}^d} -\ell(\gamma, \alpha, \beta), \quad (2.25)$$

où ℓ est la log-vraisemblance du modèle (2.22), définie par

$$\ell(\gamma, \alpha, \beta) = \sum_{i=1}^n \log(\lambda_{d_i}^*(t_i)) - \sum_{j=1}^d \Lambda_j(T), \quad (2.26)$$

avec

$$\lambda_{d_i}^*(t_i) = m_{d_i}(t_i) \times \gamma_{d_i} + \sum_{(t_k, d_k): t_k < t_i} \alpha_{d_i, d_k} e^{-\beta_{d_i}(t_i - t_k)} \quad \text{et} \quad \Lambda_j(T) = \int_0^T \lambda_j^*(u) du.$$

Le paramètre λ est ici remplacé par le paramètre d'échelle γ , qui représente l'ajustement de la fonction $m(\cdot)$ pour tenir compte du poids des paramètres α et β .

Le temps d'exécution de l'algorithme pour résoudre le problème (2.25) est similaire à celui du premier modèle, le nombre de paramètres et l'historique considéré sont identiques et l'ajout de deux évaluations de $m(\cdot)$ à chaque calcul de l'intensité conditionnelle ne semble pas significatif dans l'évaluation de l'intensité conditionnelle.

Comme pour le premier modèle, une étape de pré-calibrage a été menée pour éviter une convergence trop lente de l'algorithme et prévenir la convergence vers des optima locaux. Pour cette étape, le paramètre d'échelle a été initialement estimé comme suit

$$\hat{\gamma}_k = \frac{\hat{\lambda}_k}{\bar{N}^{(k)}},$$

avec :

- $\hat{\lambda}_k$ est le paramètre d'intensité de base estimé après calibrage du modèle initial (où l'intensité est supposée constante) ;
- $\bar{N}^{(k)}$ étant le nombre moyen de défaillances mensuelles observées dans le secteur k .

Avec cet estimateur pour γ_k , nous avons pour tout $t \geq 0$,

$$\frac{\mu_k(t)}{m_k(t)} = \frac{\hat{\lambda}_k}{\bar{N}^{(k)}},$$

ce qui implique que le niveau de la fonction $\mu_k(\cdot)$ est analogue à l'intensité de base estimée dans le modèle initial, ajustée par la fonction $m_k(\cdot)$. Cette dernière traduit l'évolution dans le temps de l'environnement économique et son impact sur le nombre de défaillances observées dans le secteur considéré.

Les résultats du pré-calibrage, qui seront conservés pour la suite des travaux, sont présentés dans le tableau 2.5 ci-après. Ils sont similaires à ceux obtenus avec le premier modèle (cf. tableau 2.2). Cependant, on note que pour la plupart des secteurs, les valeurs sont légèrement inférieures avec le modèle actuel. Cette diminution peut s'expliquer par le fait que le modèle actuel attribue une partie des variations historiques au paramètre μ , réduisant ainsi le rôle de la composante d'excitation. De plus, la vraisemblance obtenue avec le pré-calibrage du modèle actuel est supérieure à celle obtenue avec le pré-calibrage du premier modèle pour tous les secteurs.

Secteur d'activité	γ	α	β
(A)	0.9994	0.0010	0.0033
(I)	0.3952	1.1318	1.9154
(C)	0.1863	3.1020	3.7332
(E)	0.6407	0.6822	1.6478
(S)	0.3684	2.9899	4.5118

TABLE 2.5 : Résultat du pré-calibrage sur chaque secteur

La définition des paramètres initiaux suit les mêmes règles que celles définies pour le modèle (2.18). Les paramètres γ_k , à l'instar des paramètres λ_k , sont initialisés avec les paramètres issus du pré-calibrage.

Comme pour le premier modèle, c'est l'approche statistique qui fournit les meilleurs résultats en termes de vraisemblance et de prédictions. Les paramètres conservés pour la suite, arrondis à 10^{-4} pour une meilleure lisibilité, sont les suivants

$$\gamma = \begin{pmatrix} 0.9929 \\ 0.2784 \\ 0.3953 \\ 0.6919 \\ 0.4312 \end{pmatrix}, \quad \alpha^T = \begin{pmatrix} 0.0006 & 1.3013 & 0.1085 & 0.0931 & 0.0259 \\ 0.0009 & 0.0001 & 0.5527 & 0.1525 & 2.6982 \\ 0.0001 & 0.3574 & 2.9797 & 0.0000 & 2.4325 \\ 0.0002 & 0.3019 & 2.6711 & 0.4725 & 1.3843 \\ 0.0000 & 0.1177 & 0.0017 & 0.0000 & 3.1444 \end{pmatrix}, \quad \text{et} \quad \beta = \begin{pmatrix} 0.7061 \\ 3.7146 \\ 6.8195 \\ 1.6408 \\ 7.9696 \end{pmatrix}. \quad (2.27)$$

L'analyse des paramètres obtenus révèle des dynamiques intéressantes concernant les interactions entre secteurs identifiées par le modèle. Notamment, les défaillances d'entreprises dans le secteur agricole semblent être principalement dictées par l'évolution de l'environnement économique global et non par des effets de contagion liés aux interactions avec d'autres entreprises. Cependant, bien que les défauts dans le secteur de l'agriculture soient peu sensibles aux perturbations des autres secteurs, leur survenance semble tout de même avoir un impact sur ceux-ci.

À l'inverse, le secteur des services marchands est fortement affecté par les défaillances des autres secteurs, tandis qu'il influence relativement peu les autres secteurs. Ces asymétries ont du sens d'un point de vue économique, car l'agriculture se situe au début de la chaîne d'approvisionnement, tandis que les services marchands en sont la dernière étape. Cependant, il convient de nuancer cette observation. En effet, une défaillance d'une entreprise devrait théoriquement influencer ses fournisseurs, qui risquent de ne pas être payés.

La figure 2.13 présente un *barplot* des paramètres α_k pour le modèle initial (2.18) et le modèle final (2.22). À quelques exceptions près, les inter-dépendances sectorielles observées après le calibrage du second modèle sont globalement similaires à celles identifiées avec le premier modèle. Parmi ces exceptions, on peut citer l'excitation du secteur industriel vis-à-vis de l'agriculture, qui était jugée nulle

par le premier modèle malgré des corrélations élevées entre ces deux secteurs, ou encore l'excitation entre le secteur de la construction et les services non marchands, où les corrélations suggèrent également un lien significatif. Nous estimons donc que les changements apportés par rapport au premier modèle sont appropriés.

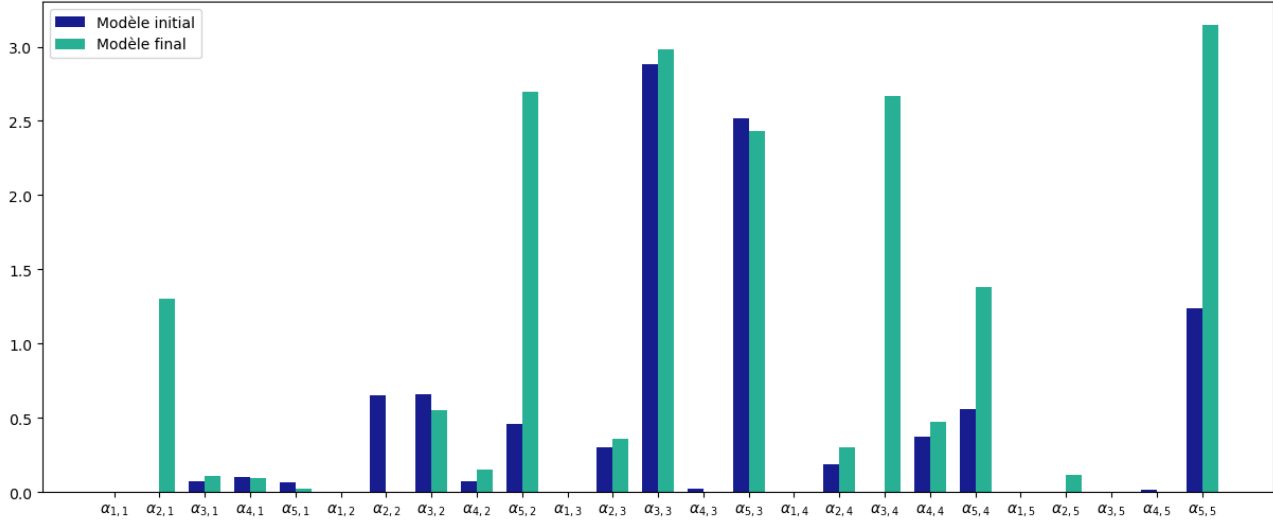


FIGURE 2.13 : Comparaison du paramètre α pour les deux modèles

Une autre observation notable est que les paramètres d'excitation estimés dans ce modèle sont plus élevés que pour le premier modèle, ce qui peut paraître contre-intuitif, car une partie des fluctuations dans les défaillances est désormais expliquée par l'intensité de base du processus. Cela s'explique toutefois par le fait que, dans le premier modèle, les paramètres d'excitation devaient capturer l'ensemble des évolutions dans le nombre de défaillances et captaient donc les tendances de manière plus globale, ce qui avait pour effet de contenir leurs valeurs à un niveau modéré. Dans le modèle actuel, les paramètres d'excitation capturent les interactions de façon plus précise, étant libérés de la nécessité de représenter l'évolution globale. Toutefois, il convient de souligner que les interprétations de ces paramètres doivent être abordées avec prudence en raison des instabilités observées sur le calibrage des paramètres en section 2.3.3.

2.3 Évaluation des performances du modèle

Cette section présente et analyse les résultats obtenus par le modèle (2.22). L'objectif principal est d'évaluer la qualité des prédictions produites par le modèle, en examinant à la fois sa précision et sa capacité à s'ajuster aux données de défaillances observées. Pour conclure, une discussion sur les avantages et les limites du modèle sera proposée en vue d'identifier d'éventuelles pistes d'amélioration.

2.3.1 Analyse des résultats

Une méthode couramment utilisée pour évaluer la précision d'un modèle est le *Backtesting*. Cela consiste à effectuer des prédictions sur des périodes passées, puis à comparer ces prédictions aux données effectivement observées. Dans notre cas, le modèle a été calibré sur la période 2010-2019, ce qui nous permet d'utiliser les trois années suivantes (2020-2023) comme période de test pour évaluer la précision du modèle sans risque d'*overfitting*. Ces années test sont indépendantes de la période

de calibrage et permettent une évaluation plus objective des performances prédictives. Il n'est pas possible de tester notre modèle sur des années antérieures à 2010, faute de données sur les indicateurs définissant $\mu(\cdot)$.

Afin de quantifier l'erreur de prédiction, nous avons choisi d'utiliser la *Mean Absolute Percentage Error* (MAPE), une métrique largement reconnue pour sa facilité d'interprétation. Compte tenu des variations importantes dans le nombre de défaillances au fil du temps et entre différents secteurs, nous privilégions l'utilisation d'une mesure d'erreur relative plutôt qu'absolue. La MAPE permet de normaliser les erreurs par rapport aux valeurs observées, ce qui la rend particulièrement adaptée pour comparer les performances des prédictions à travers différentes échelles. Un autre avantage motivant notre choix pour la MAPE est qu'elle est moins sensible aux valeurs aberrantes que d'autres métriques comme la *Root Mean Square Error* (RMSE) ou la *Mean Absolute Error* (MAE). Cela est particulièrement important dans notre contexte, où les défaillances suivent un schéma irrégulier d'un mois à l'autre, tout en maintenant une tendance générale sous-jacente. Enfin, nos données ne comportant pas de valeurs proches de zéro, le risque de voir la MAPE « exploser » est nul. Ainsi, la MAPE nous semble bien adaptée à notre situation et devrait fournir une mesure fiable et cohérente de l'erreur de prédiction du modèle.

Notons :

- n est le nombre total de valeurs prédites ;
- A_i est la valeur réellement observée à la date i ;
- F_i est la valeur prédite à la date i ;

alors la MAPE est définie par la formule suivante

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right|. \quad (2.28)$$

La MAPE est comprise entre 0 et 1. Plus sa valeur est proche de 0, plus la précision des prédictions est élevée. À l'inverse, une MAPE proche de 1 reflète des prédictions de très mauvaise qualité. Le tableau 2.6 ci-dessous présente une interprétation de la MAPE, telle que proposée dans LEWIS (1982).

MAPE	< 10%	10% - 20%	20% - 50%	≥ 50%
Précision	Excellente	Bonne	Raisonnable	Mauvaise

TABLE 2.6 : Interprétation de la MAPE - LEWIS, 1982

Ainsi, notre modèle sera jugé performant si la MAPE associée est inférieure à 10%.

Plusieurs simulations du processus sont effectuées sur diverses périodes passées pour évaluer la précision des prédictions. Ces simulations sont générées comme le prolongement de l'historique des données afin de supprimer l'effet de bord. La moyenne empirique des trajectoires simulées est ensuite retenue comme prédiction. Nous commençons par une visualisation graphique des résultats obtenus. La figure 2.14 illustre les résultats obtenus sur chaque secteur d'activité pour des prédictions démarrant en janvier 2022 et jusqu'à juin 2024. Nous considérons que juin 2024 est la dernière date connue sur les défaillances puisque les études mensuelles publiées par la Banque de France fournissent des données provisoires, nous prenons donc une marge de 3 mois pour nous assurer qu'il ne manque pas de tardifs.



FIGURE 2.14 : Prédications du modèle (2.22) avec les paramètres (2.27) de janvier 2022 à juin 2024

Visuellement, les résultats sont assez satisfaisants. Le modèle parvient à bien modéliser la tendance de reprise sur le nombre de défaillances observées depuis 2022. On note cependant que, pour le secteur *Enseignement, santé humaine, action sociale et services aux ménages*, où le paramètre d'intensité de base a été maintenu constant faute de meilleures alternatives, les effets de contagion avec les autres secteurs ne semblent pas suffisants pour reproduire correctement la tendance observée.

Par ailleurs, dans le secteur de l'*Industrie*, on constate que la courbe prédite dépasse celle observée à partir de la fin de l'année 2023. Cette situation peut être attribuée à la construction de l'intensité de base pour ce secteur, qui accorde une importance significative à l'indicateur GSCPI avec un décalage de trois mois. Ce décalage est nécessaire pour permettre des prédictions à court terme. Cependant, cet indicateur est conçu pour refléter en temps réel les tensions dans les chaînes d'approvisionnement.

Ainsi, à la fin de 2023, l'intensité de base est légèrement surestimée par rapport à l'état réel des conditions du secteur, en raison de ce retard de trois mois.

Le tableau 2.7 présente les MAPE calculés pour ces prédictions, confirmant les observations graphiques. En particulier, les secteurs de l'*Industrie* et de l'*Enseignement, santé humaine, action sociale et services aux ménages* affichent les MAPE les plus élevées. Ces valeurs restent néanmoins tout à fait acceptables. La moyenne globale de 0,10 témoigne d'une bonne performance du modèle sur l'ensemble des secteurs.

Période de prédiction	(A)	(I)	(C)	(E)	(S)	Moyenne
01/2022 - 06/2024	0.1042	0.1337	0.0780	0.1209	0.0675	0.1009

TABLE 2.7 : MAPE du modèle (2.22)

Les prédictions réalisées sur un horizon de projection étendu fournissent des résultats corrects, suggérant qu'une estimation précise de la fonction $\mu(\cdot)$ joue un rôle central dans la qualité des prévisions. Cela souligne l'importance d'affiner la définition de cette fonction en y intégrant des indicateurs sectoriels plus spécifiques. Par ailleurs, ces résultats laissent entendre qu'avec de bonnes projections sur les valeurs futures des indicateurs économiques, il serait possible de prédire les défaillances à plus long terme avec une précision honorable en utilisant le modèle (2.22). Les figures 2.15 et 2.16, illustrant les résultats obtenus pour une période allant de 2011 à 2023, montrent que le modèle parvient à capter les grandes tendances sur cette longue période.

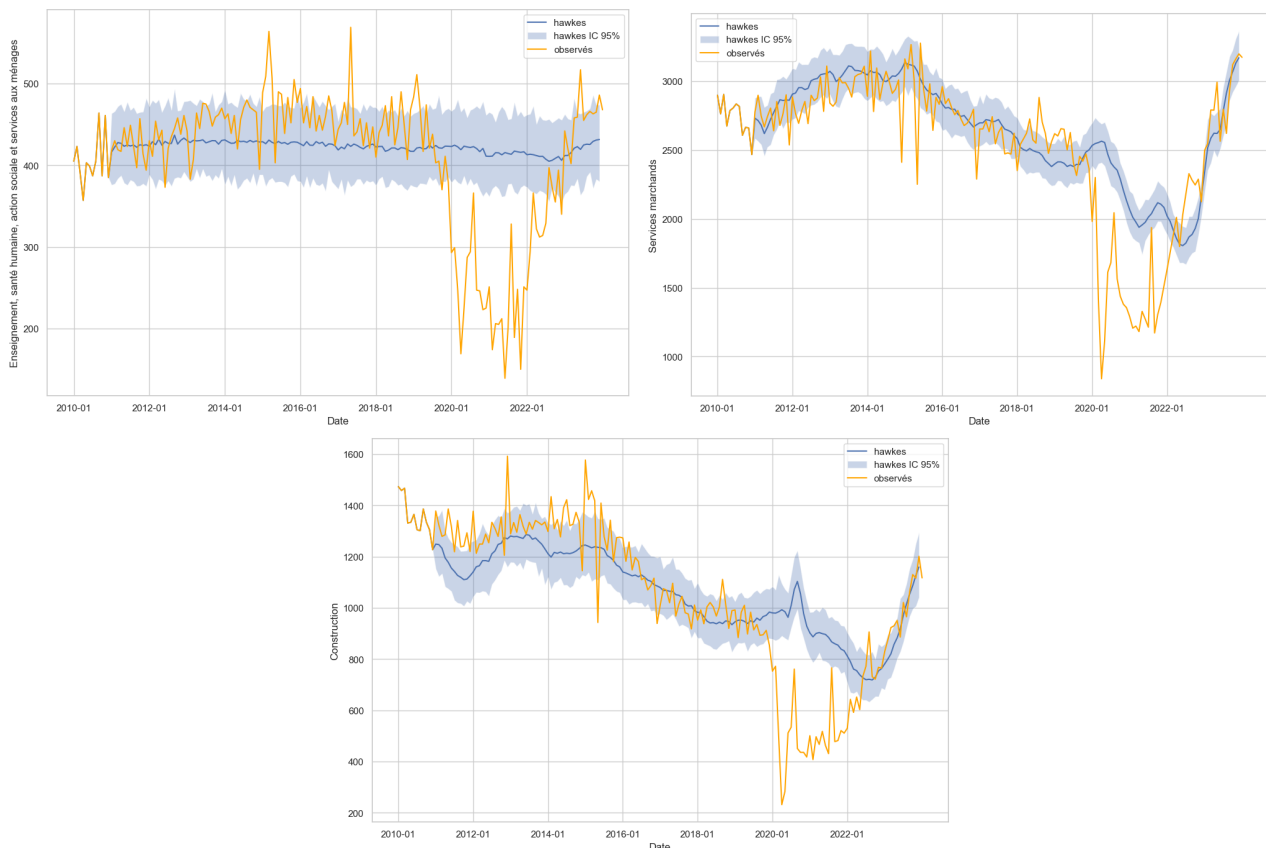


FIGURE 2.15 : Prédications du modèle (2.22) avec les paramètres (2.27) de 2011 à 2023

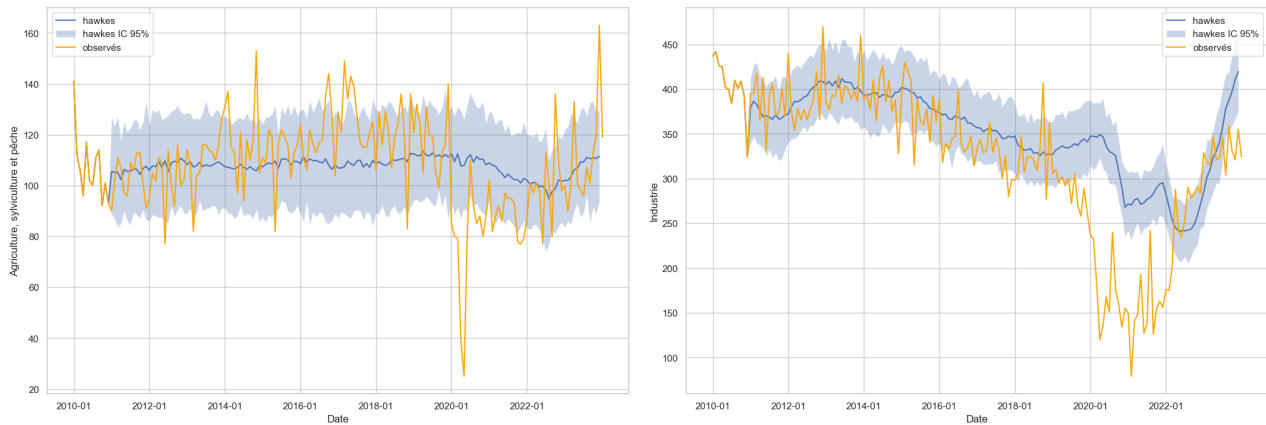


FIGURE 2.16 : Prédications du modèle (2.22) avec les paramètres (2.27) de 2011 à 2023

Bien que globalement satisfaisant, le modèle peine à reproduire correctement la baisse soudaine du nombre de défaillances observée durant la période Covid et post-Covid, en grande partie liée aux mesures de soutien gouvernementales. La figure 2.17 présente les prédictions pour deux secteurs d'activité en démarrant au milieu du « creux » dans l'évolution du nombre de défaillances. On constate que, en janvier 2021, même après une année entière marquée par un niveau exceptionnellement bas de défaillances, le modèle n'ajuste pas suffisamment l'intensité conditionnelle de chaque processus. La décroissance de la composante d'excitation générée par les événements récents reste trop faible pour dicter une diminution assez forte de l'intensité conditionnelle pour compenser la surévaluation de l'intensité de base à cette période.

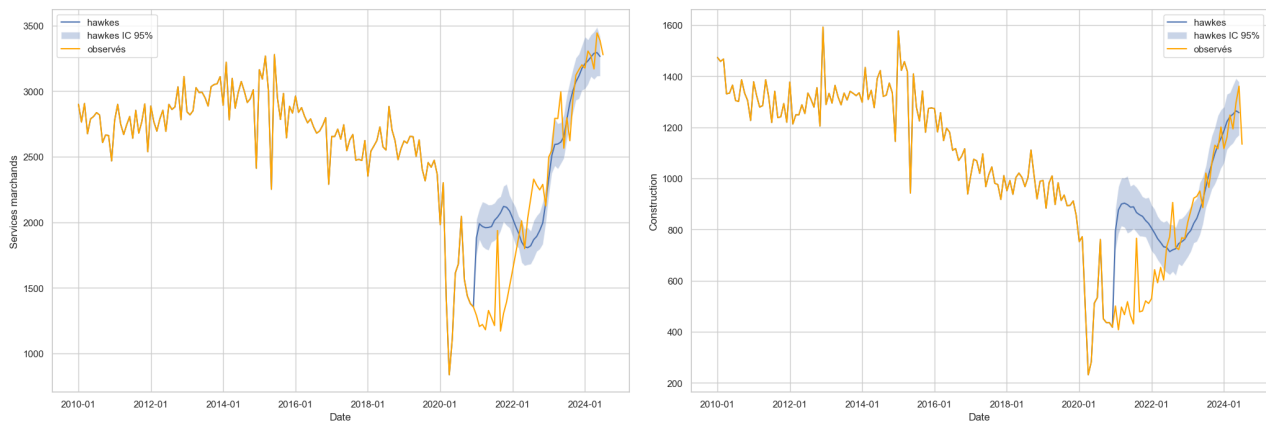


FIGURE 2.17 : Prédications du modèle (2.22) de janvier 2021 à juin 2024

Le tableau 2.8 ci-après présente les MAPE calculés pour différentes périodes de prédiction. Les résultats montrent clairement que les performances du modèle se dégradent durant la période perturbée par la crise sanitaire (2020-2022), avec une augmentation sensible des MAPE lorsqu'on inclut cette période dans les prévisions.

Pour la période de prédiction allant de janvier 2021 à décembre 2022, la MAPE moyenne s'élève à 0.2522, ce qui signifie que, durant la période liée à la crise sanitaire, la différence moyenne entre les valeurs prédites et les valeurs observées est de 25%. Ce niveau est nettement supérieur à celui obtenu sur une période plus longue, indiquant que la dégradation observée est atténuée par les bonnes performances du modèle sur les périodes en dehors de cette fenêtre perturbée. En effet, la MAPE moyenne sur l'ensemble de la période 2011-2023 reste relativement basse, à 0.1297.

Période de prédiction	(A)	(I)	(C)	(E)	(S)	Moyenne
01/2011 - 12/2023	0.1279	0.1418	0.1368	0.1392	0.1097	0.1297
01/2021 - 06/2024	0.1173	0.2256	0.1723	0.2244	0.1463	0.1772
01/2021 - 12/2022	0.1320	0.2994	0.1723	0.3281	0.2321	0.2522

TABLE 2.8 : MAPE du modèle (2.22)

Jusqu'à présent, les résultats suggèrent que nos prévisions sont largement dictées par la fonction $\mu(\cdot)$, ce qui pourrait laisser penser que la composante d'excitation est négligeable. Cela amène à se demander si la fonction $\mu(\cdot)$ ne serait pas suffisante à elle seule. Pour répondre à cette question, nous comparons les performances de notre modèle à celles d'un processus de Poisson non-homogène, dont l'intensité $m(\cdot)$ correspondrait à $\mu(\cdot)$ avec $\gamma_k = 1$ pour tout k . La figure 2.18 illustre les prédictions obtenues pour deux secteurs d'activité sur deux périodes distinctes.

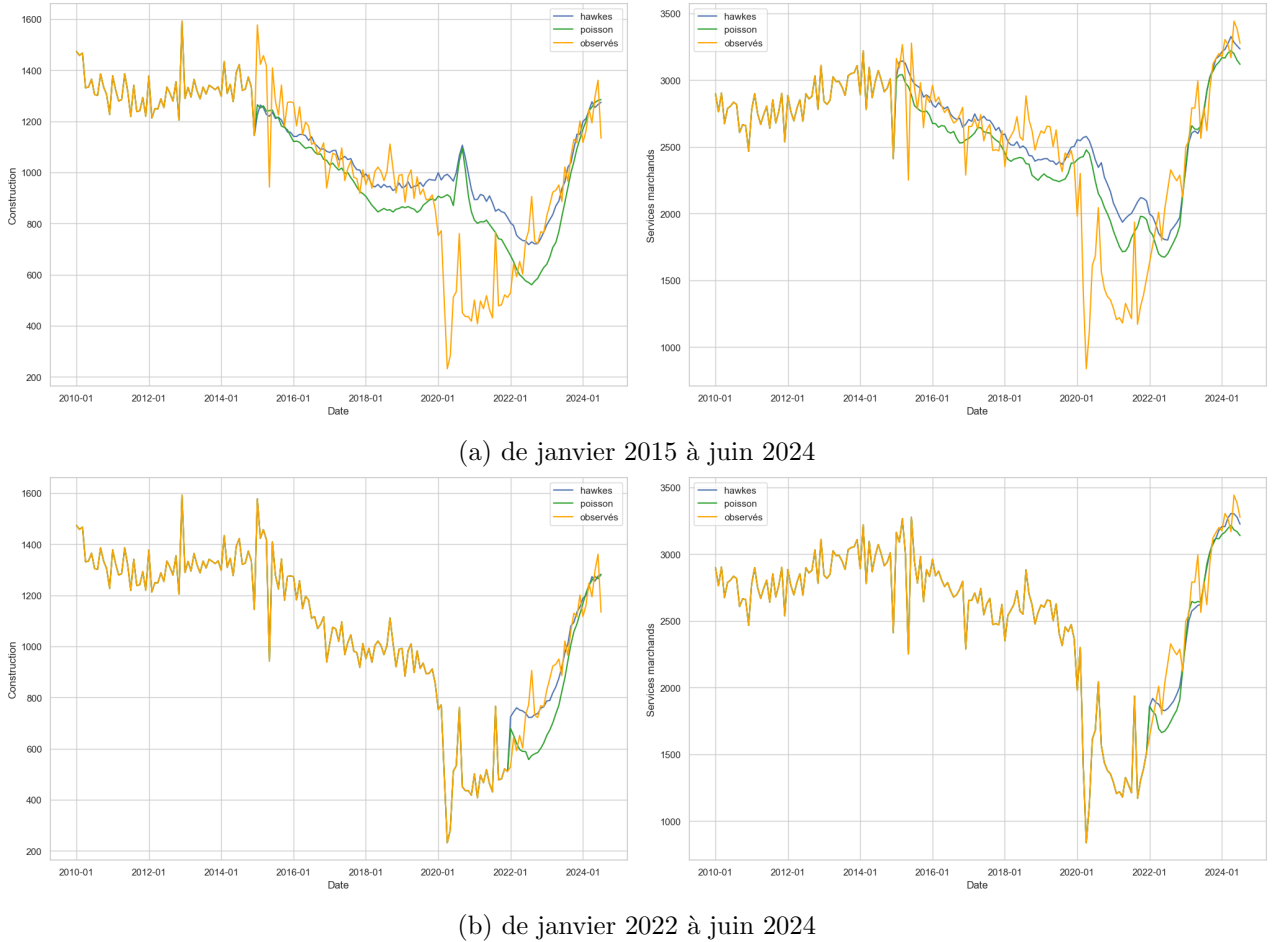


FIGURE 2.18 : Comparaison des prédictions du modèle (2.22) avec celles d'un processus de Poisson non-homogène

On observe que le processus de Poisson non-homogène fournit des résultats convenables, mais reste moins précis que le modèle (2.22). La contribution de la composante d'excitation au sein de

notre modèle se révèle donc bien réelle, bien qu'elle joue un rôle moins significatif que prévu. Ces observations sont corroborées par les valeurs de la MAPE présentées dans le tableau 2.9 ci-dessous, confirmant que, même si le processus de Poisson donne des résultats satisfaisants, le modèle de Hawkes apporte un gain de précision notable. Nous avons également inclus la MAPE du premier modèle de Hawkes (2.18) afin de mesurer les améliorations apportées par l'intégration de facteurs économiques dans l'intensité de base. En outre, un modèle « naïf » a été introduit, consistant à prédire les valeurs par la moyenne empirique des données. Ce modèle sert de point de comparaison pour évaluer l'efficacité des autres modèles plus complexes.

Modèle	(A)	(I)	(C)	(E)	(S)	Moyenne
Hawkes (2.22)	0.1247	0.1319	0.1417	0.1376	0.0951	0.1262
Hawkes (2.18)	0.1259	0.1816	0.2143	0.1333	0.1441	0.1598
Poisson	0.1259	0.1546	0.1481	0.2107	0.1015	0.1476
Naïf	0.1573	0.2599	0.2556	0.3024	0.1665	0.2283

TABLE 2.9 : MAPE des différents modèles sur la période 2011-2024

Prédictions pour l'année 2024

Les figures 2.19 et 2.20 illustrent les prévisions du modèle (2.22) pour la période de juin à décembre 2024. Selon les résultats du modèle, une stabilisation, voire une légère diminution, du nombre de défaillances mensuel est attendue dans les secteurs de l'*Agriculture, sylviculture et pêche* ainsi que de la *Construction* d'ici la fin de l'année 2024. Pour le secteur de l'*Industrie*, le modèle anticipe une augmentation des défaillances en juillet, suivie d'une diminution progressive au cours des mois suivants. Par ailleurs, les secteurs de l'*Enseignement, santé humaine, action sociale et services aux ménages* ainsi que des *Services marchands* devraient afficher un nombre de défaillances mensuel en baisse dans les mois à venir. Dans l'ensemble, ces prévisions suggèrent une tendance générale vers une stabilisation des défaillances dans les six prochains mois, pour tous les secteurs.

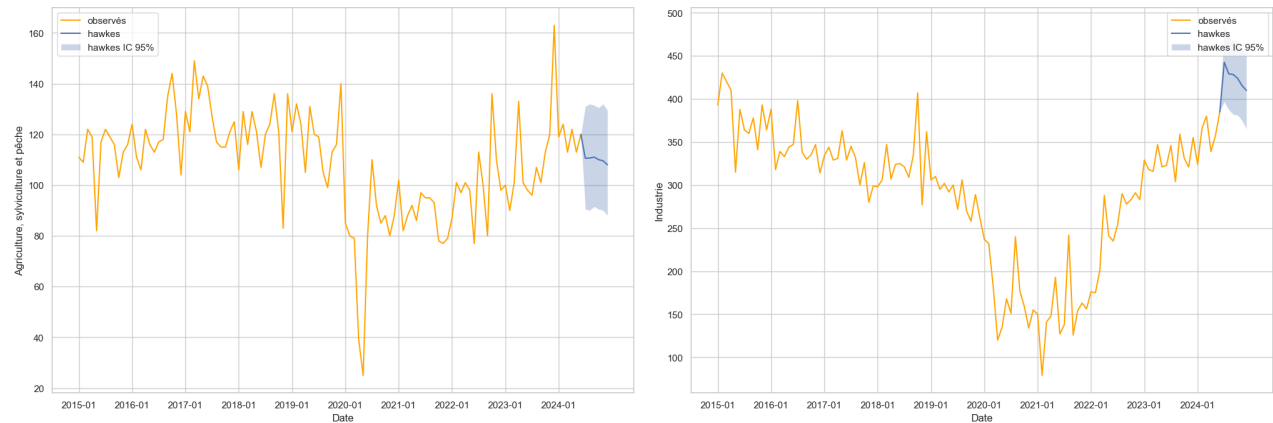


FIGURE 2.19 : Prédictions du modèle (2.22) avec les paramètres (2.27) de juin à décembre 2024



FIGURE 2.20 : Prédictions du modèle (2.22) avec les paramètres (2.27) de juin à décembre 2024

La figure 2.21 présente l'évolution du nombre de défaillances cumulées sur 12 mois glissants, tous secteurs confondus. D'après notre modèle, le nombre total de défaillances devrait atteindre le nombre record de 65225 défaillances au total en 2024, avec un intervalle de confiance à 95% entre 63537 et 66939.

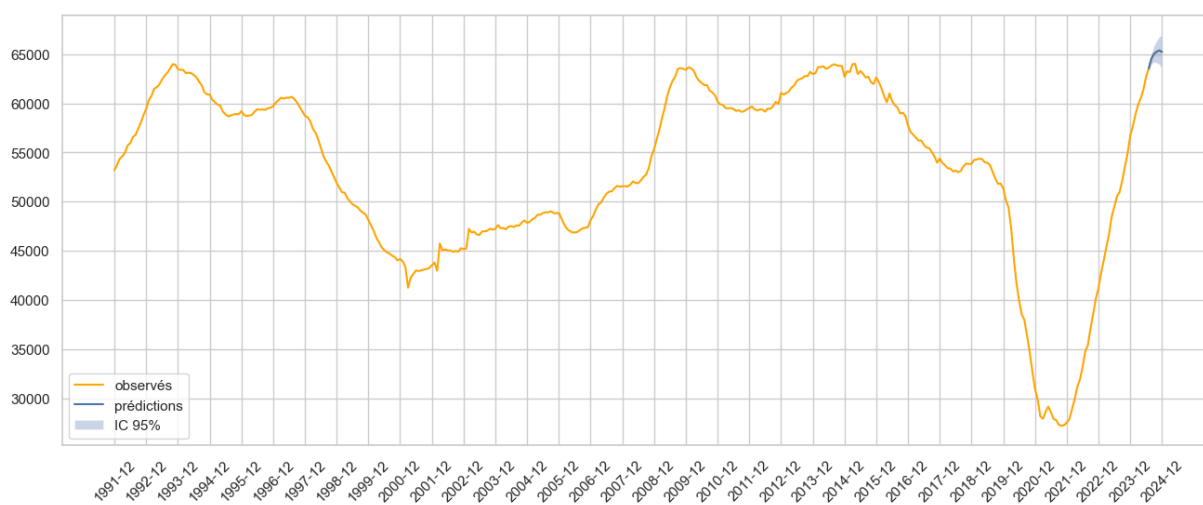


FIGURE 2.21 : Évolution du nombre de défaillances en France, cumul sur 12 mois glissants - Prédictions pour la fin d'année 2024

Ce chiffre est largement supérieur au niveau pré-pandémique moyen de 59 342. De plus, il dépasse significativement le chiffre de 60 700 annoncé par ALLIANZ TRADE (2024) dans sa dernière étude sur les défaillances d'entreprises. Cet écart peut s'expliquer par le fait que nos prévisions intègrent des données de défaillances actualisées à fin juin, ainsi qu'une évaluation de l'état de l'économie mise à jour à fin septembre, tandis que celles d'Allianz Trade reposent sur des informations disponibles à fin février. Cela souligne l'importance d'une surveillance régulière des défaillances en lien avec l'évolution de la conjoncture économique.

Cependant, la tendance observée au cours des derniers mois de l'année suggère que cette dynamique haussière est susceptible de se terminer, laissant entrevoir une stabilisation en 2025. Par ailleurs, l'assouplissement de la politique monétaire des banques centrales, qui devrait favoriser l'amélioration de la trésorerie des entreprises, soutient cette analyse.

2.3.2 Qualité d'ajustement

L'évaluation de la qualité d'ajustement d'un processus de Hawkes repose sur le théorème 2.3, présenté dans DALEY et VERE-JONES (2003).

Théorème 2.3. Soient $\{t_1, \dots, t_n\}$ les réalisations sur $[0, T]$ d'un processus ponctuel d'intensité $\lambda^*(.)$. Si $\lambda^*(.)$ est positive sur $[0, T]$ et $\Lambda(T) < +\infty$ presque sûrement, alors les points $\{\Lambda(t_1), \dots, \Lambda(t_n)\}$ forment un processus de Poisson d'intensité 1.

Ainsi, pour évaluer l'ajustement d'un processus de Hawkes, il convient de vérifier si l'ensemble $\Lambda(t_2) - \Lambda(t_1), \dots, \Lambda(t_n) - \Lambda(t_{n-1})$ suit une distribution exponentielle de paramètre 1. Cette vérification peut être réalisée à l'aide d'un diagramme quantile-quantile (Q-Q plot). La figure 2.22 montre les résultats obtenus pour le modèle (2.22) sur la période de calibrage.

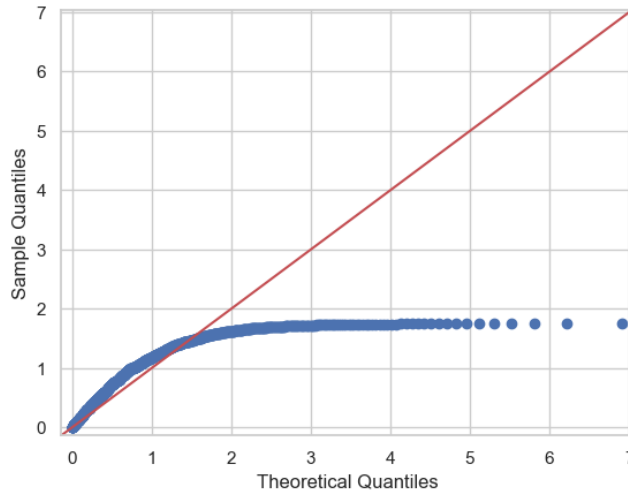


FIGURE 2.22 : Q-Q plot modèle (2.22)

Les résultats obtenus à partir de ce Q-Q plot ne sont pas concluants. Les points ne sont pas alignés sur la droite rouge $x = y$, ce qui suggère un mauvais ajustement du modèle. Bien que les points soient bien alignés au début, on observe une divergence à partir de la valeur $x = 1.8$ environ, où les points dévient sous la courbe rouge. Cela signifie que les grandes valeurs dans notre échantillon sont plus fréquentes que ce qu'une distribution exponentielle prévoit. En d'autres termes, les intervalles entre les événements sont plus longs dans les données que ceux attendus par le modèle de Hawkes.

Cette observation peut s'expliquer par le fait que les défaillances ont été réparties uniformément au sein de chaque mois, ce qui introduit un biais dans le test, rendant ce résultat attendu. Toutefois, elle pourrait également indiquer que l'effet d'auto-excitation observé dans les données est moins prononcé que ce que le modèle prédit, ou encore que le modèle utilisé n'est pas parfaitement adapté à la structure des données.

Étant donné que ce test repose sur les temps de sauts, il ne permet pas d'évaluer pleinement l'ajustement du modèle à nos données. Cependant, compte tenu des résultats obtenus sur les prédictions, cela ne remet pas en cause la suite de nos travaux.

2.3.3 Analyse de sensibilité

Cette section vise à étudier la robustesse des paramètres calibrés. Nous nous intéressons principalement au paramètre α qui reflète les interdépendances entre les différents secteurs d'activité. Nous avons observé une différence notable entre les paramètres estimés en fixant les paramètres initiaux avec l'approche statistique et l'approche économique. La figure 2.23 illustre les valeurs obtenues avec ces deux méthodes.

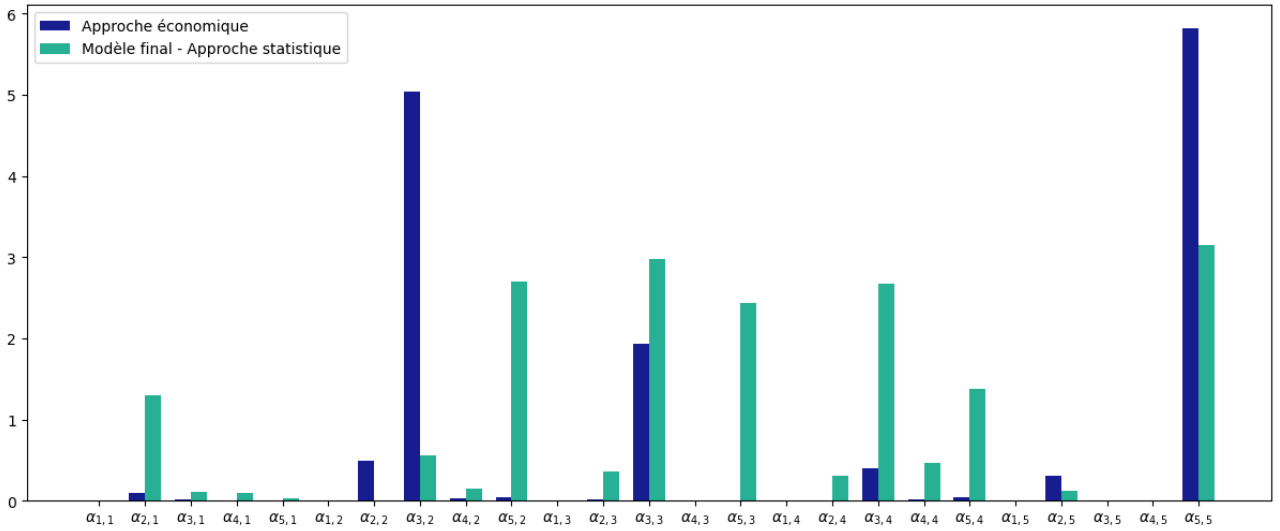


FIGURE 2.23 : Valeurs du paramètre α issus des calibrages avec l'approche statistique et économique

Pour évaluer la robustesse des paramètres (2.27), nous procéderons à des calibrages en partant du même point de départ que le modèle final, tout en faisant varier un seul paramètre à la fois. L'objectif est d'examiner si les valeurs convergent vers les mêmes résultats. Étant donné le temps de calibrage relativement long, nous ne pouvons pas tester une variation sur l'ensemble des paramètres. Nous avons effectué les variations suivantes :

- $\alpha_{1,1}$: 0.0011 \rightarrow 0.8 ;
- $\alpha_{2,1}$: 0.8739 \rightarrow 0.10. ;
- $\alpha_{3,4}$: 2.3035 \rightarrow 1.0 ;
- $\alpha_{3,5}$: 2.4741 \rightarrow 4.0 ;
- $\alpha_{5,2}$: 1.8179 \rightarrow 0.2 ;
- $\alpha_{5,5}$: 3.1254 \rightarrow 1.5 ;

Pour les paramètres $\alpha_{1,1}$ et $\alpha_{2,1}$, l'objectif est de déterminer si, en partant d'une valeur initiale éloignée de 0, le paramètre aura quand même tendance à converger vers 0. Les paramètres $\alpha_{3,4}$, $\alpha_{5,2}$

et $\alpha_{5,5}$, qui sont relativement élevés dans le modèle final, seront testés avec des valeurs initiales plus faibles afin d'observer si les paramètres calibrés le seront également. Enfin, pour le paramètre $\alpha_{3,5}$, qui est estimé à une valeur presque nulle malgré une corrélation assez élevée entre les deux secteurs, nous testons si en fixant une valeur initiale plus élevée, le calibrage aboutit à une valeur nulle. Les résultats des calibrages sont présentés dans la figure 2.24.

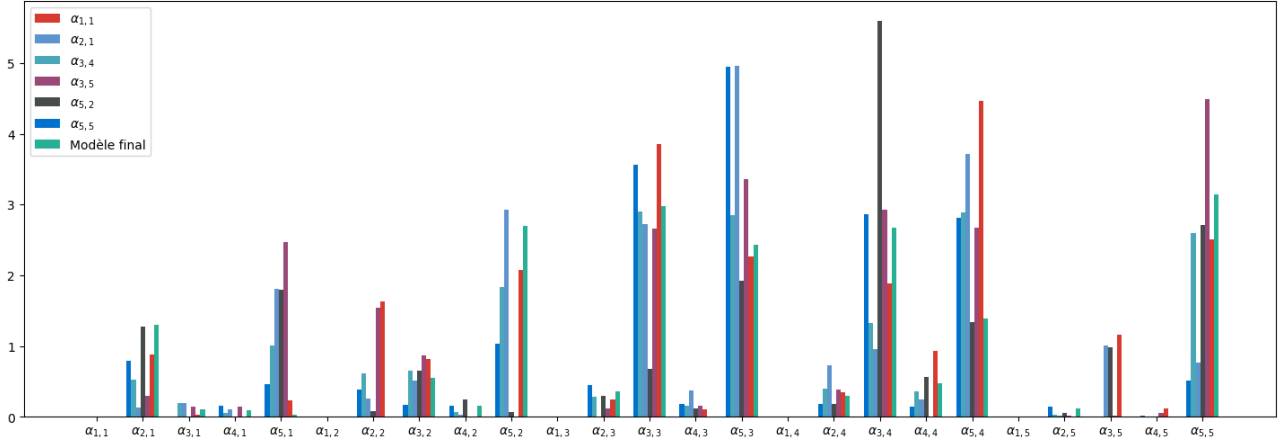


FIGURE 2.24 : Valeurs du paramètre α issus des calibrages

Les paramètres obtenus présentent des variations significatives d'un calibrage à l'autre. Même une légère variation sur le choix du point de départ peut impacter significativement l'estimation des paramètres.

Cependant, une certaine convergence émerge pour certains paramètres, qui sont systématiquement classés comme nuls ou non nuls après calibrage. Malgré cela, l'ordre de grandeur des paramètres estimés varie considérablement d'un calibrage à l'autre. Pour la suite, nous conservons les paramètres (2.27), car ils offrent la vraisemblance la plus élevée.

2.3.4 Limites

Bien que les résultats obtenus soient prometteurs, plusieurs limites subsistent, notamment en termes de temps de calcul, tant pour le calibrage des paramètres que pour les simulations. Le temps de calcul augmente avec la dimension du processus multivarié, ce qui pourrait constituer un obstacle à l'application de ce modèle en assurance-crédit. Dans nos travaux, en regroupant les défauts en cinq secteurs d'activités, nous avons observé un temps de calibrage du modèle proche de 15 heures, et un temps de simulation d'environ 5 minutes. Dans un portefeuille d'assurance-crédit réel, avec une segmentation plus fine des défauts en groupes homogènes et probablement plus de cinq segments, la complexité algorithmique et les temps de calcul seraient bien plus élevés. Ce coût computationnel semble également fortement influencé par le nombre d'événements modélisés, chaque événement contribuant à augmenter le temps de calcul. À titre comparatif, le tableau 2.10 détaille le temps d'exécution des différents modèles.

Par ailleurs, lors du calibrage des processus univariés, nous observons une convergence stable des paramètres, indépendamment des valeurs initiales. En multivarié, cependant, le modèle montre une certaine instabilité lors du calibrage : les solutions obtenues varient selon les valeurs initiales choisies, bien que les résultats des simulations demeurent similaires. Cela suggère que, bien que les valeurs des paramètres diffèrent d'une solution à l'autre, leurs effets s'équilibrent, produisant ainsi des prédictions proches. Cette observation interroge sur la robustesse des paramètres en multivarié.

Modèle	Calibrage (2010-2019)	Prédiction (n = 100 sim, T = 6 mois)
Hawkes (2.22)	$\simeq 15\text{h}$	$\simeq 5\text{min}$
Hawkes (2.18)	$\simeq 15\text{h}$	$\simeq 5\text{min}$
Poisson	Immédiat	Immédiat
Hawkes univarié (cumul 5 secteurs)	$\simeq 15\text{min}$	$\simeq 3\text{min}$

TABLE 2.10 : Temps d'exécution des différents modèle

Enfin, bien que le modèle offre des prédictions relativement précises, il repose sur une estimation de la fonction d'intensité de base $\mu(\cdot)$, qui est encore améliorable. L'intégration d'indicateurs économiques plus spécifiques à chaque secteur d'activité affinerait cette fonction, et pourrait considérablement améliorer la qualité des prédictions. De plus, des données précisant les moments exacts des sauts de défaillance pourraient également parfaire la précision globale du modèle.

Malgré ces limites, le modèle de Hawkes présente des avantages notables. En intégrant en temps réel les variations de l'environnement économique dans les prévisions, il offre des erreurs de prédiction relativement faibles, malgré le biais dans nos données et les possibilités d'optimisation pour la fonction $\mu(\cdot)$. Ces éléments laissent entrevoir un potentiel d'amélioration et d'application en assurance-crédit, à condition d'affiner les paramètres à partir de données réelles du secteur.

Chapitre 3

Impact du contexte économique sur les pertes d'un assureur-crédit

Ce dernier chapitre vise à étudier l'impact de la conjoncture économique, modélisée à travers les processus de Hawkes, sur les pertes subies par un assureur-crédit fictif. Nous commencerons par définir le modèle employé pour estimer les pertes de l'assureur ainsi que la méthodologie mise en oeuvre pour construire les données fictives. Ensuite, nous chercherons à mesurer l'intensité et la durée de l'effet de rattrapage, déjà en cours, lié à la crise du Covid-19 sur le nombre de défaillances, ainsi que l'impact prévisionnel pour l'assureur. Enfin, nous examinerons l'impact de l'allocation de l'exposition dans les différents secteurs d'activités sur les pertes de l'assureur, tout en tenant compte de l'évolution de la conjoncture économique.

3.1 Modélisation des pertes attendues

Cette section propose un modèle destiné à estimer les pertes attendues d'un assureur-crédit fictif. L'objectif est d'expliquer comment les travaux réalisés sur les processus de Hawkes peuvent être utilisés pour le pilotage des risques en assurance-crédit, en particulier dans un contexte économique incertain. Pour rappel, notre modélisation se concentre uniquement sur les pertes attendues associées aux contrats dénommés, sans aborder les pertes liées aux contrats non dénommés. De plus, compte tenu des données disponibles publiquement, nous avons uniquement modélisé les défauts résultant de l'insolvabilité totale des entreprises, ce qui exclut les défauts de type *Protracted*.

3.1.1 Présentation du modèle

Pour définir notre modèle, nous adoptons une approche *fréquence-sévérité*, couramment utilisée en assurance non-vie. Cette méthode consiste à modéliser la distribution des pertes attendues comme le produit d'une loi qui modélise la fréquence des événements et d'une loi qui modélise la sévérité des sinistres.

Bien que cette approche ne soit pas nécessairement la plus adaptée à l'assurance-crédit, puisque le coût d'un sinistre dépend fortement des factures en circulation au moment du défaut d'une entreprise et de la limite accordée par l'assureur. La sévérité d'un défaut peut varier donc selon l'instant de défaut. Elle présente néanmoins certains avantages.

D'une part, cette méthode est simple à mettre en oeuvre et permet d'obtenir des résultats clairs et intuitifs, facilitant ainsi leur analyse. D'autre part, elle nous permet de contourner la contrainte liée

à l'absence de données précises sur les expositions réelles d'un assureur-crédit. Nous permettant donc de traduire l'évolution du nombre de défauts en pertes financières pour l'assureur fictif.

Remarque 3.1. Il convient de noter que les processus de Hawkes auraient également pu être utilisés pour estimer une probabilité de défaut sur des segments homogènes d'acheteurs ayant le même profil de risque. Dans ce cadre, la perte pourrait alors être modélisée selon l'approche classique du risque de crédit, par l'intermédiaire de la formule (1.1).

Nous passons maintenant à la formalisation du modèle théorique. Celui-ci repose sur l'utilisation d'un processus de Hawkes, tel que défini en (2.22), pour modéliser la fréquence des sinistres. Cela permet de tenir compte de l'évolution dynamique des défauts au sein des différents secteurs d'activité, en fonction du contexte économique. Nous détaillons ci-après la modélisation de la perte de l'assureur sur une période donnée.

Modèle théorique

La perte de l'assureur sur la période $[0, t]$ s'exprime de la manière suivante

$$L_t = \sum_{k=1}^d L_t^{(k)}, \quad (3.1)$$

où $d = 5$ est le nombre de secteurs et $L_t^{(k)}$ est la perte de l'assureur sur $[0, t]$ associée aux défauts des acheteurs du secteur k .

On note :

- $t_0 = 0, t_1 = 1, \dots, t_N = \lfloor t \rfloor$ et $t_{N+1} = t$;
- $(N_t^{(k)})_{(t \geq 0)}$ est le processus de Hawkes associé au secteur k , modélisant la fréquence des défauts dans ce secteur ;
- $X_{k,j}(t_i)$ représente la perte individuelle issue du défaut j dans le secteur k à la date t_i ;
- $E_{k,j}$ correspond à l'exposition totale associée à l'acheteur j du secteur k , garantie par l'assureur ;
- $\sigma_k(t_i)$ est le taux de perte à la date t_i associé aux sinistres dans le secteur k ;

et pour chaque secteur k , la perte $L_t^{(k)}$ se calcule comme suit

$$L_t^{(k)} = \sum_{i=0}^N \sum_{j=N_{t_i}^{(k)}}^{N_{t_{i+1}}^{(k)}} X_{k,j}(t_i) \quad \text{avec} \quad X_{k,j}(t_i) = E_{k,j} \times \sigma_k(t_i). \quad (3.2)$$

En général, les coûts des sinistres $X_{k,j}$ sont supposés indépendants du nombre de sinistres $N^{(k)}$. Cependant, cette hypothèse est discutable dans le contexte de l'assurance-crédit. Il semble en effet intuitif que le taux de destruction $\sigma_k(t)$ soit potentiellement corrélé à la fréquence des défauts. Un nombre élevé de défauts à un instant t pourrait être le reflet d'un environnement économique défavorable, ce qui pourrait impliquer une réduction des montants recouvrés par l'assureur et entraînerait une hausse des pertes associées. À l'inverse, il est également possible que l'assureur, anticipant une telle dégradation économique, ait déjà réduit les limites de couverture, limitant ainsi les pertes potentielles.

Le modèle (3.2) permet d'introduire une dépendance entre la fréquence et la sévérité à travers le paramètre $\sigma_k(\cdot)$. Ce paramètre peut être vu comme le produit entre l'*UGD* et le taux *LGD* présentés dans

le premier chapitre de ce mémoire. Une manière simple de définir $\sigma_k(\cdot)$ serait de le considérer comme une fonction à valeurs discrètes, correspondant à différentes situations économiques. Par exemple, un scénario pourrait distinguer deux états : *crise économique* et *stabilité économique*. Le passage d'un état à l'autre serait déclenché par l'atteinte d'un seuil. Ce seuil pourrait être défini de plusieurs façons, comme par exemple en fonction de l'écart entre le nombre de défauts observés à deux dates proches ou encore en fonction de la valeur de la fonction d'intensité de base $\mu_k(\cdot)$ décrite dans le chapitre précédent.

Dans ce modèle, les expositions individuelles $E_{k,j}$, associées aux acheteurs d'un secteur k , sont supposées être identiquement distribuées selon une loi E_k , représentant le coût d'un sinistre dans ce secteur. La variance de cette loi doit refléter les disparités entre les niveaux d'exposition au sein du secteur, ces disparités étant liées aux différentes limites de couverture accordées par l'assureur, ainsi qu'à l'utilisation de ces limites par les assurés. Afin de tenir compte du fait que l'exposition ne peut dépasser la limite accordée par l'assureur, il pourrait être pertinent d'utiliser une distribution tronquée pour E_k .

La mise en œuvre du modèle (3.2) repose sur l'accès à des données précises concernant les expositions $E_{k,j}$ et les taux de recouvrement $\sigma_k(t)$ de l'assureur sur les débiteurs. Cependant, en l'absence de données spécifiques pour calibrer de manière rigoureuse ces paramètres, il est nécessaire de simplifier le modèle afin de contourner ces limitations. Nous avons ainsi choisi de supposer que le taux de recouvrement $\sigma_k(t)$ est constant dans le temps et d'assigner une valeur fixe \bar{E}_k aux expositions individuelles $E_{k,j}$, propre à chaque secteur mais identique pour tous les acheteurs d'un même secteur. Ces simplifications permettent de conserver un modèle fonctionnel et applicable à nos données. Toutes choses étant égales par ailleurs, cela nous permettra de mesurer l'impact économique sur les pertes, à travers l'évolution de la fréquence des défauts, constituant désormais l'unique paramètre stochastique du modèle.

Modèle simplifié

Dans la formule (3.2), nous simplifions l'exposition $E_{k,j}$ en la remplaçant par une valeur moyenne \bar{E}_k pour chaque secteur, et nous considérons que la fonction de recouvrement $\sigma_k(t)$ devient un paramètre constant σ .

Ainsi la perte $L_t^{(k)}$ dans le secteur k s'écrit

$$L_t^{(k)} = \sum_{i=0}^{N_t^{(k)}} E_k \times \sigma = N_t^{(k)} \times E_k \times \sigma, \quad (3.3)$$

et la perte attendue sur la période $[0, t]$ se calcule par la formule suivante

$$L_t = \sum_{k=1}^d N_t^{(k)} \times \bar{E}_k \times \sigma \quad (3.4)$$

où :

- $(N_t^{(k)})_{(t \geq 0)}$ est le processus de Hawkes associé au secteur k ;
- \bar{E}_k correspond à l'exposition moyenne garantie par l'assureur dans le secteur k ;
- σ est le taux de perte moyen des sinistres.

Pour déterminer la valeur du paramètre σ , nous nous référons à une étude publiée par l'ICISA et VERSICHERUNGSFOREN LEIPZIG (2023). Cette étude, réalisée en collaboration avec la plupart des entreprises d'assurance-crédit, se base sur des données couvrant 75% des primes du marché de l'assurance-crédit, offrant ainsi un aperçu représentatif du marché. L'étude en question se concentre sur le taux de perte LGD , qui diffère légèrement de celui défini à la remarque 1.2, car il inclut également l' UGD dans sa définition. Il peut donc s'apparenter au paramètre σ de notre modèle. L'étude porte sur les risques pour lesquels l'exposition des assureurs est élevée, et couvre la période d'observation de 2017 à 2019.

D'après les résultats de cette étude, le paramètre LGD se situe dans l'intervalle $[5.8\%, 7.8\%]$ avec un niveau de confiance de 95%. Notre modèle ne se limite pas nécessairement à des expositions élevées et partant du principe que plus l'exposition est élevée, moins l'entreprise est susceptible de faire faillite, la LGD diminue. En effet, si l'assureur a accepté de s'exposer à un tel niveau de risque, cela signifie qu'il estime que l'entreprise est solide financièrement. Cependant, notre modèle se concentre uniquement sur les défauts résultant de l'insolvabilité de l'acheteur, ce qui tend à augmenter la LGD , car l'assureur rencontrera davantage de difficultés à recouvrer les impayés. Partant de l'hypothèse que ces deux effets se compensent, nous choisissons de fixer σ à la moyenne de l'intervalle, soit 6.8%, afin de modéliser la perte associée aux défaillances d'un portefeuille d'assureur-crédit ayant des niveaux d'exposition divers.

L'exposition moyenne \bar{E}_k pour chaque secteur k est obtenue en réalisant le calcul suivant

$$\bar{E}_k = \frac{E_k}{n_k},$$

où E_k représente l'exposition totale de l'assureur fictif dans le secteur k et n_k est le nombre total d'acheteurs susceptibles de faire défaut dans le secteur k .

Le processus $(N_t^{(k)})_{(t \geq 0)}$ enregistre l'ensemble des défauts survenus dans le secteur k en France. Cependant, toutes les défaillances d'entreprise ne sont pas forcément liées au portefeuille d'assurés de notre assureur fictif, et n'impactent pas forcément la perte potentielle de ce dernier. Pour identifier les défauts qui ont un impact sur notre assureur fictif, nous introduisons un coefficient p_k . Ce coefficient représente la probabilité que l'entreprise en défaut soit un acheteur du portefeuille d'assurés de l'assureur. En multipliant le nombre total de défauts survenus par p_k , nous obtenons une estimation des défauts qui vont affecter notre assureur fictif. Ces défauts sont notés $(\tilde{N}_t^{(k)})_{(t \geq 0)}$ et sont calculés grâce à la relation suivante

$$\tilde{N}_t^{(k)} = N_t^{(k)} \times p_k.$$

Ce modèle est assez simplifié et présente donc certaines limites. L'emploi d'un paramètre σ constant et le passage d'une exposition individuelle à une exposition moyenne sur l'ensemble d'un secteur peuvent introduire des biais, notamment dans les secteurs avec des tranches d'encours variées et des profils de risque non homogènes. Néanmoins, ce modèle fournit un outil exploitable avec des données publiques pour obtenir une estimation approximative des pertes attendues d'un assureur-crédit fictif et pour évaluer l'impact de l'environnement économique.

3.1.2 Construction des données fictives

Cette section a pour but de présenter la méthodologie employée pour construire nos données d'exposition fictives. L'objectif est de reproduire une exposition au risque comparable à celle d'un des trois acteurs principaux du marché de l'assurance crédit. D'une part pour être en mesure de contrôler les résultats du modèle en utilisant les données de défaillances des années passées et d'autre part, afin de

disposer de données aussi proches que possible de celles d'un véritable assureur-crédit. Pour ce faire, nous avons collecté diverses informations publiées par l'assureur Coface dans ses rapports annuels.

Les principales informations extraites des rapports annuels de Coface sont les suivantes :

- la part de marché de Coface sur le marché mondial de l'assurance-crédit : $\delta_{\text{Marché}}^{\text{Coface, Monde}}$;
- l'exposition totale de Coface à l'échelle mondiale : $E^{\text{Coface, Monde}}$;
- la proportion de cette exposition liée à des acheteurs en France : $\delta_{\text{Expo, Coface}}^{\text{France}}$;
- l'encours moyen par débiteur du portefeuille global de Coface : $\bar{E}^{\text{Coface, Monde}}$;
- le nombre total de débiteurs associés aux assurés de Coface dans le monde : $n^{\text{Coface, Monde}}$;
- une approximation de l'allocation sectorielle de l'exposition totale de Coface au niveau mondial : $\theta^{\text{Coface, Monde}} \in \mathbb{R}^d$.

Le tableau 3.1 présente la répartition de l'exposition de Coface dans nos cinq secteurs d'activité pour l'année 2021.

Secteur d'activité	$E_k^{\text{Coface, Monde}}$	$\theta_k^{\text{Coface, Monde}}$
(A)	94 500 000 000	0.1608
(I)	199 000 000 000	0.3387
(C)	65 500 000 000	0.1105
(S)	211 100 000 000	0.3593
(E)	17 500 000 000	0.0298
Total	587 600 000 000	1.000

TABLE 3.1 : Données d'exposition de Coface en 2021

Outre les informations provenant des rapports annuels de Coface, nous avons également recueilli des données sur le stock d'unités légales par secteurs d'activité en France 2014 à 2021, ainsi que sur le chiffre d'affaires par secteurs d'activité de 2014 à 2022 à partir des bases de données de l'INSEE. Nous avons également récupéré le taux de pénétration de l'assurance-crédit dans le monde, noté τ_p , provenant des études de l'ICISA (International Credit Insurance & Surety Association) sur le marché mondial de l'assurance-crédit.

Ces informations sont très globales et ne permettent pas une application directe du modèle (3.4). Il nous manque des données spécifiques au marché français ainsi que sur le nombre de débiteurs par secteur d'activité pour estimer l'exposition moyenne d'un assureur en France dans chaque secteur. Cependant, ces données globales nous fournissent un bon point de départ pour construire les informations nécessaires au calcul de l'exposition moyenne par secteur de notre assureur-crédit fictif.

Afin de calculer les encours E_k de notre assureur fictif en France pour le secteur d'activité k , nous estimons l'allocation sectorielle de l'exposition de Coface en France par

$$\hat{\theta}^{\text{Coface, France}} = \theta^{\text{Coface, Monde}}.$$

En d'autres termes, nous supposons que la répartition sectorielle de l'exposition totale de Coface en France est identique à celle de son exposition totale à l'échelle mondiale. Par conséquent, l'exposition par secteur est calculée de la manière suivante

$$E_k = E_k^{\text{Coface, Monde}} \times \delta_{\text{Expo, Coface}}^{\text{France}},$$

où $E_k^{\text{Coface, Monde}}$ représente l'exposition totale de Coface dans le secteur k et $\delta_{\text{Expo, Coface}}^{\text{France}}$ désigne la part de l'exposition totale de Coface en France.

Avant de déterminer le nombre n_k de débiteurs par secteurs d'activité dans le portefeuille de notre assureur fictif, il est nécessaire d'établir certaines variables. Le chiffre d'affaires moyen par unité légale dans le secteur k en France $\bar{CA}_k^{\text{France}}$ est calculé comme suit

$$\bar{CA}_k^{\text{France}} = \frac{CA_k^{\text{France}}}{n_k^{\text{France}}},$$

où CA_k^{France} est le chiffre d'affaires total du secteur k en France, et n_k^{France} est le nombre total d'unités légales dans ce secteur.

Par ailleurs, nous supposons que la part de marché de Coface en France est égale à sa part de marché dans le monde, ce qui se traduit par l'équation suivante

$$\hat{\delta}_{\text{Marché}}^{\text{Coface, France}} = \delta_{\text{Marché}}^{\text{Coface, Monde}}.$$

De même, nous considérons que le taux de pénétration de l'assurance-crédit est comparable à celui du marché mondial, soit

$$\hat{\tau}_p^{\text{France}} = \tau_p.$$

En conséquence, le chiffre d'affaires total « assuré » par Coface en France, noté $CA^{\text{Coface, France}}$, est estimé par la formule suivante

$$\widehat{CA}^{\text{Coface, France}} = CA^{\text{France}} \times \hat{\delta}_{\text{Marché}}^{\text{Coface, France}} \times \hat{\tau}_p^{\text{France}},$$

où CA^{France} est le chiffre d'affaires total des entreprises en France.

Le chiffre d'affaire « assuré » par Coface en France dans chaque secteur k est donné par le calcul suivant

$$\widehat{CA}_k^{\text{Coface, France}} = \widehat{CA}^{\text{Coface, France}} \times \frac{E_k^{\text{Coface, Monde}}}{E^{\text{Coface, Monde}}} = \widehat{CA}^{\text{Coface, France}} \times \theta_k^{\text{Coface, Monde}},$$

où $\theta^{\text{Coface, Monde}} \in \mathbb{R}^d$ est l'allocation en secteurs d'activité de l'exposition totale de Coface.

Enfin, le nombre de débiteurs du portefeuille de notre assureur fictif dans le secteur k en France est obtenu à l'aide de la formule suivante

$$n_k = \hat{n}_k^{\text{Coface, France}} = \frac{\bar{CA}_k^{\text{France}}}{\widehat{CA}_k^{\text{Coface, France}}}.$$

Le coefficient p_k , qui représente la proportion d'entreprises du secteur k dont le défaut aura un impact sur notre assureur, est déterminé comme suit

$$p_k = \frac{n_k}{n_k^{\text{France}}},$$

où n_k^{France} est le nombre d'unités légales dans le secteur k en France.

Les données d'exposition fictives ont été construites selon cette méthodologie pour les années 2021, 2022 et 2023. Nous avons remarqué que les variations d'une année à l'autre dans la répartition sectorielle, les volumes d'exposition, le nombre de débiteurs ainsi que les parts de marché étaient très faibles pour les années 2021 et 2022. Ces évolutions sont également restées marginales entre 2022 et 2023. C'est pourquoi nous avons décidé de conserver l'exposition de 2023 pour l'année 2024, estimant qu'elle reflète toujours fidèlement la situation du marché. Le tableau 3.2 illustre les résultats obtenus pour l'année 2023 que nous conservons pour l'année 2024.

Secteur d'activité	n_k	E_k	\bar{E}_k	p_k
(A)	71 521	11 564 500 000	161 693.76	0.1842
(I)	10 583	23 533 000 000	2 223 665.74	0.0297
(C)	19 574	7 423 500 000	379 256.67	0.0298
(S)	40 843	24 502 600 000	599 916.16	0.0117
(E)	33 116	2 171 500 000	65 573.46	0.0032
Total	175 637	69 195 100 000	393 967.08	0.0278

TABLE 3.2 : Données d'exposition de notre assureur fictif en 2024

Afin de nous assurer que nos données fictives d'exposition sont suffisamment cohérentes pour être exploitées dans notre modèle, nous avons réalisé plusieurs vérifications. Ces contrôles visent à vérifier que les pertes estimées sont comparables à celles d'un véritable assureur-crédit. Ils seront détaillés dans la section suivante.

3.1.3 Backtesting

L'objectif de cette section est de détailler les contrôles effectués sur les données construites afin de garantir qu'elles sont suffisamment représentatives de la réalité. Ces contrôles nous permettent de vérifier si les données sont adaptées à une analyse transposable à des données réelles. Nous mettrons également en évidence les biais identifiés, qui devront être pris en compte lors de l'interprétation des résultats. En particulier, nous chercherons à évaluer si nos données permettent de reproduire de manière cohérente la perte d'un assureur fictif sur les années passées. Ces pertes doivent être comparables à celles de Coface, dont nous avons tenté de reproduire l'exposition.

Nous concentrons notre analyse sur l'année 2021, car c'est la dernière année pour laquelle nous disposons d'une vision complète des unités légales par secteur d'activité en France. Toutefois, des contrôles similaires ont été effectués pour les années 2022 et 2023, en appliquant un taux de croissance moyen des unités légales dans chaque secteur, basé sur l'évolution des cinq dernières années.

Un premier contrôle de cohérence a été réalisé sur le calcul du nombre d'unités n_k à travers le chiffre d'affaires. Cette variable est sujette aux approximations les plus importantes, car elle résulte de la combinaison de plusieurs variables elles-mêmes approximées. Nous avons comparé le nombre total obtenu à celui qui aurait été calculé en multipliant le nombre total de débiteurs de Coface dans le monde par ses parts de marché. Le ratio ainsi obtenu est de 70%, ce qui semble raisonnable dans la mesure où la France est un marché développé. Il est donc plausible que l'exposition sur les entreprises

françaises soit plus élevée que celle sur les entreprises dans les marchés émergents. De ce fait, l'encours moyen calculé en France est plus élevé que celui rapporté par Coface dans son rapport, ce qui est cohérent avec notre hypothèse.

En appliquant la méthode présentée dans la section précédente, nous obtenons une part de 2,95% d'entreprises françaises assurées, ce qui est très proche des 2,18% obtenus en multipliant les parts de marché de Coface par le taux de pénétration de l'assurance-crédit, qui correspondrait à la part des entreprises du monde assurées par Coface.

Cependant, nous avons relevé que les encours moyens par secteur d'activité paraissent élevés, en particulier au regard des données de la Banque de France. Celle-ci indique que la quasi-totalité des défaillances concerne des PME, tandis que les rapports de Coface indiquent que les PME sont les entreprises les moins assurées. De plus, les entreprises tendent à entretenir des relations commerciales avec des entreprises de taille comparable, ce qui rend peu réaliste l'idée que les grandes entreprises aient des transactions commerciales importantes avec les PME. Ce dernier point pourrait cependant expliquer une partie des écarts observés, mais il est plus probable que ces écarts soient dus à la volatilité des encours assurés par notre assureur. Le fait de supposer un encours moyen par secteur constitue en soi une approximation importante.

Malgré cela, lorsque nous calculons les pertes à partir de nos données, celles-ci sont cohérentes avec les charges publiées par Coface. Dans ULLMO (2019), il est indiqué que sur les années 2006-2014, 74% de la charge de Coface provenait de sinistres dont l'acheteur était en insolvabilité. C'est ce type de défaut que nous modélisons, nous comparerons donc nos résultats avec 74% de la charge Coface en France. Cette dernière est calculée en appliquant le ratio des parts de marché à la charge mondiale publiée dans chaque rapport annuel. Le tableau 3.3 présente les résultats obtenus, ainsi que le ratio entre la charge réelle observée et celle modélisée. Il convient de préciser que, pour la fréquence des sinistres, nous nous appuyons sur des données réelles, ce qui élimine tout biais potentiel de ce côté.

Année	\hat{L}^{Coface}	$\hat{L}^{\text{Modèle}}$	$\frac{\hat{L}^{\text{Coface}}}{\hat{L}^{\text{Modèle}}}$
2021	20 961 281	21 709 316	0.9655
2022	35 281 646	35 021 158	1.0074
2023	41 753 053	46 502 846	0.8979

TABLE 3.3 : Backtesting du modèle (3.4)

Nous considérons ces résultats satisfaisants compte tenu des approximations faites pour construire les données. La proximité entre les pertes modélisées et les pertes réelles observées valide également le choix du paramètre σ .

Les données publiques disponibles nous ont permis de construire des données d'exposition reflétant de manière très approximative les expositions d'un véritable assureur-crédit en France. Bien qu'imparfaites, elles permettent de retrouver des résultats proches des pertes réellement observées et offrent ainsi une base suffisante pour mener à bien notre analyse. Par la suite, pour évaluer l'impact de la conjoncture économique sur les pertes de notre assureur-crédit fictif, nous fixerons les données d'exposition pour l'année 2024, et nous ferons varier uniquement le nombre de défauts dans nos prévisions. Bien que les données soient construites à partir de proxys et que le modèle soit très simplifié, il permettra tout de même de convertir ces variations en pertes pour l'assureur et d'obtenir une estimation de l'impact financier potentiel.

3.1.4 Prévisions pour l'année 2024

La figure 3.1 présente la distribution de la perte attendue pour notre assureur-crédit fictif pour l'année 2024. Il convient de rappeler que la seule composante stochastique du modèle (3.4) est la fréquence des défaillances, laquelle est modélisée par un processus de Hawkes multivarié. Les résultats présentés sont issus de 500 simulations du modèle (2.22).

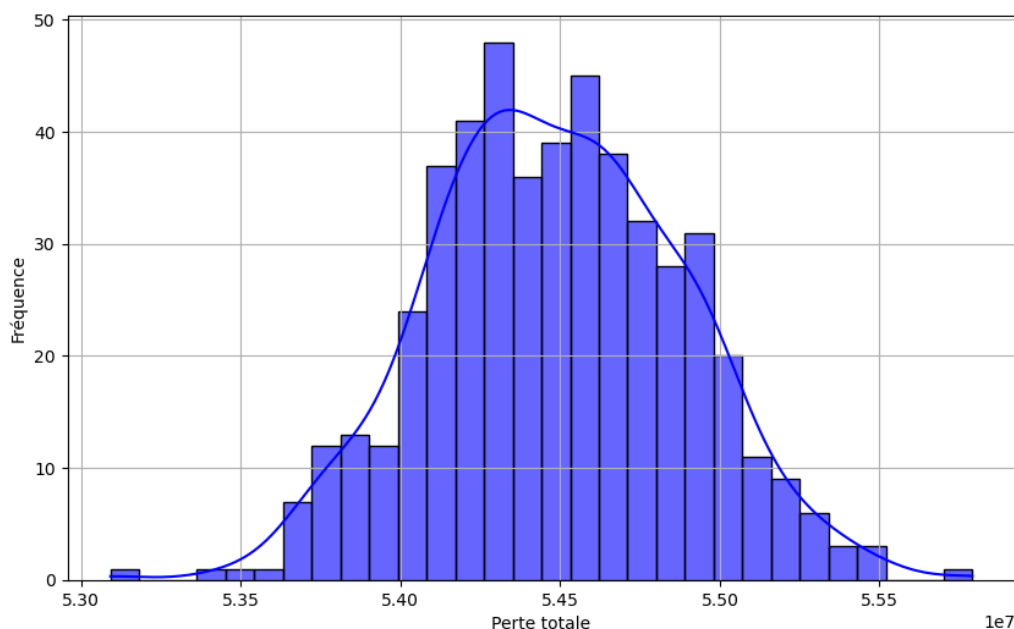


FIGURE 3.1 : Distribution de la perte attendue en 2024

Selon les prévisions du modèle, la perte attendue moyenne pour l'année 2024 s'élève à 54 485 592 €, avec un intervalle de confiance à 95% s'étendant de 53 736 596 € à 55 250 938 €. Cette estimation représente une augmentation de 17.2% par rapport à l'année 2023. Cet écart résulte exclusivement de l'évolution des défaillances, toutes choses étant égales par ailleurs.

En outre, si l'on considère le montant de la perte attendue dans un scénario où le niveau de défaillances serait resté proche de celui observé avant la pandémie de Covid-19, la perte moyenne aurait été de 49 840 180 €, ce qui correspond à un écart relatif de 9.3%.

3.2 Mesure de l'intensité de l'effet de rattrapage de la crise Covid

Notre objectif ici est de quantifier l'effet de rattrapage sur les défaillances d'entreprises attendu et déjà observé avec la fin des aides gouvernementales mises en place pendant la crise sanitaire, suite à leur retrait progressif à la fin de l'année 2021. Pour ce faire, nous analyserons comment la dynamique des défaillances a évolué pendant et après la crise sanitaire.

Si les aides de l'État n'ont eu aucun impact durable, c'est-à-dire si elles ont simplement reporté des faillites inévitables sans réellement « sauver » les entreprises, l'effet de rattrapage pourrait ajouter environ 60 000 défaillances « manquantes » entre 2020 et 2021. Cette hypothèse, bien que pessimiste et probablement irréaliste, souligne un enjeu de notre étude, qui est de distinguer deux types d'entreprises. D'une part, il s'agit d'identifier celles qui auraient normalement fait faillite pendant la période Covid, mais qui ont été temporairement sauvées grâce aux aides, et, d'autre part, celles qui ont pu être

stabilisées durablement par ces mesures et qui demeureront viables même sans aides prolongées. Par ailleurs, l'environnement économique a radicalement changé entre la période d'avant Covid et celle qui lui a succédé, augmentant les risques conjoncturels pour les entreprises. Il sera donc complexe de distinguer les défauts liés aux fluctuations économiques de ceux qui relèvent de l'effet de rattrapage.

Le modèle de Hawkes, utilisé pour modéliser les défaillances, montre des prévisions qui divergent de la réalité pendant la période 2020-2022, en raison de l'impact des mesures de soutien de l'État, non intégrées dans le modèle. Nous allons tenter d'exploiter cette inadéquation entre prévisions et observations pour estimer le nombre de défauts évités grâce aux aides publiques. Toutefois, nous restons conscients de la difficulté à distinguer les défaillances issues de l'effet de rattrapage de celles attribuables aux fluctuations conjoncturelles, d'autant que ces effets peuvent s'influencer mutuellement.

Nous proposons ici une approche relativement simplifiée, visant principalement à fournir quelques éléments d'analyse sans recourir à des modélisations complexes. Une étude approfondie de cette problématique exigerait davantage de temps et de ressources.

La première étape consiste à mesurer le nombre de défauts attendus, compte tenu des conditions économiques, mais non survenus sur la période 2020-2021. Cela nous permettra d'estimer le nombre de défaillances susceptibles de survenir en raison de cet effet de rattrapage. Pour ce faire, nous calculons l'écart entre les prédictions du modèle de Hawkes entre 2020 et 2021 et les défauts réellement observés. Nous obtenons un total de 37 232 défaillances, un chiffre inférieur aux 60 000 défaillances initialement estimées en se basant sur l'écart entre la moyenne 2010-2019 et les défaillances observées.

Nous calculons ensuite, entre janvier 2022 et juin 2024, la proportion de sinistres qui ne sont pas une conséquence directe de l'environnement économique. Pour chaque secteur k nous calculons ainsi

$$\tilde{n}^{(k)} = (1 - \gamma_k)n^{(k)},$$

où γ_k est le paramètre d'échelle du modèle (2.22) et $n^{(k)}$ les prédictions du processus de Hawkes $N^{(k)}$. Certaines défaillances comptabilisées par $\tilde{n}^{(k)}$ peuvent néanmoins résulter de façon indirecte de la conjoncture économique, à travers les inter-dépendances entre entreprises. Nous obtenons un total de 74 047 défaillances pour la période s'étendant de janvier 2022 à juin 2024.

Parmi ces défauts, certains peuvent découler des inter-dépendances entre entreprises, tandis que d'autres résultent de l'effet de rattrapage lié à la crise Covid. Distinguer ces deux types de défaillances s'avère complexe. Par exemple, la difficulté de certaines entreprises à rembourser leur Prêt Garanti par l'État (PGE) pourrait affecter la trésorerie de leurs fournisseurs, qui, bien qu'ils n'aient pas rencontré de problèmes de liquidité auparavant, peuvent être fragilisés par ces défauts de paiements.

Nous proposons donc d'explorer différentes hypothèses concernant la proportion de ces défaillances, notée τ , que l'on pourrait attribuer à l'effet de rattrapage. En fonction de ces hypothèses, nous pourrions déterminer si cet effet de rattrapage est susceptible de se poursuivre au second semestre de l'année 2024 et durant l'année 2025. Cependant, il convient de souligner que considérer τ comme constant n'est pas nécessairement approprié. En effet, au début de l'année 2022, τ devrait être plus faible qu'à la fin de cette même année, en raison du temps nécessaire pour observer l'impact de l'arrêt des mesures de soutien. Une approche plus nuancée pourrait donc être envisagée pour mieux refléter l'évolution de la situation économique et des comportements des entreprises au fil du temps.

Si nous supposons que $\tau = 20\%$, cela signifie qu'environ 15 000 défaillances survenues entre janvier 2022 et juin 2024 seraient dues à un rattrapage des défaillances qui auraient dû se produire durant la période Covid. Dans ce cas, nous nous attendons à observer encore environ 22 000 défaillances, dont 4 500 entre juillet et décembre 2024.

En prenant $\tau = 20\%$, nous estimons qu'environ 15 000 défaillances survenues entre janvier 2022 et juin 2024 seraient dues à un rattrapage des défaillances qui auraient dû se produire durant la période

Covid. Dans ce cas, nous anticipons environ 22 000 défaillances supplémentaires, dont 4 500 entre juillet et décembre 2024.

Enfin, en adoptant l'hypothèse que $\tau = 40\%$, nous estimons qu'environ 29 000 défaillances survenues entre janvier 2022 et juin 2024 sont le résultat d'un rattrapage. Dans ce cas, nous nous attendons à environ 8 000 défaillances supplémentaires entre juillet et décembre 2024, ce qui indiquerait une stabilisation dès 2025.

Étant donné les prévisions de notre modèle pour la fin de l'année 2024, qui suggèrent une stabilisation du nombre de défaillances, cette dernière hypothèse nous semble la plus plausible. Nos prévisions avec le modèle (3.4) à partir de 2023 sont assez proches de ce qui a été observé, ce qui indique que les défaillances liées au "rattrapage" sont suffisamment présentes dans l'historique pour être intégrées par le modèle. Cependant, cela peut également signifier qu'à la fin de cet effet, le modèle pourrait avoir tendance à surestimer le nombre de défaillances.

3.3 Analyse des pertes selon les expositions sectorielles

L'objectif de cette section est d'analyser la répartition de l'exposition sectorielle de notre assureur fictif et d'explorer des pistes d'optimisation en fonction des prévisions de sinistralité à venir. L'étude se fera en maintenant un niveau constant d'exposition, ce qui implique de conserver à peu près le même volume de primes pour l'assureur. Nous commencerons par une analyse rétrospective sur l'année passée pour déterminer quelle aurait été l'allocation optimale des expositions sous différentes contraintes ainsi que l'impact de cette optimisation sur les pertes attendues. Ensuite, nous chercherons à déterminer l'allocation optimale dans le contexte actuel, avec pour objectif principal la minimisation des pertes attendues pour l'année 2024. L'idée principale est de voir comment, à exposition constante, l'assureur peut minimiser ses pertes potentielles.

3.3.1 Problème d'optimisation

Nous cherchons à résoudre le problème de minimisation suivant

$$\begin{cases} \min_{\theta \in \mathbb{R}^d} L_t(\theta), \\ s.c : \sum_{k=1}^d \theta_k = 1 \text{ et } \theta_k \geq 0, \forall k \end{cases} \quad (3.5)$$

où L_t , la perte de l'assureur sur la période $[0, t]$, est définie par le modèle simplifié (3.4) par

$$L_t = \sum_{k=1}^d L_t^{(k)} := \sum_{k=1}^d N_t^{(k)} \times p_k(\theta_k) \times \bar{E}_k \times \sigma.$$

En remarquant que

$$p_k(\theta_k) = \frac{n_k(\theta_k)}{n_k^{\text{France}}} = \frac{\widehat{\text{CA}}^{\text{Coface, France}}}{\text{CA}_k^{\text{France}}} \times \theta_k,$$

et en définissant

$$x_k := N_t^{(k)} \times \bar{E}_k \times \sigma \times \frac{\widehat{\text{CA}}^{\text{Coface, France}}}{\text{CA}_k^{\text{France}}},$$

pour simplifier les notations. Le problème (3.5) se réécrit

$$\begin{cases} \min_{\theta \in \mathbb{R}^d} \sum_{k=1}^d x_k \theta_k, \\ s.c : \sum_{k=1}^d \theta_k = 1 \text{ et } \theta_k \geq 0, \forall k \end{cases} \quad (3.6)$$

Le problème (3.6) est clairement un problème de programmation linéaire, c'est-à-dire un problème d'optimisation avec critère et contraintes affines. Selon un théorème démontré dans CIARLET (1990), la solution d'un tel problème est un sommet du simplexe définissant l'ensemble des contraintes : $\{\theta \in \mathbb{R}_+^d : \sum_{k=1}^d \theta_k = 1\}$. C'est-à-dire un des vecteurs de la base canonique de \mathbb{R}^d .

La solution du problème (3.6), consiste donc en une allocation totale de l'exposition dans le secteur où l'on observe les plus faibles taux de défaut et d'exposition moyenne en cas de sinistre. Cela ne paraît pas réaliste en pratique. Il est en effet peu probable qu'un assureur puisse modifier radicalement la répartition de ses expositions entre les différents secteurs. De plus, un assureur n'a pas nécessairement la capacité de concentrer toute son exposition sur un seul secteur, en raison de la concurrence avec d'autres assureurs et de la taille limitée du marché. Par exemple, allouer l'intégralité de l'exposition au secteur agricole pourrait aboutir à un nombre de débiteurs dans ce secteur supérieur au nombre total d'exploitations agricoles en France. Enfin, d'un point de vue de gestion des risques, il est généralement préférable de diversifier son portefeuille d'exposition. Une allocation trop concentrée augmente le risque de pertes importantes en cas de sinistralité dans ce secteur unique.

Pour la recherche d'une allocation d'exposition optimisée mais réaliste, nous allons ajouter des contraintes au problème (3.5). Le nouveau problème d'optimisation est le suivant

$$\begin{cases} \min_{\theta \in \mathbb{R}^d} \sum_{k=1}^d x_k \theta_k, \\ s.c : \sum_{k=1}^d \theta_k = 1, \\ |\theta_k - \theta_k^0| \leq s \text{ et } \theta_k \geq 0, \forall k, \end{cases} \quad (3.7)$$

où θ_k^0 est la part de l'exposition totale de l'assureur dans le secteur k avant optimisation et s est un seuil à définir.

Les contraintes

$$|\theta_k - \theta_k^0| \leq s, \forall k = 1, \dots, d,$$

permettent de garantir que l'allocation dans chaque secteur ne s'écarte pas de manière excessive de l'allocation actuelle. Cela assure que l'allocation dans un secteur respecte les contraintes du marché de l'assurance-crédit, tout en évitant des changements trop radicaux dans le portefeuille de contrats de l'assureur. Elle permet également de maintenir une diversification des risques similaire à celle existante. Cette contrainte trouve un équilibre entre la recherche de l'optimisation des pertes attendues et la préservation d'une allocation d'exposition cohérente avec les réalités du marché.

L'ajout de ces contraintes rend la solution du problème moins immédiate et, comme la plupart des cas de programmation linéaire, n'a pas de solution analytique directe. En effet, la présence de contraintes d'inégalité implique la possibilité d'avoir des contraintes dites « saturées », c'est-à-dire que la solution optimale atteint la borne de ces contraintes. Les contraintes saturées influencent directement la structure de la solution optimale. Elles fixent certaines des variables θ_k à des valeurs spécifiques

(comme 0 ou $\theta_k^0 \pm s$), réduisant ainsi la dimension de l'espace dans lequel l'optimisation se poursuit. Cela complique la résolution analytique, car il n'est pas possible d'ignorer ces contraintes actives sans risquer d'obtenir une solution non valide.

Pour résoudre ce type de problème, on peut utiliser l'algorithme *Coin-or Branch and Cut*. C'est l'un des solveurs les plus populaires pour ce type de problème.

3.3.2 Application sur la sinistralité passée de l'assureur fictif

Nous résolvons le problème d'optimisation (3.7) sur les données de sinistralité de l'assureur fictif pour l'année 2023. La table 3.4 présente les résultats obtenus avec un seuil s fixé à 5% puis à 1%.

	θ^0	$s = 0.05$		$s = 0.01$	
		θ	Δ	θ	Δ
(A)	0.1671	0.2171	0.05	0.1771	0.01
(I)	0.3401	0.2901	-0.05	0.3301	-0.01
(C)	0.1073	0.0573	-0.05	0.0973	-0.01
(S)	0.3541	0.3541	0.00	0.3541	0.00
(E)	0.0314	0.0814	0.05	0.0414	0.01
$\hat{L}^{\text{Modèle}}$	46 502 846	41 331 542	- 5 171 304	45 469 670	- 1 033 176

TABLE 3.4 : Allocation optimale pour l'année 2023

Les résultats obtenus dans la table 3.4 montrent que, quelle que soit la valeur du seuil s , l'allocation d'exposition optimale tend à accroître les parts dans le secteur agricole et les services non marchands, qui présentent les taux de défaut et les expositions moyennes en cas de sinistre les plus faibles. En revanche, l'allocation est réduite dans les secteurs où le rapport entre l'exposition et le nombre de défauts est moins favorable pour l'année 2023, notamment dans l'industrie et la construction.

Avec un seuil $s = 5\%$, cette redistribution aboutit à une réduction significative des pertes espérées, $\hat{L}^{\text{Modèle}}$, estimée à 5 171 304 euros, soit une diminution de plus de 11 % par rapport à la perte initiale. Pour $s = 1\%$, les pertes attendues avec l'allocation « optimale » se réduisent également, permettant une économie estimée à 1 033 176 euros.

Ces résultats illustrent l'intérêt d'une réallocation des expositions pour réduire les pertes espérées tout en identifiant les secteurs où le risque est le plus faible. Toutefois, la faisabilité de ces changements dépend des capacités réelles de l'assureur à ajuster sa présence sur certains marchés, en tenant compte des contraintes de diversification et des réalités du marché de l'assurance. En examinant l'allocation des secteurs de Coface au cours des trois dernières années, on constate qu'elle a varié de moins de 1% sur chaque secteur chaque année, ce qui souligne une inertie potentielle dans les ajustements des expositions.

3.3.3 Allocation optimale dans le contexte actuel

Dans la sous-section précédente, la résolution du problème d'optimisation s'est faite dans un contexte rétrospectif, c'est-à-dire en utilisant les données historiques de sinistralité pour estimer les pertes et déterminer une allocation optimale d'exposition. La résolution d'un tel problème est assez triviale, car elle repose sur des données passées dont les résultats sont déjà connus.

Cependant, dans une perspective de prédiction et compte tenu du contexte actuel incertain, l'assureur fait face à des incertitudes importantes quant au nombre de défauts futurs, ce qui complexifie le problème d'optimisation. La méthodologie doit désormais intégrer cette incertitude en modélisant les variations potentielles de la sinistralité. Pour cela, nous étudierons l'allocation optimale selon divers scénarios sur le nombre de défauts.

Afin de modéliser cette incertitude, nous analysons deux scénarios distincts en 2024, comme illustré dans le tableau 3.5. Le **scénario central** repose sur la moyenne des simulations, tandis que le **scénario pessimiste** correspond au quantile à 99.5% des simulations. Dans les deux cas, on observe que les défaillances évoluent selon la même tendance dans tous les secteurs d'activité.

	θ^0	scénario central			scénario pessimiste		
		$N^{(k)}$	θ	Δ	$N^{(k)}$	θ	Δ
(A)	0.1671	1 371	0.1771	0.01	1 535	0.1771	0.01
(I)	0.3401	4 701	0.3301	-0.01	5 030	0.3301	-0.01
(C)	0.1073	14 988	0.0973	-0.01	15 873	0.0973	-0.01
(S)	0.3541	38 668	0.3541	0.00	40 105	0.3541	0.00
(E)	0.0314	5 498	0.0414	0.01	5 863	0.0414	0.01
$\hat{L}^{\text{Modèle}}$		54 485 592	53 138 834	- 1 346 758	57 703 379	56 281 529	- 1 421 850

TABLE 3.5 : Allocation optimale pour l'année 2024

Les secteurs identifiés comme les moins risqués demeurent les mêmes qu'en 2023, et les conclusions relatives à l'allocation optimale s'appliquent également aux prévisions pour 2024. En particulier, bien que l'intérêt d'une réallocation des expositions pour réduire les pertes attendues soit reconnu, la faisabilité de tels ajustements reste conditionnée par les capacités opérationnelles de l'assureur à adapter ses positions sur certains marchés.

Ainsi, bien qu'une redistribution de l'exposition puisse, théoriquement, constituer une stratégie intéressante pour piloter le risque d'un assureur-crédit, notre modélisation simplifiée conduit à une allocation maximale vers les secteurs présentant le rapport le plus favorable entre exposition moyenne et nombre de défauts.

Une particularité de l'assurance-crédit réside dans la possibilité d'ajuster à tout moment les limites de couverture des contrats, un aspect essentiel pour maîtriser l'exposition globale au risque d'un assureur-crédit. Toutefois, notre modèle identifie des tendances sectorielles trop générales pour mettre en lumière des disparités dans l'évolution du nombre de défauts. Par conséquent, une modélisation plus fine, ciblant des sous-secteurs spécifiques, pourrait révéler des variations distinctes des défauts sectoriels et permettre une détection plus précise des entreprises à risque élevé. Cela offrirait à l'assureur un

cadre pour ajuster de manière ciblée les limites de couverture, contribuant ainsi à une gestion du risque plus proactive et granulaire.

Par ailleurs, pour maximiser l'efficacité de cette approche, il faudrait l'actualiser sur une base mensuelle ou trimestrielle. Une telle fréquence d'actualisation permettrait de tirer pleinement parti des avantages offerts par le modèle de Hawkes. En intégrant les données récentes sur les défauts et les indicateurs économiques, l'assureur pourrait mieux anticiper les tendances émergentes et réagir rapidement aux changements dans le paysage économique, renforçant ainsi sa capacité à gérer le risque de manière dynamique et informée.

Conclusion

Ce mémoire avait pour objectif d'étudier l'impact de la conjoncture économique sur l'activité d'un assureur-crédit, en proposant une modélisation dynamique du nombre de défauts en fonction des fluctuations économiques à l'aide des processus de Hawkes.

La première partie de nos travaux s'est concentrée sur l'évaluation de la capacité d'un modèle de Hawkes multivarié à modéliser les relations entre les défauts par secteurs d'activité. Les résultats obtenus suggèrent que ce type de modèle est adapté pour capturer les dépendances entre les défauts. Une modélisation sur une segmentation plus fine des entreprises en tenant compte d'autres critères tels que le niveau d'exposition, la notation de l'acheteur ou encore la localisation géographique, pourrait constituer une piste intéressante pour des recherches futures.

Le modèle proposé dans cette étude, qui prend en compte l'environnement global de risque des entreprises, s'est révélé satisfaisant. L'intensité de base du processus parvient à refléter l'environnement économique spécifique à chaque secteur, tandis que les paramètres d'excitation permettent de modéliser les dépendances entre les défauts. L'un des principaux avantages de ce modèle réside dans sa capacité à suivre le risque de manière dynamique, en intégrant les temps de survenance réels des événements dans son historique. Cette caractéristique améliore la précision des prévisions à court terme.

De plus, en définissant l'intensité de base comme une combinaison linéaire d'indicateurs économiques, le modèle parvient à intégrer les évolutions macroéconomiques sans nécessiter de réajustements des paramètres, contrairement aux modèles à facteurs. Toutefois, cette approche est principalement adaptée à une modélisation à court terme. Un backtesting du modèle a montré que, lorsqu'il est alimenté par des indicateurs économiques prospectifs, il est capable de produire des prévisions fiables à long terme. Cela suggère qu'en le couplant à un générateur de scénarios économiques, il pourrait être utilisé pour des projections à plus long terme.

Cependant, le modèle a montré ses limites pour modéliser le nombre de défauts durant la crise Covid, notamment en raison des difficultés à intégrer et à quantifier les mesures de soutien de l'État via des indicateurs économiques. Cette sur-estimation a néanmoins permis d'évaluer les défauts qui auraient dû se produire en l'absence de ces interventions, offrant ainsi une estimation de l'intensité de l'effet de rattrapage attendu post-Covid.

Les processus de Hawkes permettent de capter des effets de contagions intra et inter-segments. Le fait que ces derniers soient représentés de façon intuitive à travers les paramètres fait de ces processus un outil intéressant pour le pilotage des risques. Cependant, les modèles de Hawkes présentent certaines limites. La première réside dans le fait que le nombre de paramètres croît considérablement avec l'augmentation du nombre de variables, ce qui engendre des problématiques de complexité algorithmique, de difficulté d'interprétation et d'erreurs dans l'estimation des paramètres. Une seconde limite concerne sa sensibilité aux données. Comme pour tout modèle, cela affecte les paramètres estimés, mais l'historique des données peut également influencer le processus lors des prédictions.

Le modèle de pertes simplifié que nous avons développé avait pour but de déterminer un mon-

tant de pertes associées aux défauts modélisés par le processus de Hawkes. Les résultats obtenus ont montré que ce modèle permet de calculer des pertes similaires à celles observées chez un véritable assureur-crédit, et cela avec des données approximatives. Des travaux futurs, en présence de données réelles d'assureur-crédit, pourraient se pencher sur la performance du modèle de pertes théorique, en modélisant notamment la dépendance entre le taux de recouvrement et la fréquence des défauts, en comparaison avec les résultats des modèles à facteurs.

Enfin, l'étude de l'allocation optimale de l'exposition d'un assureur-crédit entre différents secteurs d'activité a montré que cette optimisation est fortement contrainte par les conditions de marché, réduisant ainsi la marge de manœuvre de l'assureur. Dans ce contexte, la gestion des limites de couverture reste l'outil principal pour maîtriser les risques.

Bibliographie


- ALLIANZ TRADE (mars 2024). Forte hausse des défaillances d'entreprises attendue en France et dans le monde en 2024.
- AU GROUP (2024). Le marché de l'Assurance-Crédit.
- BANQUE DE FRANCE (2024). Stat Info - Les défaillances d'entreprises.
- BESSY-ROLAND, Y., BOUMEZOUED, A. et HILLAIRET, C. (2021). Multivariate Hawkes process for cyber insurance. *Annals of Actuarial Science* 15.1, 14–39.
- BYRD, R. H., LU, P., NOCEDAL, J. et ZHU, C. (1995). A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal on Scientific Computing* 16.5, p. 1190-1208.
- CAJA, A. (2014). Contribution à la mesure des engagements et du besoin en capital pour un assureur crédit. Thèse de doct. École doctorale Sciences économiques et gestion (Lyon).
- CHEN, Y. (2016a). Likelihood Function for Multivariate Hawkes Processes.
- CHEN, Y. (2016b). Thinning Algorithms for Simulating Point Processes.
- CIARLET, P. G. (1990). Introduction à l'analyse matricielle et à l'optimisation. Masson.
- CORNILLON, P., HENGARTNER, N., MATZNER-LOBER, E. et ROUVIÈRE, L. (2023). Régression avec R. Pratique R. Edp Science.
- DALEY, D. J. et VERE-JONES, D. (2003). An introduction to the theory of point processes. Vol. I. Second. Probability and its Applications (New York). Springer-Verlag.
- DAS, S., DUFFIE, D., KAPADIA, N. et SAITA, L. (fév. 2007). Common Failings: How Corporate Defaults Are Correlated. *The Journal of Finance* 62, p. 93 -117.
- DUFFIE, D., FILIPOVIĆ, D. et SCHACHERMAYER, W. (2003). Affine Processes and Applications in Finance. *The Annals of Applied Probability* 13.3, p. 984-1053.
- ERRAIS, E., GIESECKE, K. et GOLDBERG, L. (2010). Affine Point Processes and Portfolio Credit Risk. *SIAM J. Financial Math.* 1, p. 642-665.
- FILIMONOV, V. et SORNETTE, D. (2015). Apparent criticality and calibration issues in the Hawkes self-excited point process model: application to high-frequency financial data. *Quantitative Finance* 15.8, p. 1293-1314.
- HAWKES, A. G. (1971). Spectra of Some Self-Exciting and Mutually Exciting Point Processes. *Biometrika* 58.1, p. 83-90.
- HILLAIRET, C., LOPEZ, O., D'OULTREMONT, L. et SPOORENBERG, B. (2022). Cyber-contagion model with network structure applied to insurance. *Insurance: Mathematics and Economics* 107, p. 88-101.
- ICISA et VERSICHERUNGSFOREN LEIPZIG (2023). Credit Insurance : Loss Given Default Study.
- INSEE (2022). Insee Résultats - Les comptes de la Nation en 2021.
- LAUB, P. J., LEE, Y., POLLETT, P. et TAIMRE, T. (2024). Hawkes Models And Their Applications.
- LAUB, P. J., LEE, Y. et TAIMRE, T. (2021). The Elements of Hawkes Processes. Springer Cham.
- LAUB, P. J., TAIMRE, T. et POLLETT, P. K. (2015). Hawkes Processes.

- LEWIS, C. (1982). Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting. Butterworth scientific. Butterworth Scientific.
- LEWIS, P. A. W et SHEDLER, G. S. (1979). Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics Quarterly* 26.3, p. 403-413.
- MERTON, R. C. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *The Journal of Finance* 29.2, p. 449-470.
- NELDER, J. A. et MEAD, R. (jan. 1965). A Simplex Method for Function Minimization. *The Computer Journal* 7.4, p. 308-313.
- OGATA, Y. (1981). On Lewis' simulation method for point processes. *IEEE Trans. Inf. Theory* 27, p. 23-30.
- OGATA, Y. (1988). Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes. *Journal of the American Statistical Association* 83.401, p. 9-27.
- OZAKI, T. (1979). Maximum likelihood estimation of Hawkes' self-exciting point processes. *Annals of the Institute of Statistical Mathematics* 31, p. 145-155.
- RASMUSSEN, J. (2013). Bayesian inference for Hawkes processes. *Methodology and Computing in Applied Probability* 15, p. 623-642.
- RIZOIU, M.-A., LEE, Y., MISHRA, S. et XIE, L. (2017). A Tutorial on Hawkes Processes for Events in Social Media.
- SAUNDERS, A. et ALLEN, L. (2010). The Credit Risk of Portfolios and Correlations. *Credit Risk Management in and Out of the Financial Crisis*. John Wiley & Sons, Ltd. Chap. 8, p. 148-165.
- SWISS RE INSTITUTE (sept. 2023). Credit and surety in the age of economic uncertainty.
- THOMPSON, A. et JESSOP, N. (2018). A Cost of Capital Approach to Estimating Credit Risk Premia. *Moody's Analytics Research*.
- ULLMO, F. (2019). Modélisation des sinistres futurs sur un portefeuille d'assurance-crédit. Mém. de mast. ISFA.
- VASICEK, O. A. (déc. 2002). The Distribution of Loan Portfolio Value. *Risk*.
- ZHU, C., BYRD, R. H., LU, P. et NOCEDAL, J. (1997). Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw.* 23, p. 550-560.

Annexe A

Annexes

A.1 Exemple de grille de notation

	Exemples de types d'entreprises	Risque de défaut *
1	Entreprise d'envergure nationale ou multinationale Bonne visibilité de l'activité à long terme Equipe de direction étoffée Structure financière très solide Performances exceptionnelles et récurrentes quel que soit l'environnement économique	Très faible  Très élevé
2	Entreprise d'envergure nationale ou multinationale Visibilité correcte de l'activité à long terme Equipe de direction étoffée Structure financière solide Excellentes performances récurrentes quel que soit l'environnement économique	
3	Bonne visibilité de l'activité à moyen terme Structure financière solide Très bonnes performances et très bonne capacité à résister aux évolutions de l'environnement économique	
4	Visibilité correcte de l'activité à moyen terme Structure financière satisfaisante Bonnes performances opérationnelles et bonne capacité à résister aux évolutions de l'environnement économique	
5	Information partielle et/ou visibilité partielle de l'activité à moyen terme et/ou structure financière correcte et/ou performances correctes	
6	Information partielle et/ou visibilité sur l'activité limitée et/ou structure financière moyenne et/ou performances moyennes	
7	Information partielle et/ou visibilité sur l'activité limitée et/ou structure financière faible et/ou faibles performances	
8	Information et/ou visibilité sur l'activité très limitée et/ou structure financière insuffisante et/ou rentabilité insuffisante	
9	Entreprise en grande difficulté pour faire face à ses engagements et/ou information très insuffisante.	
10	Affaire en procédure collective (redressement judiciaire, liquidation judiciaire, sauvegarde)	
NA	Entreprise en sommeil ou ne réalisant pas d'achat ou entreprise n'achetant pas pour son compte	

* Risque de défaut : indicateur de solidité financière

La notation repose sur notre analyse de la situation économique et financière de chaque entreprise en fonction des informations portées à notre connaissance. Elle varie en fonction de nouvelles informations sur la situation de l'entreprise concernée. Cette notation n'est en aucun cas une recommandation, pour contracter avec une entreprise. Chaque fournisseur assuré reste libre de définir sa politique commerciale comme il l'entend.

FIGURE A.1 : Echelle grille de notation Allianz Trade

A.2 Algorithme de *thinning* de Lewis

Pour la simulation d'un processus de Poisson non-homogène avec une intensité $\lambda(\cdot)$, on peut utiliser l'algorithme 3 proposé par LEWIS et SHEDLER (1979). Cet algorithme s'appuie sur le théorème suivant dont la démonstration est détaillée dans CHEN (2016b).

Théorème A.1. *Considérons $\bar{N} = (\bar{N}_t)_{t \geq 0}$ un processus de Poisson homogène d'intensité $\bar{\lambda}$. Notons t_1, \dots, t_n les temps de sauts du processus \bar{N} observés sur $[0, T]$. Si $\lambda(\cdot)$ est une fonction déterministe positive telle que*

$$\forall t \in [0, T], \quad \lambda(t) \leq \bar{\lambda}.$$

Si pour $k \in \{1, \dots, n\}$, on supprime le point t_k avec probabilité $1 - \frac{\lambda(t_k)}{\bar{\lambda}}$, alors les points conservés forment un processus de Poisson non-homogène d'intensité $\lambda(\cdot)$ sur $[0, T]$.

Algorithme 3 : Simulation d'un processus de Poisson non-homogène d'intensité $\lambda(\cdot)$ sur $[0, T]$

Initialisation : $t \leftarrow 0, k \leftarrow 0, \lambda_{sup} = \sup_{0 \leq t \leq T} \lambda(t)$;

tant que $t < T$ **faire**

 Simuler $\tau \sim \mathcal{E}(\lambda_{sup})$;

$t \leftarrow t + \tau$;

 Simuler $U \sim \mathcal{U}(0, 1)$;

si $U \leq \frac{\lambda(t)}{\lambda_{sup}}$ **et** $t < T$ **alors**

 On accepte le point candidat : $t_{k+1} \leftarrow t$;

$k \leftarrow k + 1$;

renvoyer $(t_i)_{i=1, \dots, k}$

A.3 Méthode BFGS

L'algorithme Broyden–Fletcher–Goldfarb–Shanno (BFGS) est un algorithme d'optimisation de type *quasi-Newton*. Une méthode quasi-Newton repose sur le même principe que l'algorithme de Newton-Raphson, l'inverse de la Jacobienne est remplacée par une approximation plus facile à calculer. Ce type de méthode est privilégié lorsque l'on ne sait pas calculer analytiquement la hessienne de la fonction à minimiser ou que son évaluation est trop coûteuse. Les étapes principales de l'algorithme BFGS pour un problème de minimisation de la forme $\min_{\theta \in \mathbb{R}^n} f(\theta)$ sont décrites dans l'algorithme 4.

Algorithme 4 : Algorithme BFGS pour $\min_{\theta \in \mathbb{R}^n} f(\theta)$

Initialisation : choisir un point initial $\theta^{(0)} \in \mathbb{R}^n$ et une approximation de la matrice Hessienne B_0 ;

tant que $t < T$ **faire**

Déterminer p_k tel que $B_k p_k = -\nabla f(\theta^{(k)})$;

Calculer α_k par une recherche linéaire dans la direction de descente p_k ;

$\theta^{(k+1)} \leftarrow \theta^{(k)} + \alpha_k p_k$;

$s_k \leftarrow \theta^{(k+1)} - \theta^{(k)}$;

$y_k \leftarrow \nabla f(\theta^{(k+1)}) - \nabla f(\theta^{(k)})$;

$B_{k+1} \leftarrow B_k + \frac{y_k y_k^\top}{y_k^\top s_k} - \frac{B_k s_k s_k^\top B_k}{s_k^\top B_k s_k}$;

renvoyer $(t_i)_{i=1,\dots,k}$

En pratique, on prend souvent $B_0 = I$ où I est la matrice identité et si l'on ne dispose pas de formule explicite pour calculer le gradient, celui-ci est approximé par différences finies.

A.4 Résultats calibrages

A.4.1 Répartition des temps de sauts

Secteur	Historique	Répartition des temps de sauts	θ_{start}	θ
Agriculture	2008-01- 2009-12	uniforme	108.5, 0.01, 0.05	95.42, 0.02, 0.11
Agriculture	2008-01- 2009-12	uniforme	108.5, 0.1, 0.3	95.42, 0.02, 0.11
Agriculture	2008-01- 2009-12	uniforme	108.5, 1.0, 2.0	108.5, 0.0, 4.12
Agriculture	2008-01- 2009-12	uniforme	108.5, 10.0, 15.0	108.49, 0.0, 0.0
Agriculture	2008-01- 2009-12	uniforme	108.5, 100.0, 101.0	108.5, 0.0, 28057704.9
Industrie	2008-01- 2009-12	uniforme	419.0, 0.01, 0.05	305.29, 0.07, 0.21
Industrie	2008-01- 2009-12	uniforme	419.0, 0.1, 0.3	305.29, 0.07, 0.21
Industrie	2008-01- 2009-12	uniforme	419.0, 1.0, 2.0	305.29, 0.07, 0.21
Industrie	2008-01- 2009-12	uniforme	419.0, 10.0, 15.0	419.0, 0.0, 17.94
Industrie	2008-01- 2009-12	uniforme	419.0, 100.0, 101.0	419.0, 0.0, 534.03
Construction	2008-01- 2009-12	uniforme	1328.92, 0.01, 0.05	1050.76, 0.04, 0.11
Construction	2008-01- 2009-12	uniforme	1328.92, 0.1, 0.3	1050.76, 0.04, 0.11
Construction	2008-01- 2009-12	uniforme	1328.92, 1.0, 2.0	1050.76, 0.04, 0.11
Construction	2008-01- 2009-12	uniforme	1328.92, 10.0, 15.0	1050.76, 0.04, 0.11
Construction	2008-01- 2009-12	uniforme	1328.92, 100.0, 101.0	1328.92, 0.0, 244.04
Enseignement, ...	2008-01- 2009-12	uniforme	371.33, 0.01, 0.05	331.81, 0.02, 0.13
Enseignement, ...	2008-01- 2009-12	uniforme	371.33, 0.1, 0.3	331.81, 0.02, 0.13
Enseignement, ...	2008-01- 2009-12	uniforme	371.33, 1.0, 2.0	284.35, 0.33, 1.34
Enseignement, ...	2008-01- 2009-12	uniforme	371.33, 10.0, 15.0	371.33, 0.0, 31.58
Enseignement, ...	2008-01- 2009-12	uniforme	371.33, 100.0, 101.0	371.33, 0.0, 509.11
Services marchands	2008-01- 2009-12	uniforme	2720.54, 0.01, 0.05	2159.65, 0.05, 0.18
Services marchands	2008-01- 2009-12	uniforme	2720.54, 0.1, 0.3	2159.65, 0.05, 0.18
Services marchands	2008-01- 2009-12	uniforme	2720.54, 1.0, 2.0	1451.64, 1.27, 2.68
Services marchands	2008-01- 2009-12	uniforme	2720.54, 10.0, 15.0	1451.64, 1.27, 2.68
Services marchands	2008-01- 2009-12	uniforme	2720.54, 100.0, 101.0	2720.54, 0.0, 227.4

Secteur	Historique	Répartition des temps de sauts	θ_{start}	θ
Agriculture	2008-01- 2009-12	uniforme	54.25, 0.01, 0.05	95.42, 0.02, 0.11
Agriculture	2008-01- 2009-12	uniforme	10.85, 0.1, 0.3	95.42, 0.02, 0.11
Agriculture	2008-01- 2009-12	uniforme	1.08, 1.0, 2.0	108.5, 0.0, 2.54
Agriculture	2008-01- 2009-12	uniforme	0.22, 10.0, 15.0	95.42, 0.02, 0.11
Agriculture	2008-01- 2009-12	uniforme	0.11, 100.0, 101.0	108.49, 0.0, 0.0
Industrie	2008-01- 2009-12	uniforme	209.5, 0.01, 0.05	305.29, 0.07, 0.21
Industrie	2008-01- 2009-12	uniforme	41.9, 0.1, 0.3	305.29, 0.07, 0.21
Industrie	2008-01- 2009-12	uniforme	4.19, 1.0, 2.0	305.29, 0.07, 0.21
Industrie	2008-01- 2009-12	uniforme	0.84, 10.0, 15.0	305.29, 0.07, 0.21
Industrie	2008-01- 2009-12	uniforme	0.42, 100.0, 101.0	305.29, 0.07, 0.21
Construction	2008-01- 2009-12	uniforme	664.46, 0.01, 0.05	1050.76, 0.04, 0.11
Construction	2008-01- 2009-12	uniforme	132.89, 0.1, 0.3	1050.76, 0.04, 0.11
Construction	2008-01- 2009-12	uniforme	13.29, 1.0, 2.0	1050.76, 0.04, 0.11
Construction	2008-01- 2009-12	uniforme	2.66, 10.0, 15.0	1050.76, 0.04, 0.11
Construction	2008-01- 2009-12	uniforme	1.33, 100.0, 101.0	1050.76, 0.04, 0.11
Enseignement, ...	2008-01- 2009-12	uniforme	185.67, 0.01, 0.05	331.81, 0.02, 0.13
Enseignement, ...	2008-01- 2009-12	uniforme	37.13, 0.1, 0.3	331.81, 0.02, 0.13
Enseignement, ...	2008-01- 2009-12	uniforme	3.71, 1.0, 2.0	0.0, 4.39, 4.34
Enseignement, ...	2008-01- 2009-12	uniforme	0.74, 10.0, 15.0	331.81, 0.02, 0.13
Enseignement, ...	2008-01- 2009-12	uniforme	0.37, 100.0, 101.0	284.35, 0.33, 1.34
Services marchands	2008-01- 2009-12	uniforme	1360.27, 0.01, 0.05	2159.66, 0.05, 0.18
Services marchands	2008-01- 2009-12	uniforme	272.05, 0.1, 0.3	2159.65, 0.05, 0.18
Services marchands	2008-01- 2009-12	uniforme	27.21, 1.0, 2.0	0.0, 11.17, 11.13
Services marchands	2008-01- 2009-12	uniforme	5.44, 10.0, 15.0	1451.65, 1.27, 2.68
Services marchands	2008-01- 2009-12	uniforme	2.72, 100.0, 101.0	1451.64, 1.27, 2.68

Secteur	Historique	Répartition des temps de sauts	θ_{start}	θ
Agriculture	2008-01- 2009-12	hawkes	108.5, 0.01, 0.05	94.38, 0.02, 0.12
Agriculture	2008-01- 2009-12	hawkes	108.5, 0.1, 0.3	94.79, 0.02, 0.11
Agriculture	2008-01- 2009-12	hawkes	108.5, 1.0, 2.0	79.14, 2.71, 9.95
Agriculture	2008-01- 2009-12	hawkes	108.5, 10.0, 15.0	53.69, 8.88, 17.51
Agriculture	2008-01- 2009-12	hawkes	108.5, 100.0, 101.0	26.64, 26.63, 35.17
Industrie	2008-01- 2009-12	hawkes	419.0, 0.01, 0.05	302.91, 0.07, 0.21
Industrie	2008-01- 2009-12	hawkes	419.0, 0.1, 0.3	305.36, 0.07, 0.2
Industrie	2008-01- 2009-12	hawkes	419.0, 1.0, 2.0	166.31, 5.7, 9.4
Industrie	2008-01- 2009-12	hawkes	419.0, 10.0, 15.0	119.71, 14.75, 20.58
Industrie	2008-01- 2009-12	hawkes	419.0, 100.0, 101.0	58.8, 40.23, 46.7
Construction	2008-01- 2009-12	hawkes	1328.92, 0.01, 0.05	1049.33, 0.04, 0.11
Construction	2008-01- 2009-12	hawkes	1328.92, 0.1, 0.3	1050.36, 0.04, 0.11
Construction	2008-01- 2009-12	hawkes	1328.92, 1.0, 2.0	371.83, 13.21, 18.28
Construction	2008-01- 2009-12	hawkes	1328.92, 10.0, 15.0	253.3, 27.4, 33.8
Construction	2008-01- 2009-12	hawkes	1328.92, 100.0, 101.0	109.77, 75.73, 82.47
Enseignement, ...	2008-01- 2009-12	hawkes	371.33, 0.01, 0.05	329.94, 0.02, 0.14
Enseignement, ...	2008-01- 2009-12	hawkes	371.33, 0.1, 0.3	331.22, 0.02, 0.13
Enseignement, ...	2008-01- 2009-12	hawkes	371.33, 1.0, 2.0	197.8, 4.69, 9.99
Enseignement, ...	2008-01- 2009-12	hawkes	371.33, 10.0, 15.0	137.9, 14.82, 23.51
Enseignement, ...	2008-01- 2009-12	hawkes	371.33, 100.0, 101.0	63.05, 43.65, 52.47
Services marchands	2008-01- 2009-12	hawkes	2720.54, 0.01, 0.05	2158.22, 0.05, 0.18
Services marchands	2008-01- 2009-12	hawkes	2720.54, 0.1, 0.3	2153.62, 0.05, 0.18
Services marchands	2008-01- 2009-12	hawkes	2720.54, 1.0, 2.0	653.22, 17.75, 23.31
Services marchands	2008-01- 2009-12	hawkes	2720.54, 10.0, 15.0	380.76, 41.68, 48.41
Services marchands	2008-01- 2009-12	hawkes	2720.54, 100.0, 101.0	166.06, 98.94, 105.3

Secteur	Historique	Répartition des temps de sauts	θ_{start}	θ
Agriculture	2008-01- 2009-12	hawkes	54.25, 0.01, 0.05	94.78, 0.02, 0.11
Agriculture	2008-01- 2009-12	hawkes	10.85, 0.1, 0.3	94.0, 0.02, 0.12
Agriculture	2008-01- 2009-12	hawkes	1.08, 1.0, 2.0	57.64, 45.55, 97.01
Agriculture	2008-01- 2009-12	hawkes	0.22, 10.0, 15.0	0.0, 67.23, 67.01
Agriculture	2008-01- 2009-12	hawkes	0.11, 100.0, 101.0	17.55, 5622.56, 6702.95
Industrie	2008-01- 2009-12	hawkes	209.5, 0.01, 0.05	301.4, 0.08, 0.22
Industrie	2008-01- 2009-12	hawkes	41.9, 0.1, 0.3	163.73, 4.96, 8.1
Industrie	2008-01- 2009-12	hawkes	4.19, 1.0, 2.0	0.0, 39.51, 39.41
Industrie	2008-01- 2009-12	hawkes	0.84, 10.0, 15.0	0.0, 228.21, 228.18
Industrie	2008-01- 2009-12	hawkes	0.42, 100.0, 101.0	38.81, 2930.25, 3229.07
Construction	2008-01- 2009-12	hawkes	664.46, 0.01, 0.05	1052.38, 0.04, 0.11
Construction	2008-01- 2009-12	hawkes	132.89, 0.1, 0.3	439.16, 11.1, 16.54
Construction	2008-01- 2009-12	hawkes	13.29, 1.0, 2.0	0.0, 39.78, 39.75
Construction	2008-01- 2009-12	hawkes	2.66, 10.0, 15.0	0.0, 542.74, 542.73
Construction	2008-01- 2009-12	hawkes	1.33, 100.0, 101.0	0.0, 1509.61, 1531.59
Enseignement, ...	2008-01- 2009-12	hawkes	185.67, 0.01, 0.05	330.71, 0.02, 0.14
Enseignement, ...	2008-01- 2009-12	hawkes	37.13, 0.1, 0.3	202.42, 4.48, 9.78
Enseignement, ...	2008-01- 2009-12	hawkes	3.71, 1.0, 2.0	0.0, 30.67, 30.63
Enseignement, ...	2008-01- 2009-12	hawkes	0.74, 10.0, 15.0	0.0, 213.55, 213.39
Enseignement, ...	2008-01- 2009-12	hawkes	0.37, 100.0, 101.0	34.95, 3024.93, 3335.61
Services marchands	2008-01- 2009-12	hawkes	1360.27, 0.01, 0.05	2148.33, 0.05, 0.19
Services marchands	2008-01- 2009-12	hawkes	272.05, 0.1, 0.3	2123.74, 0.05, 0.19
Services marchands	2008-01- 2009-12	hawkes	27.21, 1.0, 2.0	0.0, 41.8, 41.76
Services marchands	2008-01- 2009-12	hawkes	5.44, 10.0, 15.0	893.06, 1652.11, 2459.33
Services marchands	2008-01- 2009-12	hawkes	2.72, 100.0, 101.0	0.0, 1135.87, 1142.52

Secteur	Historique	Répartition des temps de sauts	θ_{start}	θ
Agriculture	2011-01- 2012-12	uniforme	102.46, 0.01, 0.05	102.45, 0.0, 0.25
Agriculture	2011-01- 2012-12	uniforme	102.46, 0.1, 0.3	99.57, 0.01, 0.47
Agriculture	2011-01- 2012-12	uniforme	102.46, 1.0, 2.0	102.46, 0.0, 4.36
Agriculture	2011-01- 2012-12	uniforme	102.46, 10.0, 15.0	102.46, 0.0, 32.19
Agriculture	2011-01- 2012-12	uniforme	102.46, 100.0, 101.0	102.46, 0.0, 28189093.5
Industrie	2011-01- 2012-12	uniforme	387.88, 0.01, 0.05	383.7, 0.0, 0.0
Industrie	2011-01- 2012-12	uniforme	387.88, 0.1, 0.3	387.87, 0.0, 2.0
Industrie	2011-01- 2012-12	uniforme	387.88, 1.0, 2.0	387.87, 0.0, 2.0
Industrie	2011-01- 2012-12	uniforme	387.88, 10.0, 15.0	387.87, 0.0, 29.76
Industrie	2011-01- 2012-12	uniforme	387.88, 100.0, 101.0	387.87, 0.0, 10085892.2
Construction	2011-01- 2012-12	uniforme	1301.08, 0.01, 0.05	1286.83, 0.0, 0.0
Construction	2011-01- 2012-12	uniforme	1301.08, 0.1, 0.3	1301.08, 0.0, 3.44
Construction	2011-01- 2012-12	uniforme	1301.08, 1.0, 2.0	1211.49, 0.3, 4.33
Construction	2011-01- 2012-12	uniforme	1301.08, 10.0, 15.0	1301.08, 0.0, 26.51
Construction	2011-01- 2012-12	uniforme	1301.08, 100.0, 101.0	1301.08, 0.0, 231.86
Enseignement, ...	2011-01- 2012-12	uniforme	428.62, 0.01, 0.05	419.68, 0.0, 0.0
Enseignement, ...	2011-01- 2012-12	uniforme	428.62, 0.1, 0.3	419.68, 0.0, 0.0
Enseignement, ...	2011-01- 2012-12	uniforme	428.62, 1.0, 2.0	428.57, 0.0, 0.0
Enseignement, ...	2011-01- 2012-12	uniforme	428.62, 10.0, 15.0	428.63, 0.0, 42.75
Enseignement, ...	2011-01- 2012-12	uniforme	428.62, 100.0, 101.0	428.63, 0.0, 502.44
Services marchands	2011-01- 2012-12	uniforme	2802.83, 0.01, 0.05	2710.05, 0.0, 0.0
Services marchands	2011-01- 2012-12	uniforme	2802.83, 0.1, 0.3	2710.05, 0.0, 0.0
Services marchands	2011-01- 2012-12	uniforme	2802.83, 1.0, 2.0	2506.03, 0.42, 3.88
Services marchands	2011-01- 2012-12	uniforme	2802.83, 10.0, 15.0	2802.83, 0.0, 23.29
Services marchands	2011-01- 2012-12	uniforme	2802.83, 100.0, 101.0	2802.83, 0.0, 228.41

Secteur	Historique	Répartition des temps de sauts	θ_{start}	θ
Agriculture	2011-01- 2012-12	uniforme	51.23, 0.01, 0.05	99.57, 0.01, 0.47
Agriculture	2011-01- 2012-12	uniforme	10.25, 0.1, 0.3	102.46, 0.0, 1.74
Agriculture	2011-01- 2012-12	uniforme	1.02, 1.0, 2.0	102.45, 0.0, 0.25
Agriculture	2011-01- 2012-12	uniforme	0.2, 10.0, 15.0	102.45, 0.0, 0.25
Agriculture	2011-01- 2012-12	uniforme	0.1, 100.0, 101.0	102.45, 0.0, 0.25
Industrie	2011-01- 2012-12	uniforme	193.94, 0.01, 0.05	383.7, 0.0, 0.0
Industrie	2011-01- 2012-12	uniforme	38.79, 0.1, 0.3	387.87, 0.0, 2.0
Industrie	2011-01- 2012-12	uniforme	3.88, 1.0, 2.0	387.87, 0.0, 2.0
Industrie	2011-01- 2012-12	uniforme	0.78, 10.0, 15.0	387.87, 0.0, 2.0
Industrie	2011-01- 2012-12	uniforme	0.39, 100.0, 101.0	387.87, 0.0, 5.63
Construction	2011-01- 2012-12	uniforme	650.54, 0.01, 0.05	1300.93, 0.0, 0.0
Construction	2011-01- 2012-12	uniforme	130.11, 0.1, 0.3	0.0, 8.86, 8.81
Construction	2011-01- 2012-12	uniforme	13.01, 1.0, 2.0	0.0, 8.86, 8.81
Construction	2011-01- 2012-12	uniforme	2.6, 10.0, 15.0	1301.08, 0.0, 9.06
Construction	2011-01- 2012-12	uniforme	1.3, 100.0, 101.0	1211.49, 0.3, 4.33
Enseignement, ...	2011-01- 2012-12	uniforme	214.31, 0.01, 0.05	419.68, 0.0, 0.0
Enseignement, ...	2011-01- 2012-12	uniforme	42.86, 0.1, 0.3	428.57, 0.0, 0.0
Enseignement, ...	2011-01- 2012-12	uniforme	4.29, 1.0, 2.0	419.68, 0.0, 0.0
Enseignement, ...	2011-01- 2012-12	uniforme	0.86, 10.0, 15.0	428.57, 0.0, 0.0
Enseignement, ...	2011-01- 2012-12	uniforme	0.43, 100.0, 101.0	428.62, 0.0, 4.61
Services marchands	2011-01- 2012-12	uniforme	1401.42, 0.01, 0.05	2710.05, 0.0, 0.0
Services marchands	2011-01- 2012-12	uniforme	280.28, 0.1, 0.3	2506.03, 0.42, 3.88
Services marchands	2011-01- 2012-12	uniforme	28.03, 1.0, 2.0	0.0, 12.42, 12.37
Services marchands	2011-01- 2012-12	uniforme	5.61, 10.0, 15.0	2802.83, 0.0, 9.98
Services marchands	2011-01- 2012-12	uniforme	2.8, 100.0, 101.0	2506.04, 0.42, 3.88

Secteur	Historique	Répartition des temps de sauts	θ_{start}	θ
Agriculture	2011-01- 2012-12	hawkes	102.46, 0.01, 0.05	102.46, 0.0, 0.27
Agriculture	2011-01- 2012-12	hawkes	102.46, 0.1, 0.3	96.26, 0.04, 0.63
Agriculture	2011-01- 2012-12	hawkes	102.46, 1.0, 2.0	85.67, 3.18, 19.37
Agriculture	2011-01- 2012-12	hawkes	102.46, 10.0, 15.0	54.74, 7.07, 15.11
Agriculture	2011-01- 2012-12	hawkes	102.46, 100.0, 101.0	26.98, 25.05, 33.85
Industrie	2011-01- 2012-12	hawkes	387.88, 0.01, 0.05	383.7, 0.0, 0.0
Industrie	2011-01- 2012-12	hawkes	387.88, 0.1, 0.3	387.87, 0.0, 2.9
Industrie	2011-01- 2012-12	hawkes	387.88, 1.0, 2.0	232.29, 5.92, 14.71
Industrie	2011-01- 2012-12	hawkes	387.88, 10.0, 15.0	122.09, 16.26, 23.64
Industrie	2011-01- 2012-12	hawkes	387.88, 100.0, 101.0	60.7, 39.93, 47.23
Construction	2011-01- 2012-12	hawkes	1301.08, 0.01, 0.05	1285.85, 0.0, 0.0
Construction	2011-01- 2012-12	hawkes	1301.08, 0.1, 0.3	1301.08, 0.0, 2.96
Construction	2011-01- 2012-12	hawkes	1301.08, 1.0, 2.0	529.9, 12.85, 21.62
Construction	2011-01- 2012-12	hawkes	1301.08, 10.0, 15.0	290.41, 28.67, 36.84
Construction	2011-01- 2012-12	hawkes	1301.08, 100.0, 101.0	112.36, 68.38, 74.76
Enseignement, ...	2011-01- 2012-12	hawkes	428.62, 0.01, 0.05	419.43, 0.0, 0.0
Enseignement, ...	2011-01- 2012-12	hawkes	428.62, 0.1, 0.3	401.41, 0.07, 1.12
Enseignement, ...	2011-01- 2012-12	hawkes	428.62, 1.0, 2.0	331.47, 5.36, 23.58
Enseignement, ...	2011-01- 2012-12	hawkes	428.62, 10.0, 15.0	141.25, 15.64, 23.27
Enseignement, ...	2011-01- 2012-12	hawkes	428.62, 100.0, 101.0	64.88, 36.95, 43.48
Services marchands	2011-01- 2012-12	hawkes	2802.83, 0.01, 0.05	2710.34, 0.0, 0.0
Services marchands	2011-01- 2012-12	hawkes	2802.83, 0.1, 0.3	2704.88, 0.0, 0.0
Services marchands	2011-01- 2012-12	hawkes	2802.83, 1.0, 2.0	882.33, 20.52, 29.9
Services marchands	2011-01- 2012-12	hawkes	2802.83, 10.0, 15.0	453.97, 49.44, 58.93
Services marchands	2011-01- 2012-12	hawkes	2802.83, 100.0, 101.0	176.61, 104.94, 111.92

Secteur	Historique	Répartition des temps de sauts	θ_{start}	θ
Agriculture	2011-01- 2012-12	hawkes	51.23, 0.01, 0.05	91.7, 0.14, 1.27
Agriculture	2011-01- 2012-12	hawkes	10.25, 0.1, 0.3	89.27, 0.15, 1.14
Agriculture	2011-01- 2012-12	hawkes	1.02, 1.0, 2.0	52.88, 51.06, 105.36
Agriculture	2011-01- 2012-12	hawkes	0.2, 10.0, 15.0	32.27, 1648.49, 2405.3
Agriculture	2011-01- 2012-12	hawkes	0.1, 100.0, 101.0	11.66, 4469.0, 5024.53
Industrie	2011-01- 2012-12	hawkes	193.94, 0.01, 0.05	383.04, 0.0, 0.0
Industrie	2011-01- 2012-12	hawkes	38.79, 0.1, 0.3	270.88, 3.98, 13.15
Industrie	2011-01- 2012-12	hawkes	3.88, 1.0, 2.0	0.0, 31.76, 31.71
Industrie	2011-01- 2012-12	hawkes	0.78, 10.0, 15.0	0.0, 215.2, 215.18
Industrie	2011-01- 2012-12	hawkes	0.39, 100.0, 101.0	23.28, 4750.69, 5036.98
Construction	2011-01- 2012-12	hawkes	650.54, 0.01, 0.05	1300.93, 0.0, 0.0
Construction	2011-01- 2012-12	hawkes	130.11, 0.1, 0.3	1300.93, 0.0, 0.0
Construction	2011-01- 2012-12	hawkes	13.01, 1.0, 2.0	0.0, 42.28, 42.23
Construction	2011-01- 2012-12	hawkes	2.6, 10.0, 15.0	0.0, 539.67, 539.54
Construction	2011-01- 2012-12	hawkes	1.3, 100.0, 101.0	0.0, 1555.12, 1578.15
Enseignement, ...	2011-01- 2012-12	hawkes	214.31, 0.01, 0.05	419.37, 0.0, 0.0
Enseignement, ...	2011-01- 2012-12	hawkes	42.86, 0.1, 0.3	279.89, 5.91, 16.99
Enseignement, ...	2011-01- 2012-12	hawkes	4.29, 1.0, 2.0	0.0, 26.87, 26.82
Enseignement, ...	2011-01- 2012-12	hawkes	0.86, 10.0, 15.0	0.0, 231.22, 230.72
Enseignement, ...	2011-01- 2012-12	hawkes	0.43, 100.0, 101.0	35.64, 2800.63, 3052.59
Services marchands	2011-01- 2012-12	hawkes	1401.42, 0.01, 0.05	2708.53, 0.0, 0.0
Services marchands	2011-01- 2012-12	hawkes	280.28, 0.1, 0.3	1062.69, 16.32, 26.23
Services marchands	2011-01- 2012-12	hawkes	28.03, 1.0, 2.0	0.0, 46.26, 46.21
Services marchands	2011-01- 2012-12	hawkes	5.61, 10.0, 15.0	0.0, 835.61, 835.6
Services marchands	2011-01- 2012-12	hawkes	2.8, 100.0, 101.0	0.0, 1098.76, 1104.49

A.4.2 Pré-calibrage

Secteur	Historique	θ_{start}	θ
Agriculture	2010-01- 2019-12	113.8, 0.0, 0.0	101.67, 0.0, 0.01
Agriculture	2010-01- 2019-12	113.8, 0.01, 0.05	101.67, 0.0, 0.01
Agriculture	2010-01- 2019-12	113.8, 0.1, 0.3	101.67, 0.0, 0.01
Agriculture	2010-01- 2019-12	113.8, 0.5, 0.8	87.7, 0.16, 0.68
Agriculture	2010-01- 2019-12	113.8, 1.0, 2.0	113.73, 0.0, 0.0
Agriculture	2010-01- 2019-12	113.8, 2.0, 4.0	113.73, 0.0, 0.0
Agriculture	2010-01- 2019-12	113.8, 3.0, 4.0	113.73, 0.0, 0.0
Agriculture	2010-01- 2019-12	113.8, 5.0, 9.0	113.73, 0.0, 0.0
Agriculture	2010-01- 2019-12	113.8, 10.0, 15.0	113.73, 0.0, 0.0
Agriculture	2010-01- 2019-12	113.8, 100.0, 101.0	113.8, 0.0, 3494817.26
Industrie	2010-01- 2019-12	363.18, 0.0, 0.0	143.52, 1.21, 1.99
Industrie	2010-01- 2019-12	363.18, 0.01, 0.05	143.52, 1.21, 1.99
Industrie	2010-01- 2019-12	363.18, 0.1, 0.3	143.52, 1.21, 1.99
Industrie	2010-01- 2019-12	363.18, 0.5, 0.8	143.52, 1.21, 1.99
Industrie	2010-01- 2019-12	363.18, 1.0, 2.0	143.52, 1.21, 1.99
Industrie	2010-01- 2019-12	363.18, 2.0, 4.0	143.52, 1.21, 1.99
Industrie	2010-01- 2019-12	363.18, 3.0, 4.0	143.52, 1.21, 1.99
Industrie	2010-01- 2019-12	363.18, 5.0, 9.0	363.18, 0.0, 13.57
Industrie	2010-01- 2019-12	363.18, 10.0, 15.0	363.17, 0.0, 0.29
Industrie	2010-01- 2019-12	363.18, 100.0, 101.0	363.18, 0.0, 509.11
Construction	2010-01- 2019-12	1200.72, 0.0, 0.0	1200.65, 0.0, 0.18
Construction	2010-01- 2019-12	1200.72, 0.01, 0.05	223.7, 3.28, 4.02
Construction	2010-01- 2019-12	1200.72, 0.1, 0.3	223.7, 3.28, 4.02
Construction	2010-01- 2019-12	1200.72, 0.5, 0.8	223.7, 3.28, 4.02
Construction	2010-01- 2019-12	1200.72, 1.0, 2.0	223.7, 3.28, 4.02
Construction	2010-01- 2019-12	1200.72, 2.0, 4.0	223.7, 3.28, 4.02
Construction	2010-01- 2019-12	1200.72, 3.0, 4.0	223.7, 3.28, 4.02
Construction	2010-01- 2019-12	1200.72, 5.0, 9.0	223.7, 3.28, 4.02
Construction	2010-01- 2019-12	1200.72, 10.0, 15.0	223.7, 3.28, 4.02
Construction	2010-01- 2019-12	1200.72, 100.0, 101.0	1200.71, 0.0, 264.04
Enseignement, ...	2010-01- 2019-12	444.69, 0.0, 0.0	381.52, 0.01, 0.07
Enseignement, ...	2010-01- 2019-12	444.69, 0.01, 0.05	381.52, 0.01, 0.07
Enseignement, ...	2010-01- 2019-12	444.69, 0.1, 0.3	381.52, 0.01, 0.07
Enseignement, ...	2010-01- 2019-12	444.69, 0.5, 0.8	284.9, 0.39, 1.08
Enseignement, ...	2010-01- 2019-12	444.69, 1.0, 2.0	284.9, 0.39, 1.08
Enseignement, ...	2010-01- 2019-12	444.69, 2.0, 4.0	444.42, 0.0, 0.0
Enseignement, ...	2010-01- 2019-12	444.69, 3.0, 4.0	444.43, 0.0, 0.0
Enseignement, ...	2010-01- 2019-12	444.69, 5.0, 9.0	444.43, 0.0, 0.0
Enseignement, ...	2010-01- 2019-12	444.69, 10.0, 15.0	444.69, 0.0, 19.92
Enseignement, ...	2010-01- 2019-12	444.69, 100.0, 101.0	444.69, 0.0, 527.26

Secteur	Historique	θ_{start}	θ
Services marchands	2010-01- 2019-12	2763.6, 0.0, 0.0	2763.33, 0.0, 0.09
Services marchands	2010-01- 2019-12	2763.6, 0.01, 0.05	1018.23, 3.08, 4.88
Services marchands	2010-01- 2019-12	2763.6, 0.1, 0.3	1018.24, 3.08, 4.88
Services marchands	2010-01- 2019-12	2763.6, 0.5, 0.8	1018.23, 3.08, 4.88
Services marchands	2010-01- 2019-12	2763.6, 1.0, 2.0	1018.24, 3.08, 4.88
Services marchands	2010-01- 2019-12	2763.6, 2.0, 4.0	1018.23, 3.08, 4.88
Services marchands	2010-01- 2019-12	2763.6, 3.0, 4.0	1018.24, 3.08, 4.88
Services marchands	2010-01- 2019-12	2763.6, 5.0, 9.0	1018.23, 3.08, 4.88
Services marchands	2010-01- 2019-12	2763.6, 10.0, 15.0	1018.23, 3.08, 4.88
Services marchands	2010-01- 2019-12	2763.6, 100.0, 101.0	2763.59, 0.0, 227.38
Agriculture	2010-01- 2019-12	56.9, 0.0, 0.0	101.67, 0.0, 0.01
Agriculture	2010-01- 2019-12	11.38, 0.01, 0.05	101.67, 0.0, 0.01
Agriculture	2010-01- 2019-12	5.69, 0.1, 0.3	0.0, 1.36, 1.35
Agriculture	2010-01- 2019-12	5.69, 0.5, 0.8	87.7, 0.16, 0.68
Agriculture	2010-01- 2019-12	2.28, 1.0, 2.0	101.67, 0.0, 0.01
Agriculture	2010-01- 2019-12	1.14, 2.0, 4.0	87.7, 0.16, 0.68
Agriculture	2010-01- 2019-12	1.14, 3.0, 4.0	101.67, 0.0, 0.01
Agriculture	2010-01- 2019-12	0.23, 5.0, 9.0	101.67, 0.0, 0.01
Agriculture	2010-01- 2019-12	0.16, 10.0, 15.0	87.7, 0.16, 0.68
Agriculture	2010-01- 2019-12	0.11, 100.0, 101.0	87.7, 0.16, 0.68
Industrie	2010-01- 2019-12	181.59, 0.0, 0.0	362.97, 0.0, 0.0
Industrie	2010-01- 2019-12	36.32, 0.01, 0.05	363.17, 0.0, 0.29
Industrie	2010-01- 2019-12	18.16, 0.1, 0.3	0.0, 2.46, 2.45
Industrie	2010-01- 2019-12	18.16, 0.5, 0.8	143.52, 1.21, 1.99
Industrie	2010-01- 2019-12	7.26, 1.0, 2.0	143.52, 1.21, 1.99
Industrie	2010-01- 2019-12	3.63, 2.0, 4.0	143.52, 1.21, 1.99
Industrie	2010-01- 2019-12	3.63, 3.0, 4.0	143.52, 1.21, 1.99
Industrie	2010-01- 2019-12	0.73, 5.0, 9.0	143.52, 1.21, 1.99
Industrie	2010-01- 2019-12	0.52, 10.0, 15.0	143.52, 1.21, 1.99
Industrie	2010-01- 2019-12	0.36, 100.0, 101.0	143.52, 1.21, 1.99
Construction	2010-01- 2019-12	600.36, 0.0, 0.0	1200.65, 0.0, 0.18
Construction	2010-01- 2019-12	120.07, 0.01, 0.05	223.7, 3.28, 4.02
Construction	2010-01- 2019-12	60.04, 0.1, 0.3	0.0, 4.54, 4.53
Construction	2010-01- 2019-12	60.04, 0.5, 0.8	223.7, 3.28, 4.02
Construction	2010-01- 2019-12	24.01, 1.0, 2.0	0.0, 4.54, 4.53
Construction	2010-01- 2019-12	12.01, 2.0, 4.0	223.7, 3.28, 4.02
Construction	2010-01- 2019-12	12.01, 3.0, 4.0	223.7, 3.28, 4.02
Construction	2010-01- 2019-12	2.4, 5.0, 9.0	223.7, 3.28, 4.02
Construction	2010-01- 2019-12	1.72, 10.0, 15.0	223.7, 3.28, 4.02
Construction	2010-01- 2019-12	1.2, 100.0, 101.0	223.7, 3.28, 4.02

Secteur	Historique	θ_{start}	θ
Enseignement, ...	2010-01- 2019-12	222.35, 0.0, 0.0	381.52, 0.01, 0.07
Enseignement, ...	2010-01- 2019-12	44.47, 0.01, 0.05	381.52, 0.01, 0.07
Enseignement, ...	2010-01- 2019-12	22.23, 0.1, 0.3	284.9, 0.39, 1.08
Enseignement, ...	2010-01- 2019-12	22.23, 0.5, 0.8	284.9, 0.39, 1.08
Enseignement, ...	2010-01- 2019-12	8.89, 1.0, 2.0	284.9, 0.39, 1.08
Enseignement, ...	2010-01- 2019-12	4.45, 2.0, 4.0	284.9, 0.39, 1.08
Enseignement, ...	2010-01- 2019-12	4.45, 3.0, 4.0	284.9, 0.39, 1.08
Enseignement, ...	2010-01- 2019-12	0.89, 5.0, 9.0	284.9, 0.39, 1.08
Enseignement, ...	2010-01- 2019-12	0.64, 10.0, 15.0	284.9, 0.39, 1.08
Enseignement, ...	2010-01- 2019-12	0.44, 100.0, 101.0	284.9, 0.39, 1.08
Services marchands	2010-01- 2019-12	1381.8, 0.0, 0.0	2763.33, 0.0, 0.09
Services marchands	2010-01- 2019-12	276.36, 0.01, 0.05	2763.33, 0.0, 0.09
Services marchands	2010-01- 2019-12	138.18, 0.1, 0.3	1018.23, 3.08, 4.88
Services marchands	2010-01- 2019-12	138.18, 0.5, 0.8	0.0, 6.3, 6.29
Services marchands	2010-01- 2019-12	55.27, 1.0, 2.0	0.0, 6.3, 6.29
Services marchands	2010-01- 2019-12	27.64, 2.0, 4.0	1018.23, 3.08, 4.88
Services marchands	2010-01- 2019-12	27.64, 3.0, 4.0	1018.24, 3.08, 4.88
Services marchands	2010-01- 2019-12	5.53, 5.0, 9.0	1018.22, 3.08, 4.88
Services marchands	2010-01- 2019-12	3.95, 10.0, 15.0	1018.23, 3.08, 4.88
Services marchands	2010-01- 2019-12	2.76, 100.0, 101.0	1018.23, 3.08, 4.88