Mémoire présenté devant l'ENSAE Paris
pour l'obtention du diplôme de la filière Actuariat
et l'admission à l'Institut des Actuaires
le 10/03/2022

Par : **Nicolas Richard**

Titre : **Cash flow prediction in reinsurance: short and long term projection for cash requirement and FX risk management**

Confidentialité : ☒ NON ☐ OUI (Durée : ☐ 1 an ☐ 2 ans)

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

Membres présents du jury de la filière

Nom : Pierre Picard

Membres présents du jury de l'Institut des Actuaires

Entreprise : AXA Global Reinsurance

Signature :

Directeur du mémoire en entreprise :

Nom : Elie DADOUN
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)

Signature du responsable entreprise

Secrétariat:

Signature du candidat

Bibliothèque:

Ecole Nationale de la Statistique et de l'Administration Economique (ENSAE)
5, avenue Henry Le Chatelier - 91120 PALAISEAU, FRANCE

Abstract

Reinsurance companies are particularly exposed to FX risk, mainly because of the geographical dispersion of covered risks on an international scale, which involves that they deal in a large number of currencies. Because of the random nature of its liabilities, it is extremely difficult for a reinsurer to forecast exactly the dates and amounts involved in the transformation of those liabilities into effective settlements, thus allowing it to resort to perfect hedging through economic congruence. An ideal situation would be an assets/liabilities close-to-perfect matching (in amounts, durations and currencies), which would ensure the exact liquidity needed in each currency at the time the future cash flows take place. Being able to predict to a certain extent the future inflows and outflows in currencies would not only allow an optimal hedging of FX risk as well as the opportunity to devise optimal investment strategies of the available cash. This project research is an attempt towards this direction. On the basis of AXA Global Reinsurance portfolio's contractual, accounting, and financial historical data we aim at forecasting the timing and amounts of future technical cash flows labelled in currencies. We broke down this ambitious goal into two main axes of development: first the prediction of near future monthly cash transactions (1-year horizon), and second, the prediction of future yearly run-off cash flows stemming from existing technical reserves. We present in this thesis a prospective roadmap showing original paths to reach the above objectives. Some roads have been marked out with clear waypoints and are a priori unobstructed while others are still under work.

Key words : P&C reinsurance, claims reserving, FX risk, survival analysis, duration model, Cox proportional-hazards model, Monte Carlo method, Long Short-Term Memory (LSTM) neural network, credibility theory, Bühlmann-Straub model, hierarchical credibility model, IBNR allocation, cash flows prediction, reserves projection.

Résumé

Les sociétés de réassurance sont particulièrement exposées au risque de change, notamment en raison de la dispersion géographique des risques couverts à l'échelle internationale, ce qui implique qu'elles traitent dans un grand nombre de devises. En raison du caractère aléatoire de ses engagements, il est extrêmement difficile pour un réassureur de prévoir exactement les dates et les montants faisant l'objet de la transformation de ces engagements en règlements effectifs, lui permettant ainsi de recourir à une couverture parfaite par congruence économique. Une situation idéale serait une concordance actif/passif proche de la perfection (en montants, durées et devises), qui assurerait la liquidité exacte nécessaire dans chaque devise au moment où les flux de trésorerie futurs ont lieu. Être capable de prédire dans une certaine mesure les entrées et sorties futures en devises permettrait non seulement une couverture optimale du risque de change ainsi que la possibilité de concevoir des stratégies d'investissement optimales des liquidités disponibles. Ce projet de recherche constitue un effort dans cette direction. Sur la base des données historiques contractuelles, comptables et financières du portefeuille d'AXA Global Reinsurance, nous visons à prévoir le calendrier et les montants des futurs flux de trésorerie techniques libellés en devises. Nous avons décomposé cet objectif ambitieux en deux axes principaux de développement : d'abord la prédiction des transactions de trésorerie mensuelles dans un futur proche (horizon 1 an), et deuxièmement, la prédiction des futurs flux de trésorerie annuels issus des provisions techniques existantes. Nous présentons dans ce mémoire une feuille de route prospective révélant des voies d'accès originales aux objectifs ci-dessus. Certaines routes ont été clairement balisées et sont a priori dégagées tandis que d'autres sont encore en travaux.

Mots clés : réassurance IARD, provisionnement des sinistres, risque de change, analyse de survie, modèle de durée, regression de Cox (modèle de Cox à risques proportionnels), méthode de Monte Carlo, réseau de neurones Long Short-Term Memory (LSTM), théorie de la crédibilité, modèle de Bühlmann-Straub, modèle de crédibilité hiérarchique, allocation d'IBNR, prédiction des flux et besoins de trésorerie, projection des provisions.

Acknowledgment

I would like to express my sincere gratitude to my supervisor Elie DADOUN for its assistance at every stage of this research project as well as for its insightful comments and suggestions. As AXA Global Re reserving team manager, he showed to be highly competent, available and prompt to help whenever needed. I could not have expected more of a mentor to guide me through this challenging task.

My gratitude extends to the Ecole Nationale de la Statistique et de l'Administration Economique (ENSAE) for their high-quality formation in actuaries sciences.

Executive summary

Introduction

Reinsurance companies core business, which is the transfer of risks taking place at an international scale, entails many transactions of cash made in numerous currencies and linked to, among others, general expenses, taxes, financial investments, brokerage fees, minimum and deposit premiums, adjustment premiums, commissions and claims settlements. Some of these cash flows are well anticipated by reinsurers' treasury teams, while others such as adjustment premiums, commissions and claims payments generally escape previsions, leading to poor forecasts of cash needs in currencies which gets in the way of satisfactory FX risk hedging as well as optimal investment of the available cash. Even when focusing only on their run-off estimated liabilities development into claims payment cash flows, reinsurers may encounter difficulties estimating the amounts and occurrence times of those future cash flows in each currency.

AXA Global Reinsurance (AGRe), as any other reinsurance entity, has to face the challenging issue of predicting technical inflows and outflows specifics to the reinsurance business, in a large set of transactional currencies, whether on the risks acceptance business side (inward reinsurance treaties between AXA entities and AGRe) or on the retrocession one (outward reinsurance treaties between AGRe and external reinsurers).

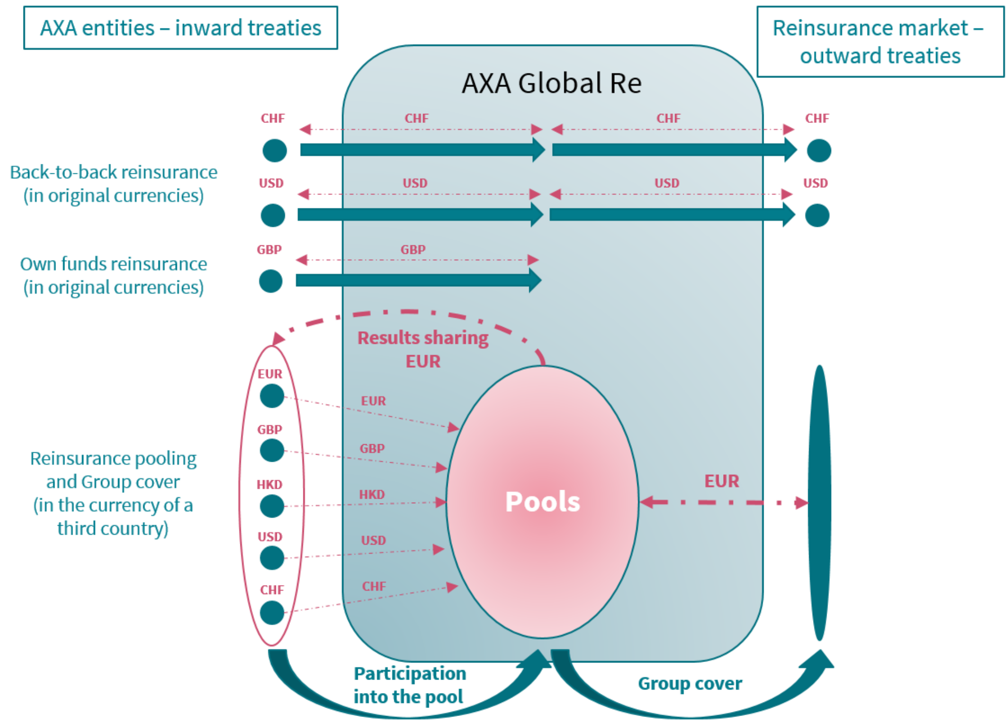


Figure 1: Currency flows in the light of AGRe risk acceptance and retrocession mechanisms

Approach

We have decomposed the complex problem of forecasting future technical cash flows into two main objectives, namely a 12-month horizon monthly cash flows prediction and a run-off projection of current reserves into yearly cash flows.

Short-term monthly technical cash flows prediction

We strove to achieve the first goal, that is the pursuit of predicting cash flows related to technical accounts materializing within the next twelve months (on both the acceptance and retrocession sides of AGRe's business), resorting to the following hypothesis, and designing a strategy comprised of 3 complementary sub-achievements:

Hypothesis: Future cash flows within the next twelve months stem from either existing but not yet settled (EBNYS) balances, or existing technical reserves, or last first notice and evaluations of claims and cash calls (not yet materialized as reserves nor considered in AGRe's accounting system).

Sub-goals:

1. Prediction of settlements issued from EBNYS technical balances thanks to duration models (Kaplan-Meier and Cox regression) calibrated on the basis of AGRe historical balances;

2. Prediction of not yet existing technical balances settlements issued from existing technical reserves and/or their past variations: for lack of reliable historical data at the time of this thesis' writing, we will present the devised methodologies, implemented models (GLM, ratios model and long short-term memory neural network model), and more generally the direction taken to that end, without numerical results;

3. Inclusion of losses first notices (recent claims and cash calls evaluations) liable to be settled within next few months but not yet contemplated in AGRe's accounting and financial systems: integrated "manually", in conformity with experts judgments, in the predictions made by the two previous courses of action.

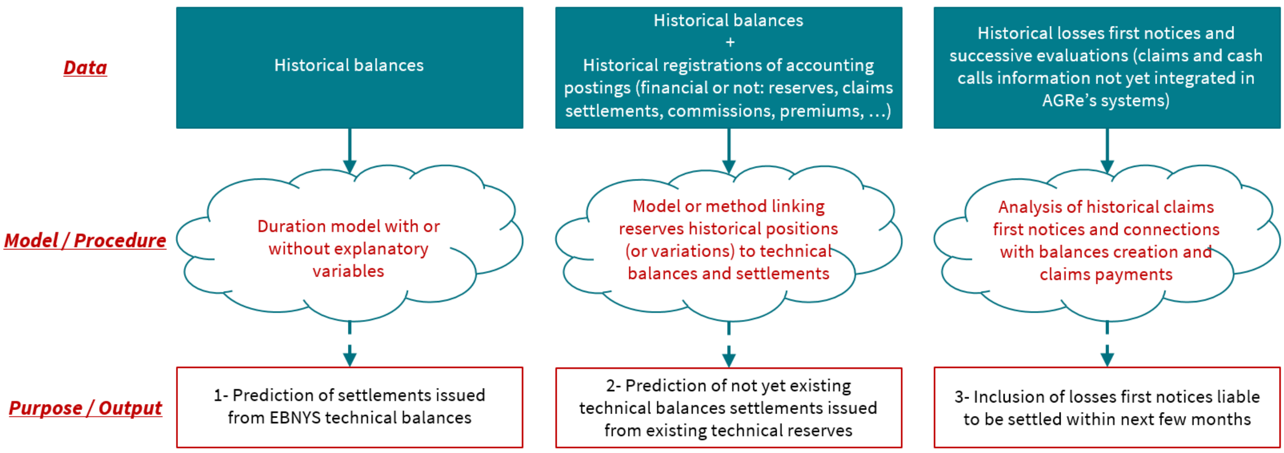


Figure 2: Short-term cash flows prediction goal breakdown - models structure and data sources

Long-term run-off yearly technical cash flows prediction

A perspective shift separates the anticipation of near future cash needs in currencies, and long-term projection of currently estimated liabilities into expected yearly claims payments in distinct currencies. Nonetheless, solving one question or the other answers the same desiderata (at different time horizon scales), which are being able to match assets with liabilities for each currency and each period of time considered, and as a consequence, to know how much and how long can one invest the unused cash in currencies at hand.

A projection of future yearly cash flows in currencies originating from existing technical reserves was already carried out to measure the market risk linked to AGRe's run-off liabilities, but presented some drawbacks. This old procedure quantified a stock of reserves to be projected by reserving segment and currency according to a questionable IBNR allocation method. Furthermore, it only took into account payment patterns at the reserving segment level to project those reserves, without distinguishing possible differences in development behaviors of liabilities labelled in different currencies. Finally, the old projection scheme handled case reserves and IBNR in the same manner without differentiating their natures and rhythms of development into claims payments.

Those observations constitute as many improvement targets that we took into consideration in the formulation of a new methodology. We list bellow the main steps we took on the trajectory to enhance current run-off provisions projection:

1. Application of credibility theory to claims development factors, in order to produce individually credibilized development patterns for each currency inside a given reserving segment, i.e. for each group of reinsurance treaties categorized in a given reserving segment and labelled in the same currency;

2. Allocation of IBNR reserves down to currency classes within AGRe's reserving segments through a Bornhuetter-Ferguson approach and counting on individually credibilized loss development factors;

3. Elaboration of new projection methods taking account of the a priory different behaviors in the liquidation patterns of file-to-file reserves and IBNR ones, as well as the information contained in both claims payment and claims incurred patterns specifics to each currency inside a reserving segment;

Results

Throughout this thesis we came across interesting intermediary results and observations, such as:

- highlight of the differentiation in the development speed of existing balances settlements, according to their characteristics (amount class in the example below);

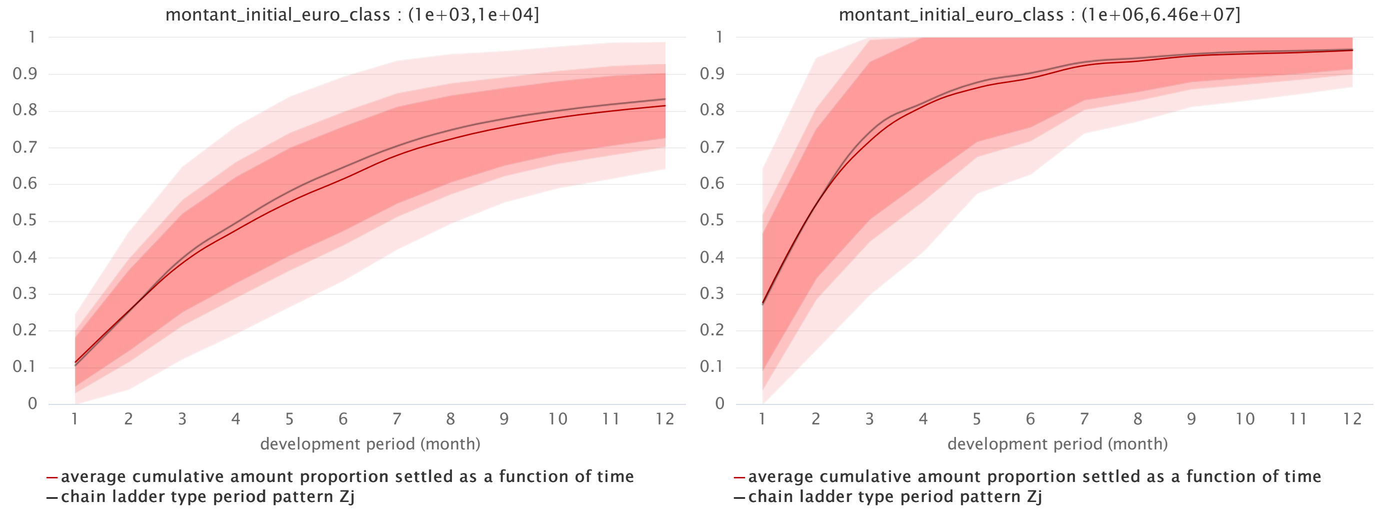


Figure 3: Average (over attaching quarter) cumulative settled amount proportion as a function of time (month) with confidence intervals (68%, 80% and 95% CI) - from 1k to 10k euros absolute amount balances (left) and more than 1M euros absolute amount balances (right)

- identification and management of right-censure in historical balances data, along with original implementation of survival analysis originating models to the prediction of existing balances settlements;
- historical financial and non-financial statements data ingenious rearrangement, together with relevant and encouraging proposals of models set-ups as endeavors to predict not yet existing balances and their settlement cash flows;
- successful application of credibility theory (Bühlmann—Straub model) to loss and reserves development analysis;

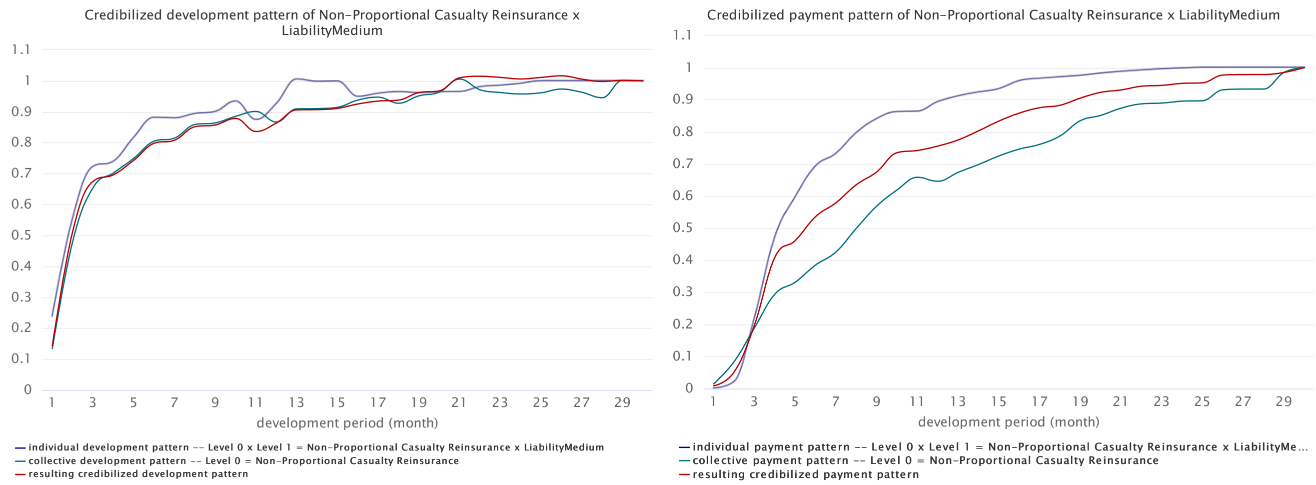


Figure 4: Developments of loss incurred (on the left) and claims payments (on the right) issued from individual, collective and credibilized development factors – Level 0 (SOLVENCY II BUSINESS LINE) value X Level 1 (LR SCOPE) value = Non-Proportional Casualty Reinsurance X LiabilityMedium

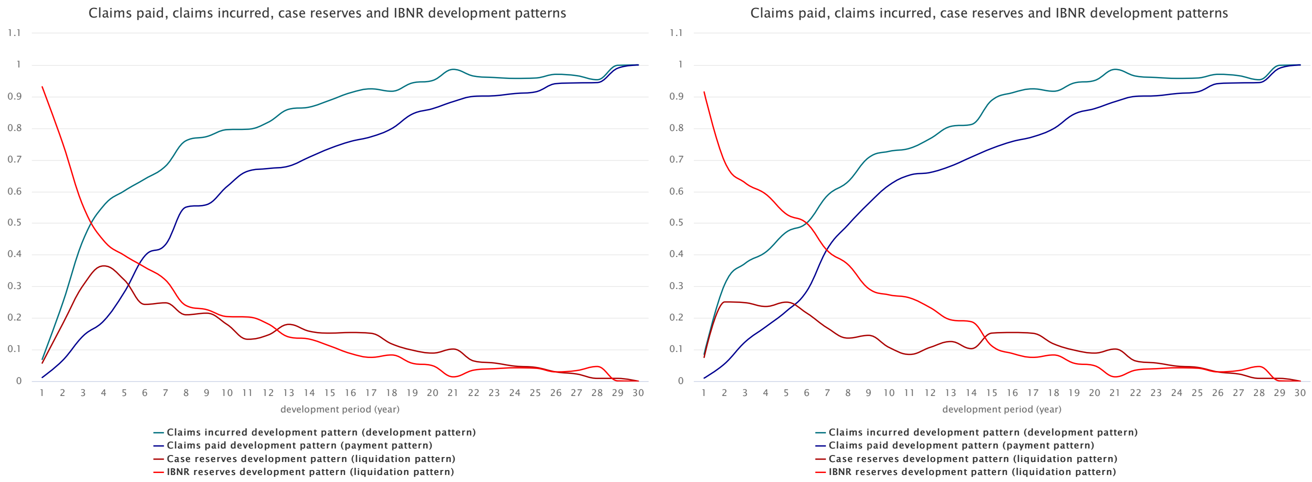


Figure 5: Credibilized developments of claims incurred, claims paid, case reserves, and IBNR – Level 0 (SOLVENCY II BUSINESS LINE) value X Level 1 (LR SCOPE) value = Non-Proportional Casualty Reinsurance X PoolLiability (on the left) and Non-Proportional Casualty Reinsurance X PoolMotor (on the right)

- innovative IBNR reserves allocation procedure to smaller groups of reinsurance treaties than the reserving segments;
- design of new methodologies of run-off liabilities projection into yearly cash flows;

Each of those in-between steps allowed us to reach particularly satisfactory monthly cash flows predictions in currencies related to positive-amounts-only or negative-amounts-only existing balances, as well run-off reserves linked yearly cash flows predictions generally closer to real ones than currently applied method.

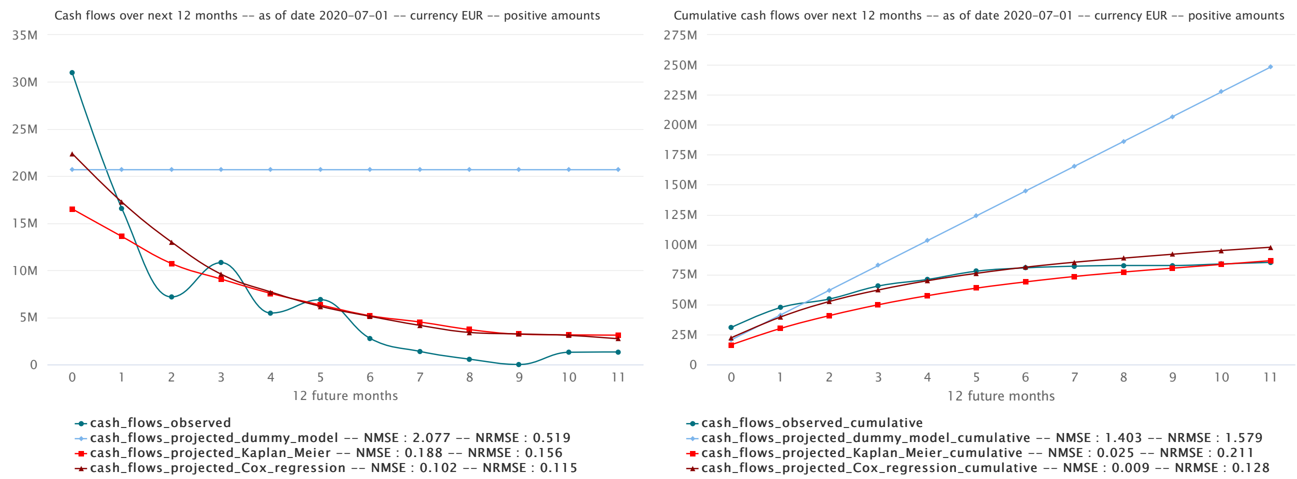


Figure 6: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2020-07-01 – currency EUR – positive amounts

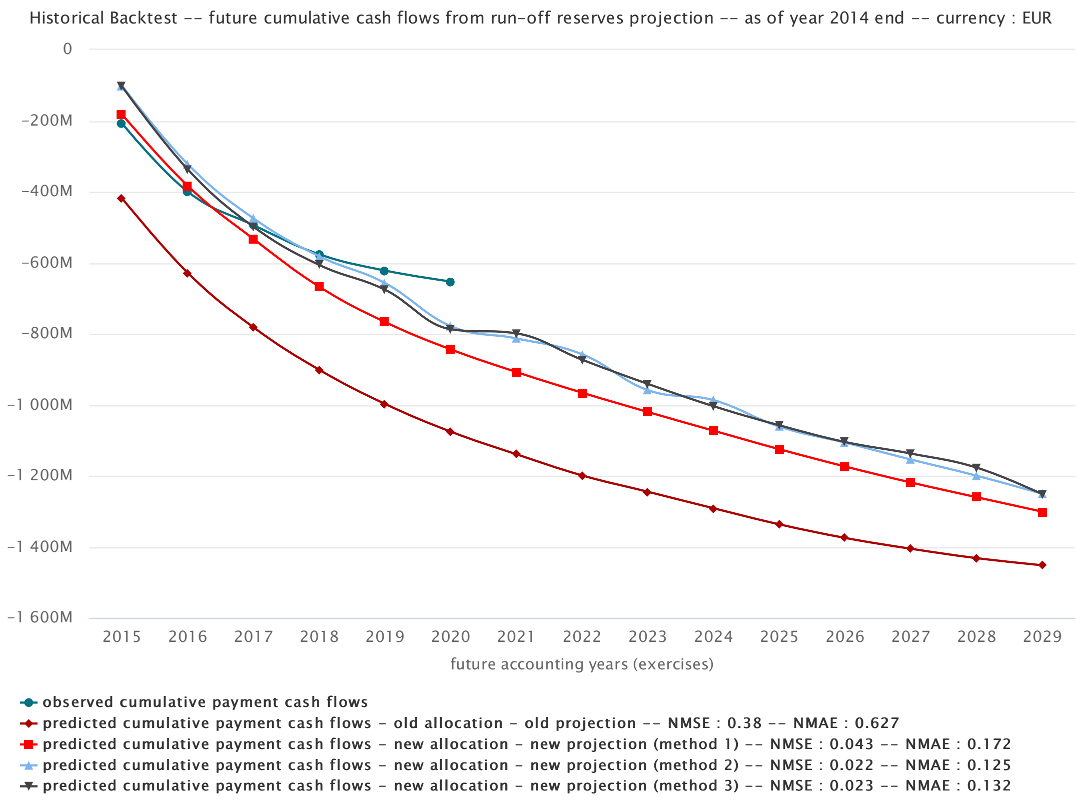


Figure 7: Reserves projections into cumulative yearly cash flows vs actually observed claims payment cumulative cash flows – as of 2014 – currency EUR

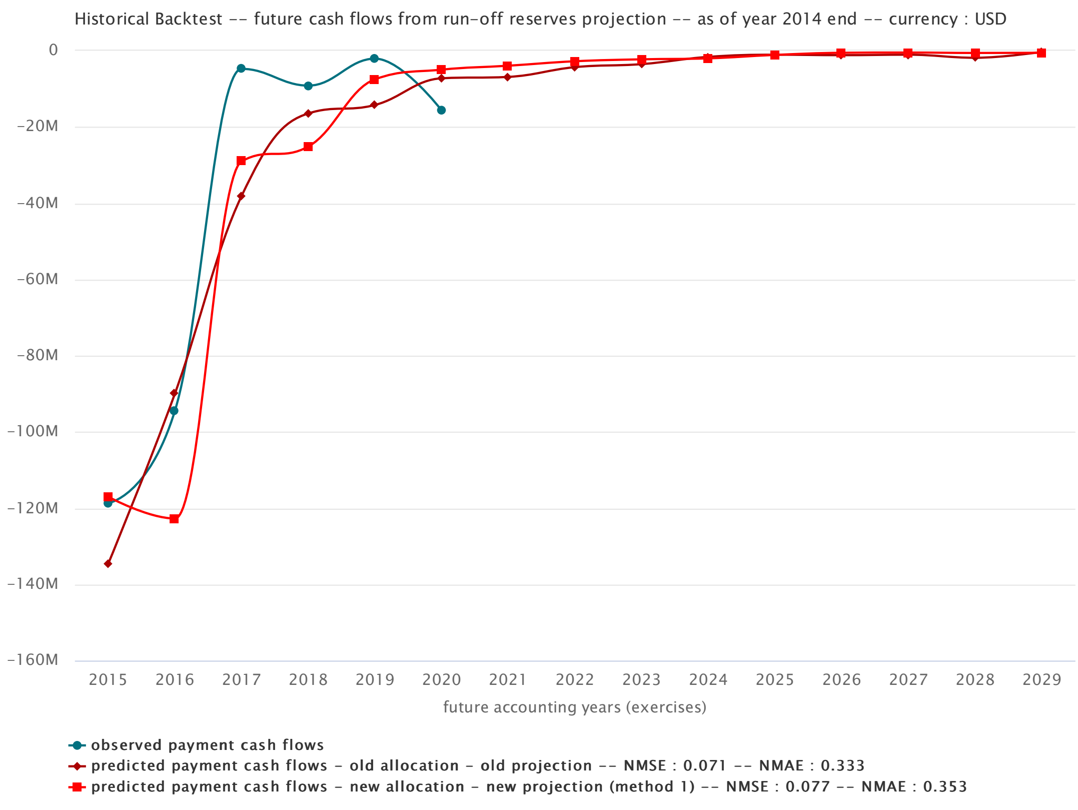


Figure 8: Reserves projections into yearly cash flows vs actually observed claims payment yearly cash flows – as of 2014 – currency USD

Note de Synthèse

Problématique

Le cœur de métier des sociétés de réassurance, qui est le transfert des risques s'effectuant à l'échelle internationale, implique de nombreuses transactions de trésorerie effectuées dans de nombreuses devises et liées, entre autres, aux frais généraux, impôts, placements financiers, frais de courtage, primes minimales et de dépôt, primes d'ajustement, commissions et règlements de sinistres. Certains de ces flux de trésorerie sont plutôt bien anticipés par les équipes de trésorerie des réassureurs, tandis que d'autres tels que les primes d'ajustement, les commissions et les règlements de sinistres échappent généralement aux prévisions, entraînant de mauvaises prédictions des besoins de liquidités en devises. Cela entrave la mise en place d'une couverture satisfaisante contre le risque de change, de même que l'implémentation d'une stratégie optimale d'investissement des liquidités disponibles. Y compris en se concentrant uniquement sur l'évolution de leurs passifs estimés à date, progressivement liquidés en flux de trésorerie de règlement des sinistres, les réassureurs rencontrent des difficultés pour estimer les montants et les occurrences de ces flux futurs dans chaque devise.

AXA Global Reinsurance (AGRe), comme toute autre entité de réassurance, doit faire face à la difficulté de prévoir les entrées et sorties de flux techniques spécifiques à l'activité de réassurance et exprimés dans un large éventail de devises transactionnelles, que ce soit du côté de l'activité d'acceptation des risques (traités de réassurance entrants entre les entités d'AXA et AGRe) ou du côté de leur rétrocession (traités de cession de risques entre AGRe et des réassureurs externes).

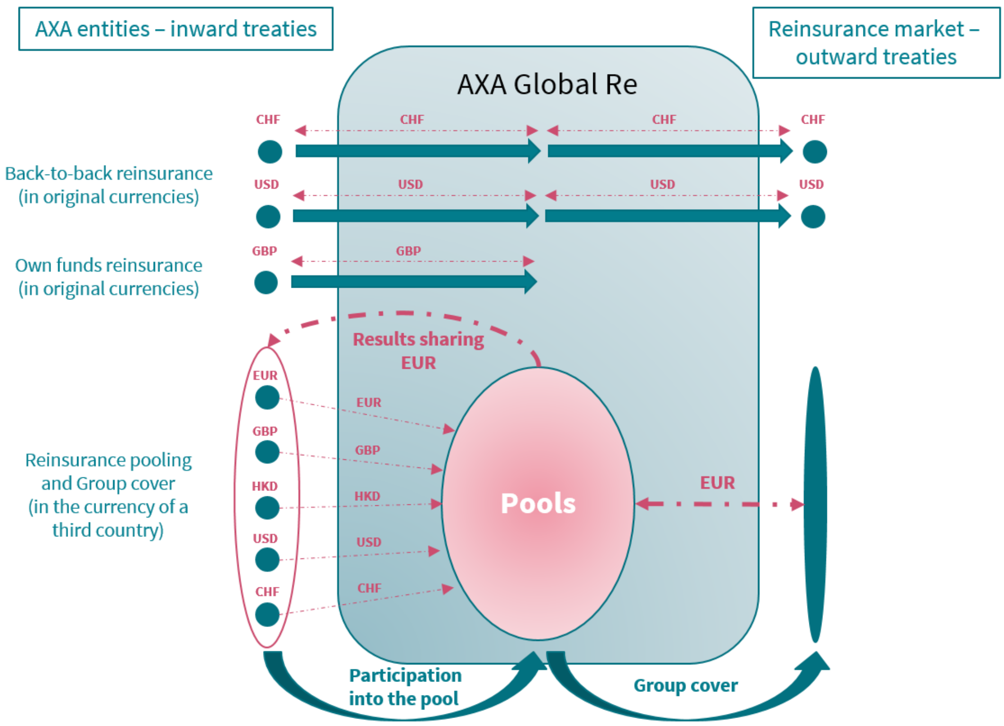


Figure 9: Flux de devises au regard des mécanismes d'acceptation et de rétrocession des risques d'AGRe

Démarche

Nous avons décomposé le problème complexe de la prévision des flux de trésorerie techniques futurs en deux objectifs principaux, à savoir une prévision des flux de trésorerie mensuels sur un horizon de 12 mois et une projection de liquidation des provisions techniques actuelles en flux de trésorerie annuels.

Prévision des flux de trésorerie techniques mensuels à court terme

Nous nous sommes efforcés d'atteindre le premier objectif, à savoir la prévision des flux de trésorerie liés aux comptes techniques se matérialisant dans les douze prochains mois (tant du côté de l'acceptation que de la rétrocession de risques au sein d'AGRe), en recourant à l'hypothèse suivante et en concevant une stratégie composée de 3 étapes complémentaires :

Hypothèse : Les flux de trésorerie futurs au cours des douze prochains mois proviennent soit des soldes existants mais non encore réglés (EBNYS), soit des réserves techniques existantes, soit des derniers avis et évaluations de sinistres et appels au comptant (non encore matérialisés sous forme de provisions, ni pris en compte dans les systèmes comptables d'AGRe).

Sous-objectifs :

1. Prédiction des règlements issus des soldes techniques EBNYS grâce à des modèles de durée (Kaplan-Meier et régression de Cox) calibrés sur la base des soldes historiques d'AGRe ;

2. Prédiction des règlements des soldes techniques non encore existants issus des provisions techniques existantes à date et/ou de leurs variations passées : faute de données historiques fiables au moment de la rédaction de ce mémoire, nous présenterons les méthodologies élaborées, les modèles mis en œuvre (GLM, modèle de ratios et modèle de réseau de neurones LSTM), et plus généralement la direction prise pour tenter d'atteindre ce deuxième sous-objectif, sans exposer de résultats numériques ;

3. Prise en considération des premiers avis de sinistres (premières évaluations de sinistres recents et appels de fonds) susceptibles d'être réglés dans les prochains mois mais pas encore envisagés dans les systèmes comptables et financiers d'AGRe : intégrés "manuellement", conformément à des jugements d'experts, aux prévisions réalisées suites aux deux étapes précédentes.

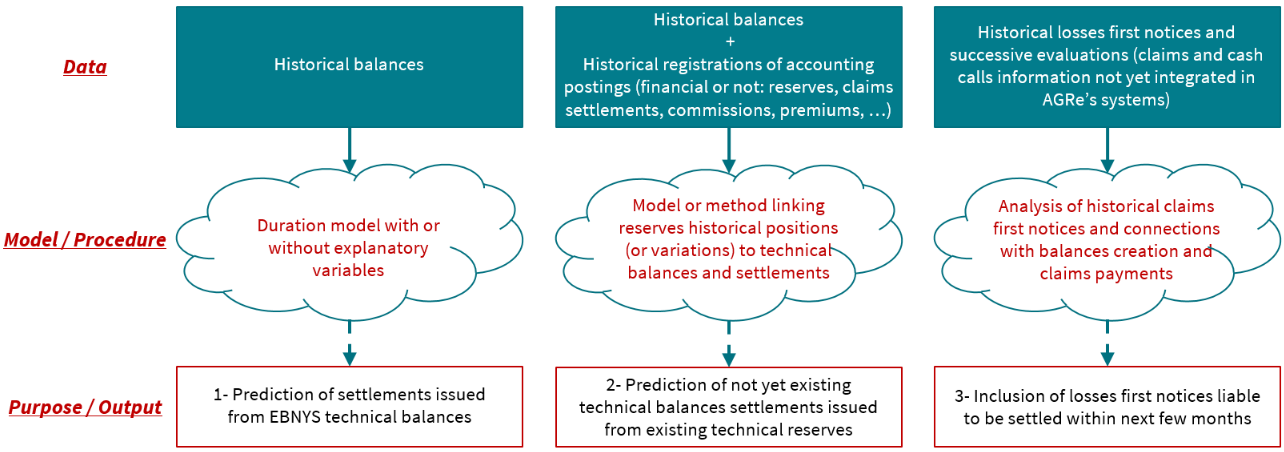


Figure 10: Répartition des objectifs de prévision des flux de trésorerie à court terme - structure des modèles et sources de données

Prévision des flux de trésorerie techniques annuels à long terme

Un changement de perspective sépare l'anticipation des besoins de trésorerie en devises dans un futur proche, et la projection à long terme du passif actuel estimé en règlements annuels de sinistres attendus dans des devises distinctes. Néanmoins, résoudre ces deux problématiques répond aux mêmes desiderata (à différentes échelles d'horizons temporels), à savoir être capable de faire se correspondre les actifs et les passifs pour chaque devise et chaque période de temps considérée, et par conséquent, pouvoir évaluer correctement les quantités de cash disponible à investir et les horizons d'investissement correspondants.

Une projection des flux de trésorerie annuels futurs en devises provenant des provisions techniques existantes a déjà été réalisée pour mesurer le risque de marché lié au passif d'AGRe, mais présentait certains inconvénients. Cette ancienne procédure quantifiait un stock de réserves à projeter par segment de provisionnement et par devise selon une méthode d'allocation des IBNR discutable. En outre, elle ne tenait compte que des développements des sinistres payés au niveau agrégé du segment de provisionnement pour projeter ces réserves, sans distinguer d'éventuelles différences dans les cadences de développement des passifs libellés dans différentes devises. Enfin, l'ancienne méthode de projection traitait de la même manière les provisions dossier-à-dossier et les IBNR sans différencier leurs natures et leurs rythmes d'évolution en règlements de sinistres.

Ces constats constituent autant d'axes d'amélioration que nous avons pris en compte dans l'élaboration d'une nouvelle méthodologie. Nous énumérons ci-dessous les principales mesures adoptées afin d'améliorer la projection actuelle des provisions techniques et de leur liquidation sous forme de transactions financières :

1. Application de la théorie de la crédibilité aux facteurs de développement des sinistres payés et rapportés, afin de générer des cadences de développement crédibilisés individuellement pour chaque devise à l'intérieur d'un segment de provisionnement donné, c'est-à-dire pour chaque groupe de traités de réassurance classés dans un même segment de provisionnement et dont la devise est la même.

2. Allocation des provisions IBNR à des classes de devises au sein des segments de provisionnement d'AGRe par le biais d'une approche Bornhuetter-Ferguson se basant sur des facteurs de développement des pertes crédibilisés.

3. Élaboration de nouvelles méthodes de projection tenant compte des comportements a priori différents dans les cadences de liquidation des réserves dossier-à-dossier et des réserves IBNR, ainsi que des informations contenues dans les cadences de développement spécifiques à chaque devise à l'intérieur de chaque segment de provisionnement.

Résultats

Tout au long de ce mémoire, nous sommes nous avons pu observer et mettre en lumière des résultats intermédiaires intéressants, telles que :

- la mise en évidence de la différenciation des vitesses d'évolution des règlements de soldes existants, selon leurs caractéristiques (classe de montant dans l'exemple ci-dessous) ;

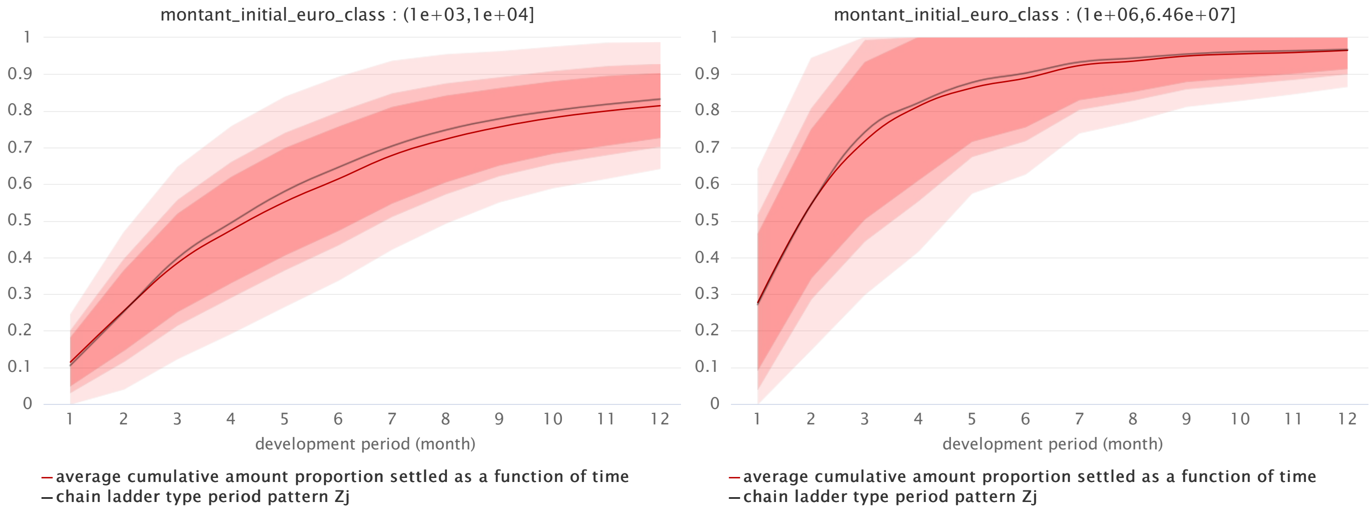


Figure 11: Proportion moyenne (sur les trimestres de rattachement) des montants réglés cumulés en fonction du temps (mois) avec intervalles de confiance (68%, 80% et 95% IC) - soldes de 1 000 à 10 000 euros en montant absolu (à gauche) et de plus de 1M d'euros en montant absolu (à droite)

- un phénomène de censure à droite dans les données des soldes historiques, ainsi que la mise en œuvre originale de modèles d'analyse de survie pour la prédiction des règlements issus de soldes existants et non réglés à date;
- un réarrangement ingénieux de la structure des données historiques financières et extra-financières, accompagné de propositions pertinentes et prometteuses d'implémentations de modèles pour tenter de prévoir les soldes non encore existants et les flux de trésorerie des règlements associés ;
- application réussie de la théorie de la crédibilité (modèle de Bühlmann—Straub) à l'analyse des développements de sinistres et de provisions;

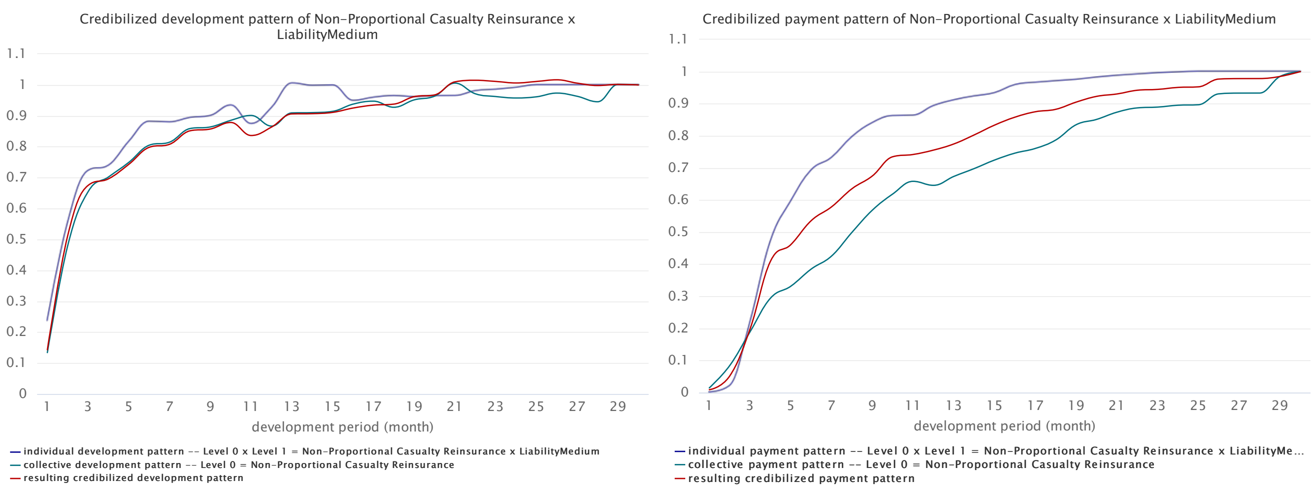


Figure 12: Cadences de développement des pertes (à gauche) et des sinistres payés (à droite) issus des facteurs de développement individuels, collectifs et crédibilisés – Niveau 0 (SOLVABILITÉ II BUSINESS LINE) X Niveau 1 (LR SCOPE) = Non-Proportional Casualty Reinsurance X PoolMotor

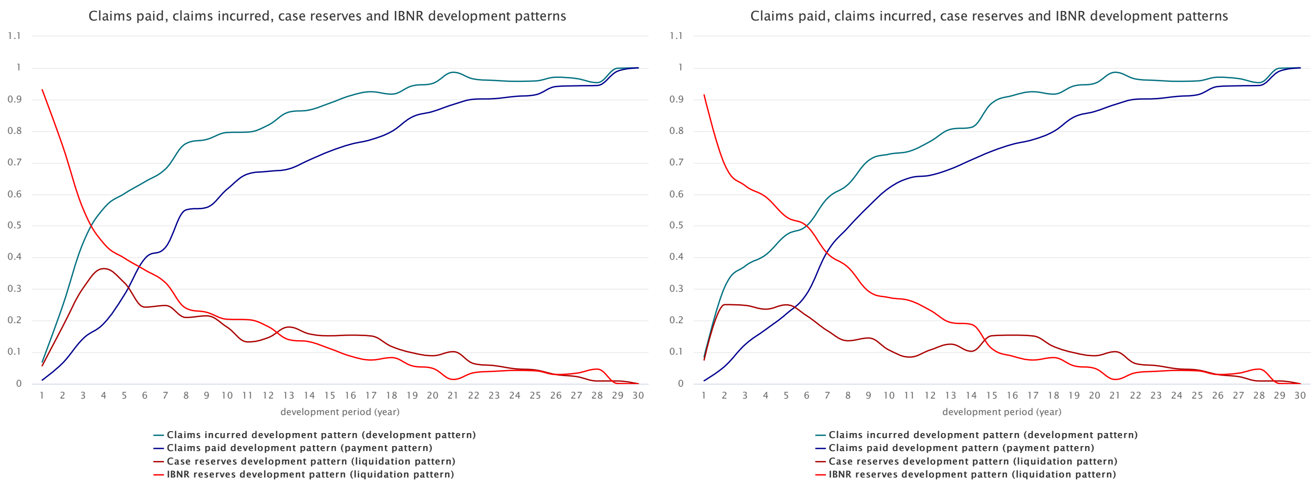


Figure 13: Cadences de développement crédibilisées de la sinistralité rapportée et payée, des provisions SAP et IBNR - Niveau 0 (SOLVABILITÉ II BUSINESS LINE) X Niveau 1 (LR SCOPE) = Non-Proportional Casualty Reinsurance X PoolLiability (à gauche) et Non-Proportional Casualty Reinsurance X PoolMotor (à droite)

- procédure innovante d'allocation des réserves IBNR à des groupes de traités de réassurance plus petits que les segments de provisionnement ;
- conception de nouvelles méthodologies de projection de liquidation des passifs à date en flux de trésorerie annuels ;

Chacune de ces étapes intermédiaires nous a permis d'obtenir des prévisions particulièrement satisfaisantes de flux de trésorerie mensuels en devises liées à des soldes existants et non réglés à montants positifs uniquement ou à montants négatifs uniquement, ainsi que des projections de flux de trésorerie annuels, liées à des cadences de liquidation de réserves, généralement plus proches des valeurs réelles par rapport à la méthode actuellement appliquée.

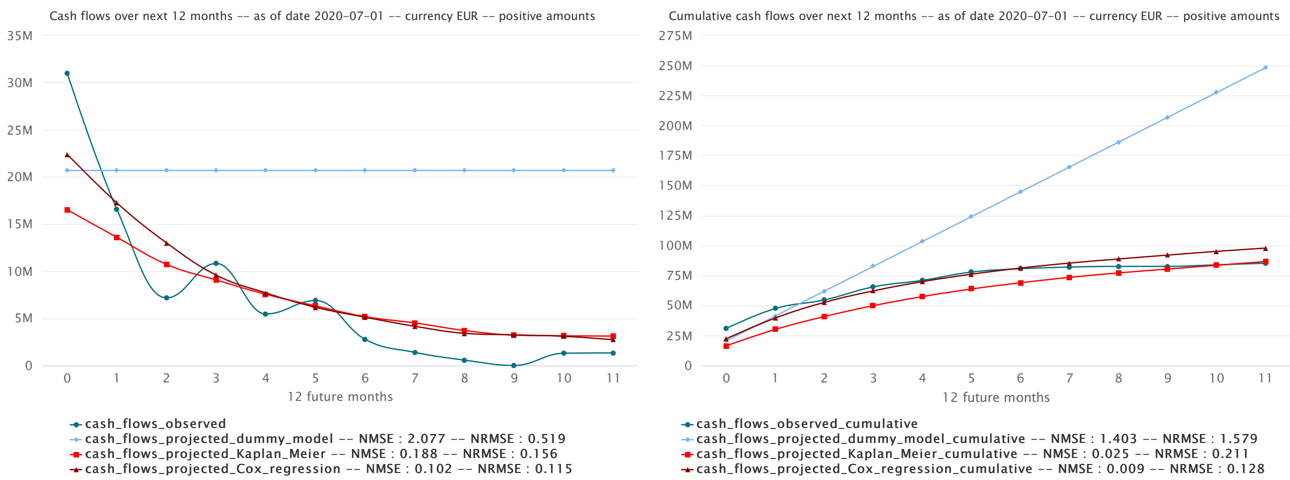


Figure 14: Flux de trésorerie individuels (à gauche) et cumulés (à droite) sur les 12 prochains mois – à la date du 01/07/2020 – devise EUR – montants positifs

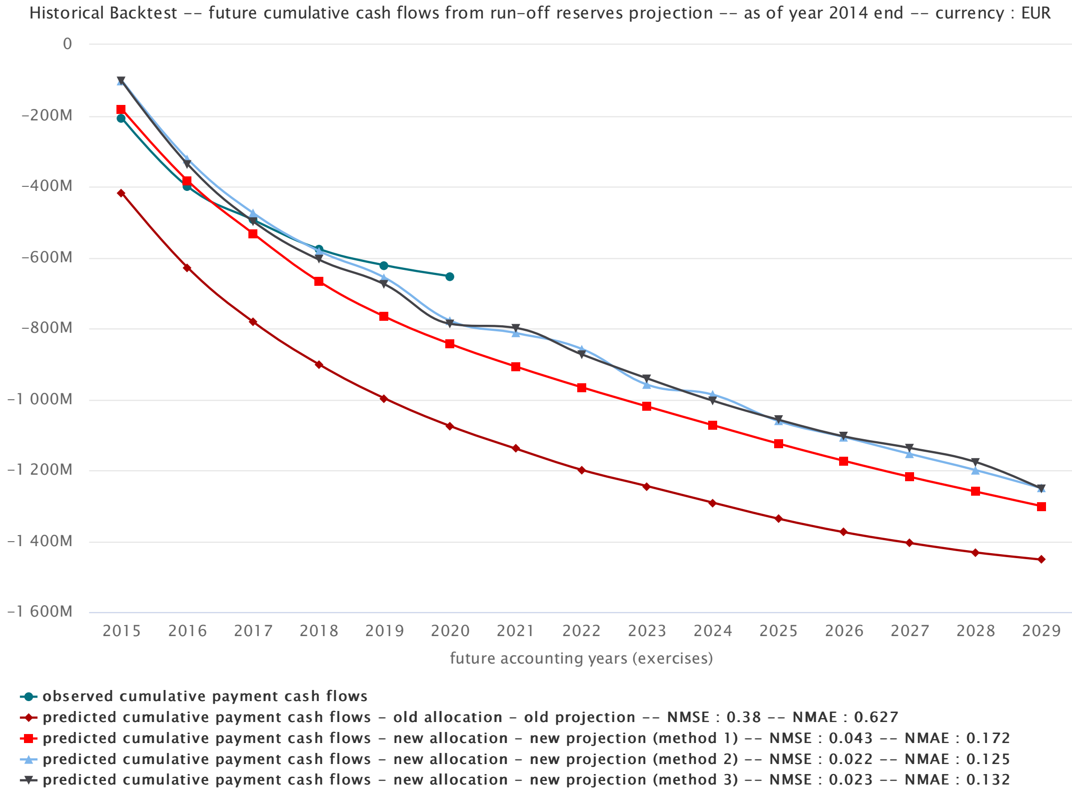


Figure 15: Projections des réserves en flux de trésorerie annuels cumulés par rapport aux flux cumulés des paiements des sinistres réellement observés - à partir de 2014 - devise EUR

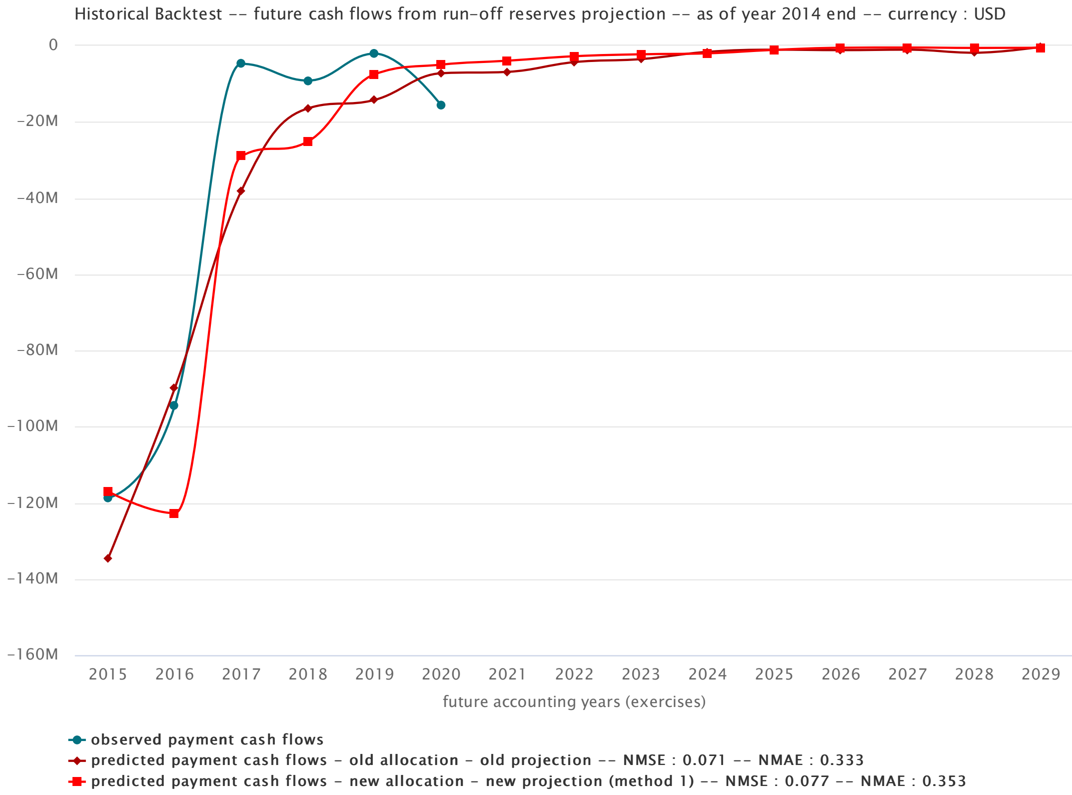


Figure 16: Projections des provisions techniques en flux de trésorerie annuels par rapport aux flux annuels de paiement des sinistres réellement observés - à partir de 2014 - devise USD

Contents

Introduction

If reinsurers were able to foresee the occurrence times of their future technical cash flows and quantify exactly in the proper currencies those future net amounts of cash transferred in and out of their business, they would be capable of both protecting themselves thoroughly from economic foreign exchange (FX) risk and identifying their available cash to invest optimally. From a reinsurer point of view, some cash flows such as those linked to general expenses, minimum and deposit premiums, taxes, dividends (related to closed exercise), financial investments products and so on are relatively easy to anticipate. This is generally not the case for claims settlements, commissions or adjustment premiums related technical cash flows, whether on the risks acceptance business side or on the retrocession one.

AXA Global Reinsurance (AGRe), as AXA Group's reinsurance platform, is responsible for AXA worldwide based entities' P&C and Life risk transfer to the reinsurance market, Group mutualization, as well as Group's capital optimization structuring, and constitutes one of the largest buyer of reinsurance in the market. As such, AGRe sees a lot of cash flows labelled in numerous currencies going through its business. It is then not a surprise that both AGRe's Risk Management and Treasury teams expressed the need to refine current simplistic prediction procedures, in order to get a better grasp of how future technical cash flows in currencies emerge, and to produce closer to reality forecasts.

Predicting technical inflows and outflows of a large set of currencies, specific to the reinsurance business, is not an easy task.
First of all, because of the random nature of events and claims impacting insurance along with reinsurance businesses, which in fine define the cash flows going through those entities.
Second, because of the inherent randomness of a reinsurer net liabilities (more so than an insurance company's) with respect to their estimated amounts in currencies as well as their extinction patterns. Even when adopting a run-off perspective, that is, considering a state of business as it is up to date, best estimate reserves are only estimations, and the pace of their transformation into claims payments are only deduced from past observations. Moreover, as it is the use at AGRe, those liabilities are generally determined through reserving actuarial methods on the basis of a specific reserving segmentation (homogeneous groups of risks / reinsurance treaties). The considered reserving segments do not necessarily integrate the currency dimension that could lead to better reserves amount and development pattern estimates by currency, and thus a better evaluation of what is to come in terms of run-off cash flows in currencies.
A third source of uncertainty, arising when trying to reach a prediction of future technical cash flows, is to be found in the time lapses of balances settlements. A claim is not instantly paid, even if both parties have agreed on its amount.

The first part on this thesis is dedicated to outlining the environment within which this research project was carried out. We will present briefly the non-life reinsurance area, AGRe's business structure and risk ceding mechanisms. We will further devote a chapter to reserves estimation general methodologies along with specificities linked to reinsurance. To close this first introductory part, we will contemplate the significance of foreign exchange risk in the reinsurance business and try to demonstrate the relevance of the issues addressed in our study.

The rest of this research project's rendering have been structured in the same manner as we tackled the subject of cash flows prediction. We indeed undertook to split this challenging issue into two main axes of development, corresponding to part II and part III of this thesis, namely a 12-month horizon monthly technical cash flows prediction and a run-off projection of current reserves into yearly cash flows. The endeavor towards the former objective principally seeks to meet the needs of AGRe' accounting and treasury departments, that is, foretell near future cash requirements in currencies. The work performed to achieve the latter objective mainly answers what AGRe's Risk Management sector calls for, that is, a trustworthy representation of AGRe's liabilities amounts and durations for all its business currencies. What's more, reaching those two goals would theoretically provide good investment opportunities, analysis we have not covered in the scope of this project.

Beyond the results obtained and presented in this report, which are limited considering the nature of AGRe's business and the data we could rely on and does on no account constitute a breakthrough in actuarial studies, the devised methods and models used in the resolution of this actuarial problem could be of great interest. We indeed offer innovative proposals and original applications of widely applied models. As an example, we apply survival analysis models to financial balances. We also resort to credibility theory for claims development patterns determination at lower levels of risks aggregation. The procedure we developed and the models we implemented can easily be generalized to other business data, i.e. other reinsurers. Results can even improve to the extent that AXA Global Re is not a big reinsurer and that major reinsurance market participants may have at disposal more historical data, more cash flows to work with, and more balanced proportions of cash transactions in different currencies. That is why our achievements can be seen as a contribution to the actuarial world.

Part I
Context of the study

1 Non-life reinsurance

1.1 Non-life insurance market

The insurance market includes many partakers, both private (individuals, businesses, companies, insurance, mutual societies, provident institutions, pension funds, etc.) and public (State, regulatory bodies, retirement plans, health plans, etc.). Some of them intervene specifically in life insurance, others in non-life insurance.

Since most of AGRe's portfolio composition is related to non-life reinsurance treaties, and the tiny fraction of life related treaties will be taken care of exactly in the same manners in our models, we will focus only on non-life insurance and reinsurance markets and practices.

We are interested here in the non-life insurance market, also called P&C - Property and Casualty insurance. Indeed, non-life insurance mainly covers risks related to property damages and general liabilities.

Let's have an overview of non-life insurance market participants:

Insured
Insured is a generic term that refers to any person or entity covered against the occurrence of an event through a contract called an "insurance policy", taken out with an insurer. Insurers make payments to insured after they experience a covered loss, damage, or an injury that qualifies for payment under the policy's terms. This could include damage to property owned by the named insured (the person who purchased the policy) or a third party.

Insurers
An insurer is a private company that assumes the risks that the insured do not want to bear in exchange for a financial flow called "insurance premium", usually paid before the occurrence of the insured event. The social entity representing the insurer can take several forms. In France, there are three: insurance companies, mutual societies and provident institutions.

Pool
Several insurers can share their risks within an insurance pool. Pooling does not reduce the random nature of the risks, but it is a process allowing mutualization (benefit brought by the fact of considering as a whole a large number of similar risks thanks to the independence between these risks, i.e. law of large numbers) and diversification (benefit provided by considering as a whole different and weakly correlated risks, i.e. having a low probability of occurring jointly over a certain period) gains on the sum of the risks. Usually, the members of a pool share its result in proportion to the premium share contributed by each of them.

Reinsurers
Reinsurance is insurance that an insurance company purchases from a reinsurance company to protect itself (at least in part) from the risk of a major claims event. With reinsurance, the company passes on ("cedes") some part of its own insurance liabilities to the reinsurer. The company that purchases the reinsurance policy is called a "ceding company" or "cedent".

The company issuing the reinsurance policy is referred to as the "reinsurer". In the classic case, reinsurance allows insurance companies to remain solvent after major claims events. In addition to its basic role in risk management, reinsurance is often used to reduce the ceding company's capital requirements. In other words, an insurer, as any risk-bearing entity, may wish to cover itself against some of the risks it has accepted to bear, even more so, it is generally hugely exposed. It then resorts to a reinsurer, who may accepts to bear the risks ceded, trough a contract called "reassurance treaty". The reinsurer may itself, in turn, wish to cover part of its risks with another reinsurer. In such a configuration, the ceding reinsurer in the retrocession process is the "retrocedent", and the end entity assuming the reinsurer's risks is known as the "retrocessionnaire". The reinsurer may be either a specialist reinsurance company, which only undertakes reinsurance business, or another generalist insurance company.

Financial markets

Both insurers and reinsurers have the possibility to transfer part of their underwriting risks to the capital markets through the creation and issuance of financial securities (Insurance-linked securities), process known as insurance securitization.

Insurance-linked securities (ILS) are broadly defined as financial instruments whose values are driven by insurance loss events. Those such instruments that are linked to property losses due to natural catastrophes represent a unique asset class, the return from which is uncorrelated and unrelated with that of the general financial market, feature particularly appreciated by investors seeking to diversify their portfolios.

The most prevalent securitized insurance contracts exchanged in capital markets include:

- Catastrophe bonds
- Embedded Value Securitization
- Extreme Mortality Securitization
- Life Settlements Securitization

The market for insurance linked securities has been very appealing for investors and insurers. One portion of insurance linked securities corresponds to the reinsurance of high severity, low probability events known as CAT bonds. These include cover for natural disasters and other uncontrollable events. These policies are grouped by their assessed risk, and then re-insured by other insurers. CAT bonds are grouped by their level of risk and sold in portfolios in security markets. This makes re-insuring these contracts more engaging because it opens a whole market for them to be sold and for risk to be spread among many investors. It is much more enticing to write expensive, risky policies and share the risk with thousands of others than it is for one firm to assume total liability.

State

The State plays several roles in the insurance market:

- Framework: it defines the legal framework applicable to the market;
- Control and regulation: it is responsible for protecting the interests of policyholders by monitoring insurance companies and financial actors. In France, the regulatory authority for insurance and financial markets is the "Autorité de Contrôle Prudentiel et de Résolution (ACPR)", attached to the "Banque de France";

- Reinsurance as a last resort: only the State is able to cover the most extreme events (uninsurable), relying on a total pooling of risks. In some cases, even if State support is not provided, national solidarity can apply. All States are not necessarily a last resort reinsurer, but France is, through the regime "Catastrophes naturelles" and the "Caisse Centrale de Réassurance (CCR)".

  Figures 17 and 18 presents non-life and life insurance and reinsurance market participants, as well as their interaction in terms of risk transfer.

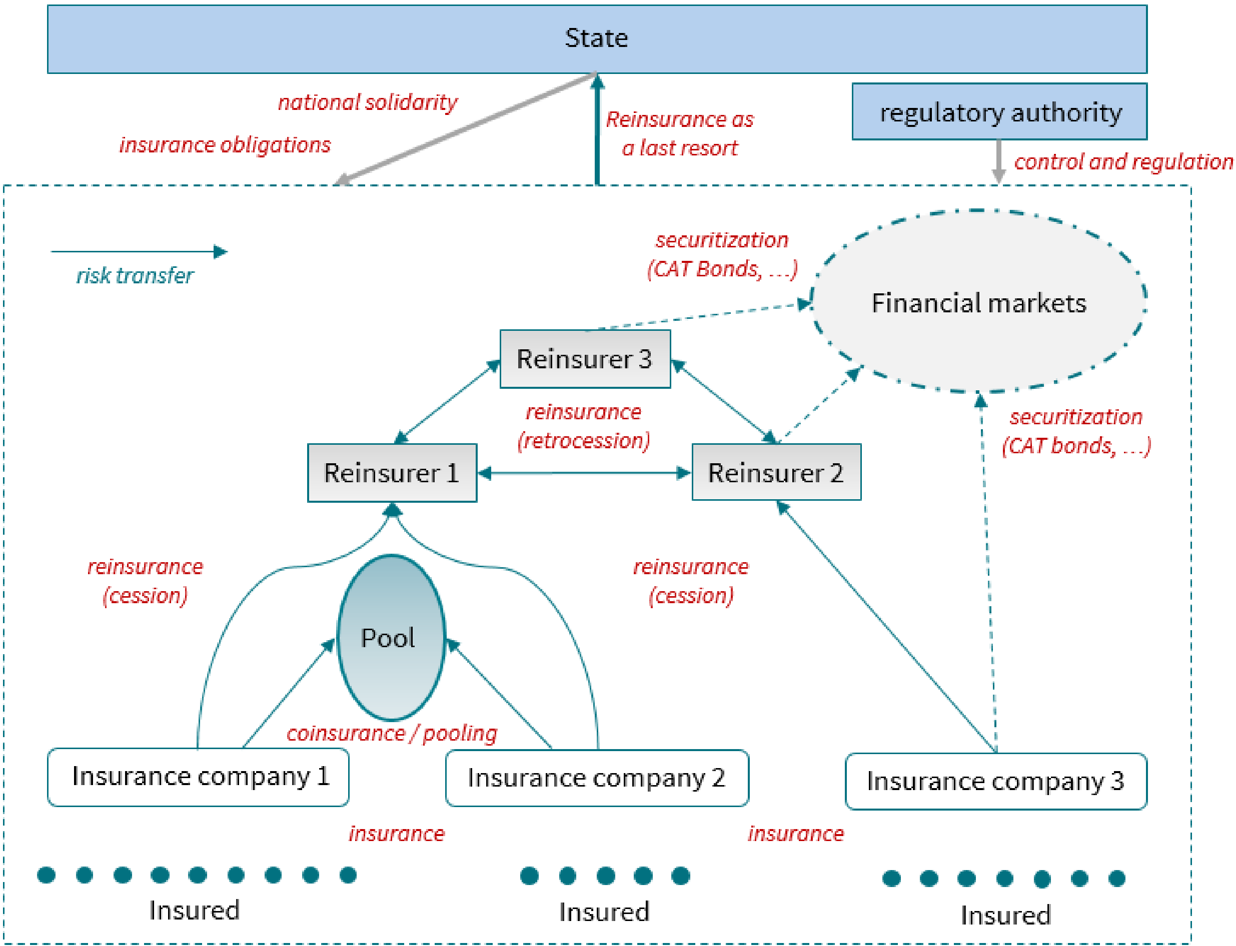


Figure 17: Non-life and life insurance and reinsurance market participants - risk transfer flows

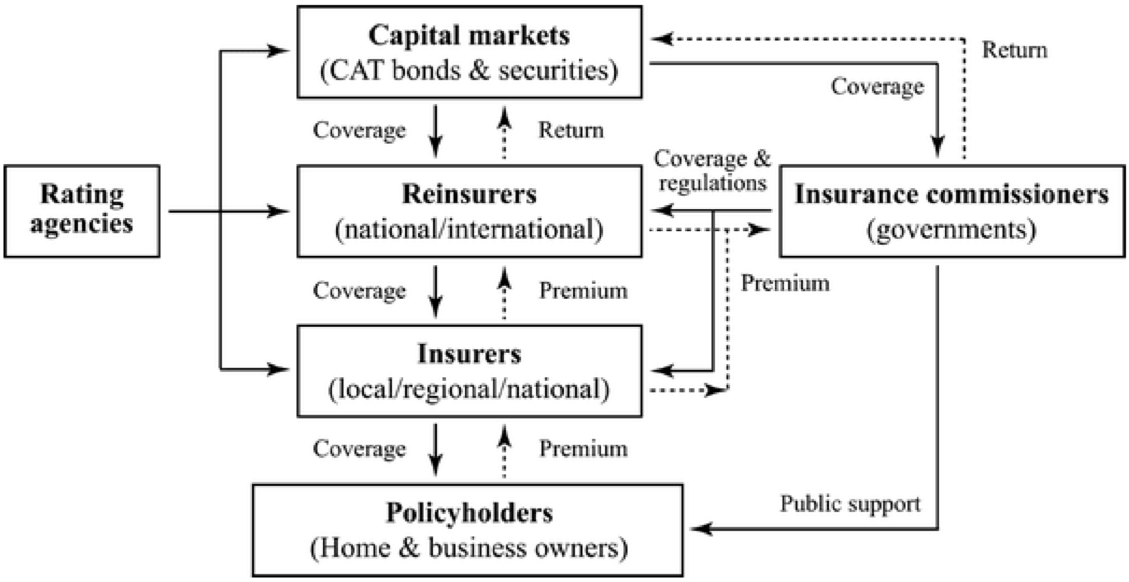


Figure 18: Overview of an insurance reinsurance system

1.2 Non-life reinsurance market

The reinsurance market is relatively small in terms of participant number: it has become very concentrated in recent years. It is also a very coded world, in order to facilitate transactions that often concern complex risks and can therefore involve many parameters. In this part, we present the main criteria characterizing a reinsurance contract binding an insurer to a reinsurer:

- the legal forms: Facultative, Obligatory, or Facultative Obligatory;
- the technical types: proportional or non-proportional.

1.2.1 Legal form

The two basic types of reinsurance are facultative reinsurance and treaty reinsurance.

Facultative reinsurance
Facultative reinsurance is issued on an individual analysis of the situation and facts of the underlying policy. It is negotiated separately for each insurance policy that is reinsured and may cover all or part of the underlying policy. By deciding coverage case by case, the reinsurer can determine if it wants the risk associated with that particular policy. Facultative reinsurance is normally purchased by ceding companies for individual risks not covered, or insufficiently covered, by their reinsurance treaties, for amounts in excess of the monetary limits of their reinsurance treaties and for unusual risks. Underwriting expenses, and in particular personnel costs, are higher for such business because each risk is individually underwritten and administered. However, as they can separately evaluate each risk reinsured, the reinsurer's underwriter can price the contract more accurately to reflect the risks involved. Ultimately, a facultative certificate is issued by the reinsurance company to the ceding company reinsuring that one policy.

Treaty reinsurance
Treaty reinsurance, on the other hand, is written to cover a particular class of policies issued by the cedent. Examples of classes covered by treaty reinsurance are all property insurance policies or all casualty insurance policies written by the cedent. Treaty reinsurance automatically passes the risk to the reinsurer for all policies that are covered by the treaty, not just

one particular policy. Treaty policies are more general than facultative policies because the reinsurance decision is based on general potential liability rather than on specific enumerated risks. It means that the ceding company and the reinsurer negotiate and execute a reinsurance contract under which the reinsurer covers the specified share of all the insurance policies issued by the ceding company which come within the scope of that contract. The reinsurance contract may obligate the reinsurer to accept reinsurance of all contracts within the scope (known as "obligatory" reinsurance), or it may allow the insurer to choose which risks it wants to cede, with the reinsurer obligated to accept such risks.

Facultative Obligatory reinsurance
This is a form of reinsurance in which a ceding company may choose to submit a given risk to the reinsurer, the latter being obligated to accept the risk up to its available retention limits, but may refuse the risk if it exceeds those limits. This form therefore exposes the reinsurer to a risk of anti-selection: the cedent may well keep its "good" risks and dispose of the bad ones. This is not, however, in the cedent medium-term interest, because as soon as the reinsurer does notice anti-selection, it will not wish to renew the treaty and the cedent will no longer be covered.

1.2.2 Technical types

In addition to the two types of reinsurance issued, there are two ways to allot coverage between the parties: either proportionally or non-proportionally.
Under proportional reinsurance, the reinsurer's share of the risk is defined for for each or for all policies, while under non-proportional reinsurance the reinsurer's liability is based on the aggregate claims incurred by the ceding company. Over the past years there has been a major shift from proportional to non-proportional reinsurance in the property and casualty fields.
Either type of coverage can be used in either facultative or treaty reinsurance contracts.

Proportional reinsurance

In the case of proportional reinsurance, the reinsured obtains coverage for only a portion or percentage of the loss or risk from the reinsurer. The proportion of coverage is typically based on the percentage of premiums paid to the reinsurer.

Under proportional reinsurance, one or more reinsurers take a stated percentage share of each policy that an insurer issues. The reinsurer will then receive that stated percentage of the premiums and will pay the stated percentage of claims. In addition, the reinsurer will allow a "ceding commission" to the insurer to cover the costs incurred by the ceding insurer (mainly acquisition and administration, as well as the expected profit that the cedent is giving up).

The arrangement may be "quota share", "surplus reinsurance" or a combination of the two. Under a quota share arrangement, a fixed percentage for all insurance policies is reinsured. Under a surplus share arrangement, the ceding company decides on a "retention limit". The cedent retains the full amount of each risk, up to the retention limit per policy or per risk, and the excess over this retention limit is supported by the reinsurer (up to a predetermined limit).

On one hand, the ceding company may seek a quota share arrangement for several reasons, one of which being an insufficiency of capital to prudently retain all of the business that it can sell. By reinsuring part of its coverage through a quota share, a cedent could grow its business without having to adjust its regulatory capital, and retain some of the profits on the additional

business via the reinsurance commission.
On the other hand, the ceding company may wish to take out a surplus reinsurance treaty to limit the losses it might incur from a small number of large claims as a result of random fluctuations in claims experience.

Non-proportional reinsurance

We speak of non-proportional reinsurance when the commitments of the insurer and the reinsurer are not proportional. This asymmetry is introduced by threshold effects. Proportional reinsurance has only two major adjustment factors: the sharing ratio and the commission. In all other aspects the reinsurer usually "follows the fortune" of the cedent. In contrast, non-proportional reinsurance allows for tailor-made solutions fitted to the targeted risk profile of cedents as close and as flexible as possible. This applies not only to the technical structure of the treaty (including reinstatements of coverage after loss events) but also to the set of conditions surrounding the treaty, including event definitions, inclusions/exclusions, and cash loss provisions.

Non-proportional treaties are carried out with respect to given amounts of loss. A deductible amount, defined either by event or in the aggregate, is set in the reinsurance contract. Any loss under the original policy exceeding that amount is borne by the reinsurer (within a capacity limit). The amount paid by the reinsurer has no relationship to the premiums received.

We generally define the following quantities, which technically characterize a non-proportional treaty:

- Retention or Priority (deductible): the claim or aggregate claim amount has to exceed this level in order to activate the treaty;
- Limit (Capacity): maximum obligation of the reinsurer per claim or event;
- Number of reinstatement: maximum number of restorations of the reinsurance limit to its full amount after a payment by the reinsurer of a loss as a result of an occurrence;
- Annual Aggregate Limit (AAL): maximum total amount of reinsured claim payable by the Reinsurer per occurrence year;
- Annual Aggregate Deductible (AAD): threshold amount below which the annual (or over the reinsurance period) loss of the ceding company covered by the treaty is not borne by the reinsurer.

Optimal reinsurance can mean multiple layers. It is indeed possible and may be better suited to meet the cedent needs to establish several layers, the limit of one corresponding to the priority of the second and so on. The interest of segmenting a non-proportional hedge into tranches is essentially commercial: this enables better negotiation for each layer of risk, possibly with different counterparties.

The main forms of non-proportional reinsurance are excess of loss and stop loss.
Excess of loss (XL) reinsurance transfers losses beyond a certain threshold (retention) from cedents to reinsurers. This can be done for single losses, events, or a combination thereof. Typically losses are covered up to a certain limit. Various limits can be staggered ("layers" of coverage). Cedents may cede all the losses in a layer or retain certain percentages of given layers. Reinsurers may demand that the cedent retain a portion of the layer as an incentive

not to overpay claims once losses reach the ceded layer. Unlimited covers are possible but uncommon (generally for Third-Party Liability).

Excess of loss reinsurance can have three forms:

- "Per Risk XL" (Working XL): designed to protect the cedent against large individual losses (see Figure 19).
- "Per Occurrence or Per Event XL" (Catastrophe or Cat XL): designed to protect the cedent against catastrophic events that involve more than one policy (see Figure 19).
- "Aggregate XL": Regardless of the physical reasons for losses, entire portfolios can be reinsured beyond a certain threshold. It generally serves as a protection against an annual cumulative loss.

Aggregate covers can also be linked to the cedent's gross premium income during a 12-month period, with limit and deductible expressed as percentages and amounts. Such covers are then known as "stop loss" contracts. For stop-loss treaties, retention and limit are indeed typically expressed in (annual) loss ratio terms for the covered portfolio. Stop-loss reinsurance, therefore, not only protects the insurer against large claims but also large number of small claims during the year. It affords what other reinsurance treaties do not: a frequency risk protection.

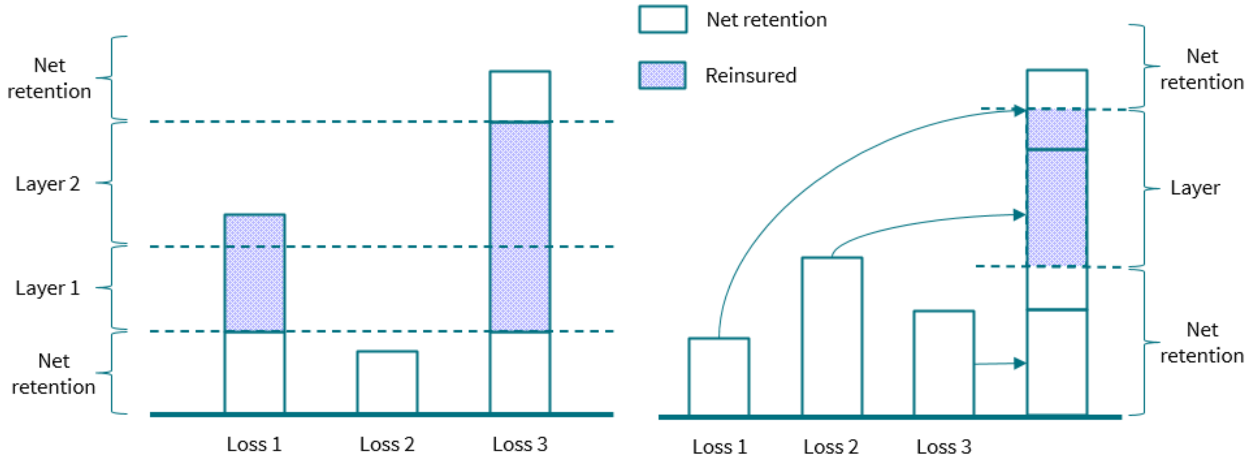


Figure 19: Per risk (left) and catastrophe (right) XL reinsurance treaty

1.2.3 Basis

Risks attaching basis

Under this basis, reinsurance is provided for claims arising from policies commencing during the period to which the reinsurance cover relates. The insurer knows there is coverage during the whole policy period even if claims are only discovered or made later on. That is, all claims from a cedent underlying policies incepting during the reinsurance contract period are covered even if they occur after the expiration date of that reinsurance contract. Any claims from a cedent underlying policies incepting outside the period of the reinsurance contract are not covered even if they occur during this period.

Losses occurring basis

It relates to a reinsurance treaty configuration under which all claims occurring during the period of the contract, irrespective of when the underlying policies incepted, are covered. Any loss occurring after the contract expiration date are not covered. As opposed to claims-made or risks attaching contracts, insurance coverage is provided for losses occurring in the defined period. This is the usual basis of cover for short tail business.

Claims-made basis
A policy which covers all claims reported to an insurer within the policy period irrespective of when they occurred.

To summarize in one sentence, for a given coverage time interval defined in a reinsurance contract established according to one of the three bases modalities, which are risk attaching, loss occurring, and claims-made, a protection is provided to the cedent if, respectively, the underlying policy takes effect, the loss occurs or the claim is made during that period of time.

2 Axa Global Reinsurance (AGRe)

2.1 General view

AXA Group was the first insurance group to create an internal reinsurance vehicle 20 years ago: AXA Global Re (formerly AXA Cessions), a wholly-owned subsidiary of the AXA Group, centralizes the Group's reinsurance business in property-casualty, life and health insurance. As part of its duties, AXA Global Re consolidates the reinsurance accounts of more than 60 entities, which then constitute the basis for calls for funds and financial flows issued between AXA Group and reinsurance companies. It is one of the largest buyers of reinsurance in the market.

AXA Global Re (AGRe) resulted from the merger in 2017 of AXA Global Life (AGL) and AXA Global P&C (AGPC), formerly in charge of life and non-life reinsurance respectively.

Although AXA Global Re has both life and non-life business, the life part is minor as manifested by its reserves stocks and earned premium proportions at half-year 2021, displayed in the table below. This is why we have focused until now on non-life reinsurance, and we will continue to do so throughout this thesis.

PERIMETER	**Reserves proportion**	**Earned premium proportion**
AXA GLOBAL RE P&C	96.3%	100.0%
AXA GLOBAL RE LIFE	3.3%	0.0%
Other	0.4%	0.0%

Table 1: AGRe's gross reserves and gross earned premium proportions for life and non-life business

AXA Global Re's missions are to contribute to the overall stability of the AXA Group's results around the world, to support the development of its entities and to allow them to benefit from the diversification of capital and retention, by optimizing and securing the reinsurance programs for all of the Group's activities and entities. It brings together both operational reinsurance functions (Customer Relationship, Underwriting, Claims, Actuarial Services), but also support functions (Risk Management, Finance, IT, HR, Legal).

As the Group's reinsurance placement hub for non-life and life reinsurance, AGRe is a key tool for optimizing and securing the operations of reinsurance ceded from the Group. This entity is responsible for:

- Analyzing the risks borne by the Group's entities;
- Accepting the risks that the entities of AXA Group do not wish to keep, then possibly transferring them to the reinsurance market;
- Designing and implementing the Group's ceded reinsurance strategy;
- Implementing risk sharing solutions (pools) and risk transfer solutions;
- Representing AXA Group with international reinsurance brokers and reinsurers in order to be the sole interlocutor of the market for negotiations and in order to centralize counterparty risk monitoring.

To summarize, AXA Global Re centralizes AXA Group skills regarding reinsurance, so that all of AXA Group's reinsurance goes through AGRe, with the exception of some facultative transfers. It's business model meets the AXA Group's needs of risk transfer, value retention and capital optimization. Those represent the three business pillars common to AGRe's P&C and Life perimeters:

- **Risk transfer / back-to-back reinsurance (Inward treaties = Outward treaties)**: transfer of risks underwritten by the Group's entities to the external reinsurance market, without retention by AXA Global Re except in case of commutation.
- **Group mutualization (Group's retentions, Pools and covers)**: this allows in particular the retention at the Group level of risks which exceeds the entities' risk appetite. Those risks are pooled and partly redistributed to the entities, partly retained by AGRe and partly covered by external reinsurance treaties (Group covers). AXA Global Re's retention is aligned with its capital level and risk appetite.
- **Group's capital optimization**: AXA Global Re can establish quota-shares contracts with certain Group entities (Germany and Mexico) in order to optimize their need for local capital. The Property pool organized by AGRe is also a capital optimization tool where entities cede CAT risks, not much diversified taken individually (storm Europe, Turkish earthquake, English flood, Mexican hurricane,...). Entities with share in the pool receive in return a diversified and global risk, whose needs in capital are reduced thanks to diversification. Here again, the retention of AXA Global Re is still aligned with its capital and its risk appetite. Loss Portfolio Transfer (LPT) with AXA XL is another tool implemented and employed by AGRe to optimize AXA Group's capital.

Segmentation and provisioning in AGRe's non-life reinsurance activity is organized around three poles: specific reinsurance, pools and Group coverage, which differ in the way the accepted risk is managed and reinsured. We will explain the specificities of these three poles. To begin with, Figure 20 presents a general diagram of AGRe commitments in terms of accepted and retroceded risks. We see that, within the pools and Group covers, accepted risks are aggregated, which is not the case for the specific retrocession.

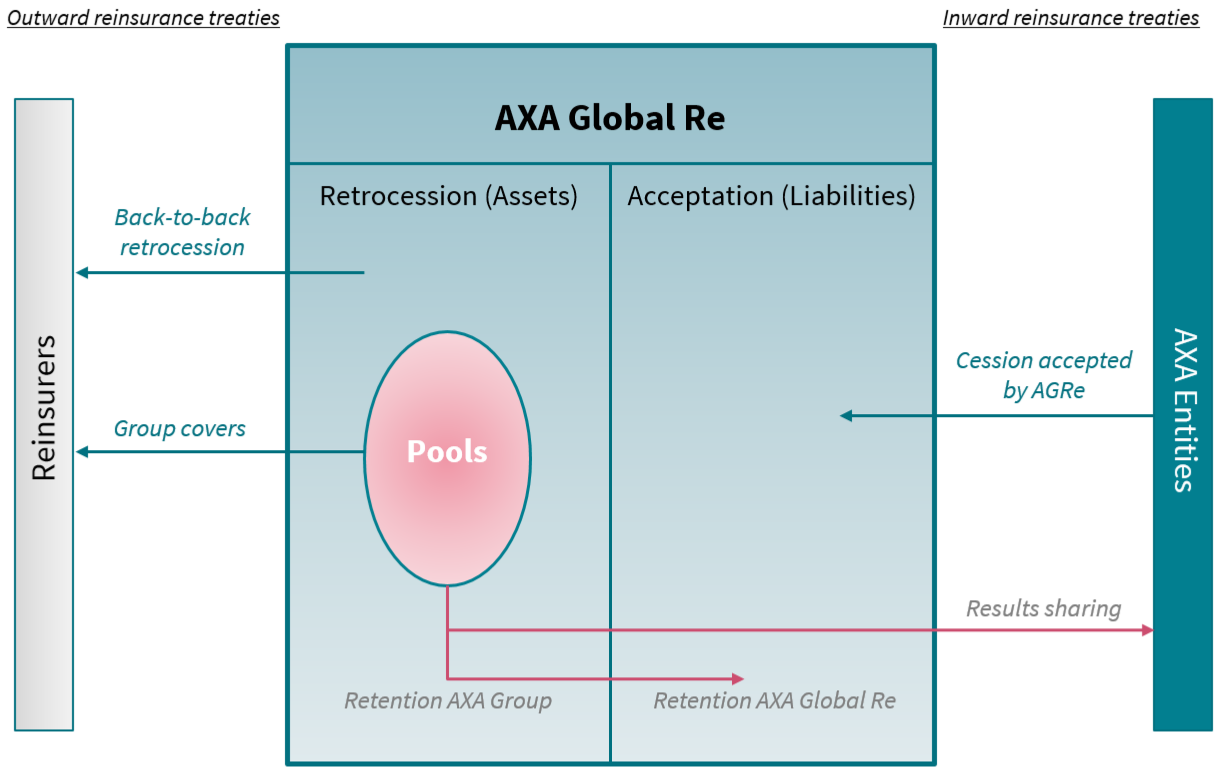


Figure 20: General diagram of AGRe commitments: fux of accepted and retroceded risks

Some of the risks are accepted then retroceded identically, i.e. the treaty is the same between the local entity and AGRe and between AGRe and the external reinsurer. This scope refers to "Non-pooled" business. AGRe's retention on this perimeter is due to past commutations with external reinsurers, or in some rare cases to "Cherry picking" (for some treaties very likely to be profitable).

Other risks are mutualized within so-called "Pools" before being retroceded. These pools are often protected by "Group covers", which are treaties underwritten with external reinsurers that applies to the risks accepted by the pools. The results of the pools may then be shared with ceding entities through quota-shares. The pools mutualize similar risks (e.g.: CAT, GTPL, MTPL,...) and each of them thus defines a reserving segment. These segments comprise the "Pooled" perimeter.

2.2 AGRe ceding mechanisms

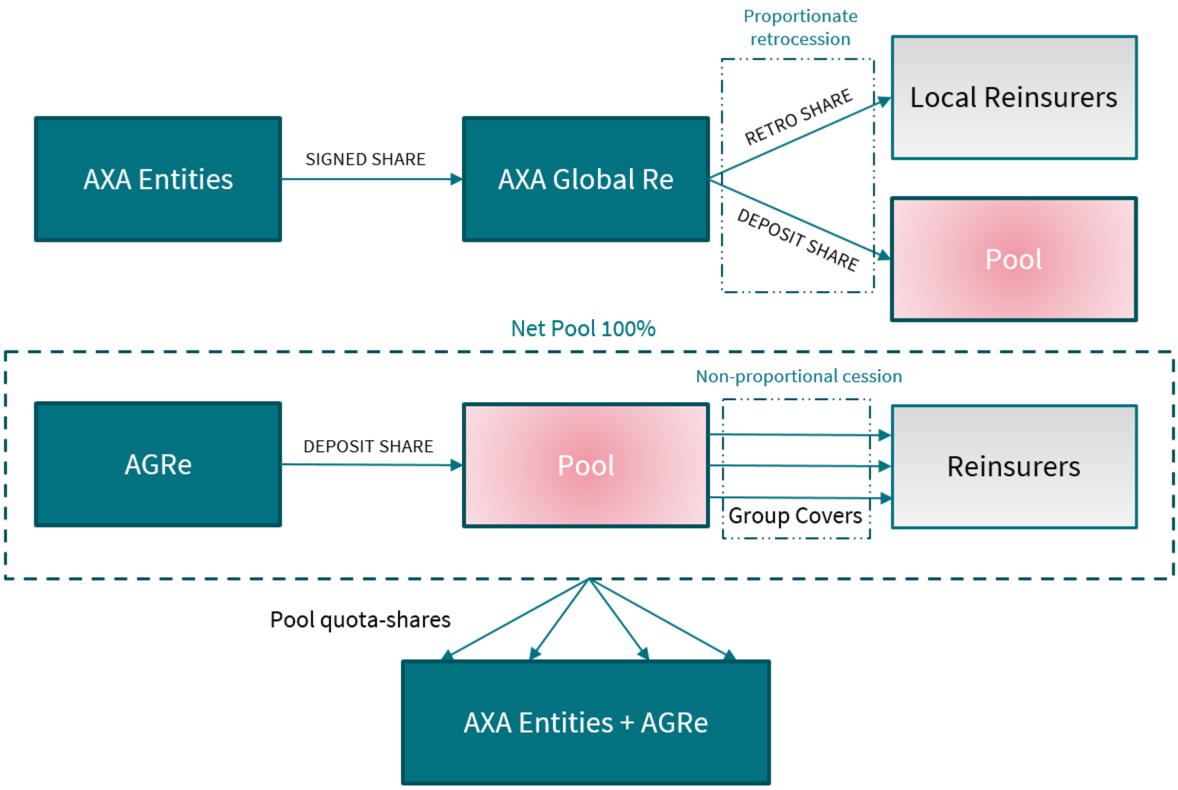


Figure 21: AGRe acceptance and retrocession schemes

AGRe's accepted business is either deposited in a pool or retroceded locally to reinsurers Most of the times, only a share of the accepted treaty is deposited in the pools (e.g. 80%), the other part is retroceded locally. Pools are protected by group covers. Net results of the pools are shared with participant of the pools (AXA entities), including AGRe. AGRe's accepted reserves cover all the businesses ceded by the entities. AGRe's net reserves correspond to the accepted reserves net of recoverable from local retrocession, group covers, and quota shares with other AXA entities participating in the pools.

- SIGNED SHARE takes values between 0 and 1 (proportion of the underlying risk ceded to and accepted by AGRe)

- RETRO SHARE (the retrocession share to external reinsurers) takes values between 0 and 1. If RETRO SHARE = 1 then the treaty is considered "not pooled", else it is labeled as "pooled" (pool or retention) except in the following cases:

- RETRO SHARE was initially equal to 1 but the contractual links with the corresponding reinsurers have been broken (commutation): the reinsurance treaty is still a "not pooled" one but its retrocession share is no more 100%;
- « Cherry picking »: concerning some treaties in the scope "Financialrisks" in 2015 and 2016 where retrocession was not performed at 100%, as well as facultatives underwritten in years 2016 to 2018 (scopes EngineeringFacDesk, LiabilityFacDesk and PropertyFacDesk);

At any time and for all underwritten reinsurance treaties we have the relationship DEPOSIT SHARE + RETRO SHARE = SIGNED SHARE.

As part of the specific retrocession, AGRe accepts certain risks ceded by AXA Group entities and transfer them identically on the reassurance market. Its role is therefore similar to that of a broker, except that the counterparty risk is borne by AXA Global Re and not by the cedent.

Figure 22 provides a second perspective one the general operation of a pool, the main points of which are as follows:

1. **Disposal**: AXA Group entities determine the risks they do not wish to keep (assessment of risk appetite) and cede them to AGRe.

2. **Local retrocession**: AGRe retocedes a portion (from 5% to 20%) of each treaty accepted locally. The retroceded treaty is identical to the accepted treaty, according to the same principle as in a specific retrocession, but the majority of the risk is retained by AXA Global Re and enters a pool. The main objective of local retrocession is to obtain a market price for each treaty.

3. **Group covers**: All the risks retained by AGRe within the framework of a pool are covered by Group protection, i.e. a reinsurance program covering all of these aggregated risks. This is where the main interest of the pools lies: there is mutualization and diversification of risks between participating entities, with the aim of reducing its random nature of the whole set of risks and therefore the cost of reinsurance.

4. **Results sharing (pools only)**: At each end of financial years, after Group covers' conditions application, a given pool perimeter generates a certain result, which is then shared between all AXA Group entities participating in that pool, via quota-share treaties.

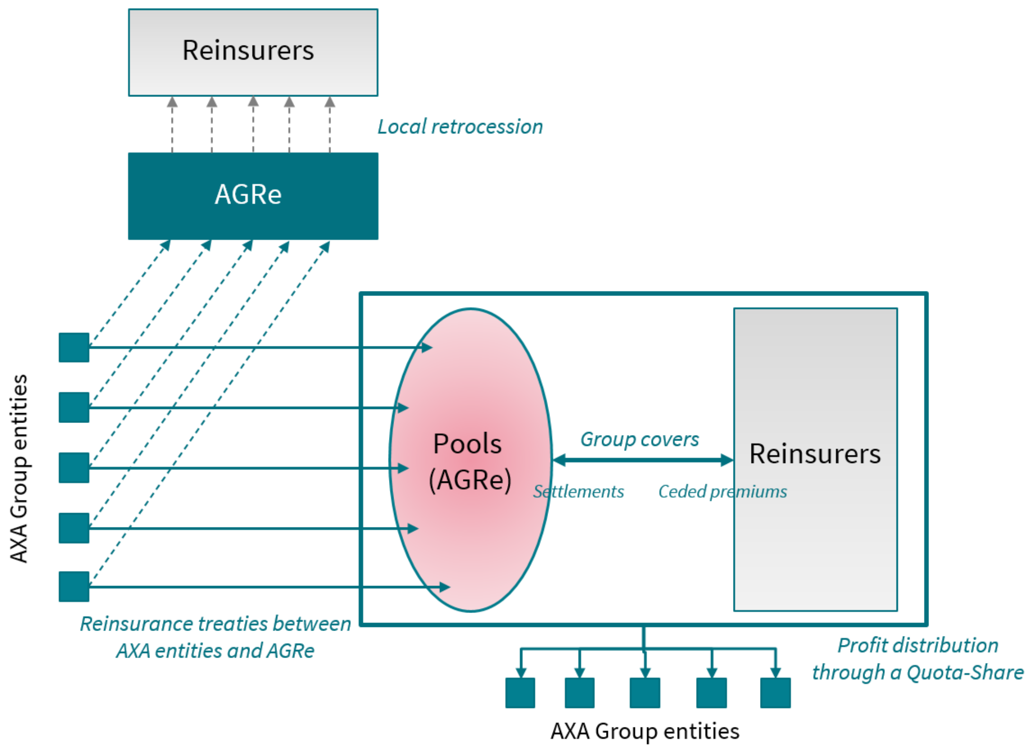


Figure 22: Pools and Group covers mechanism

The pool quota-shares may be equal to 0, in which case the mechanism is named a "retention" instead of "pool". This is the case with the Marine and the General / Motor Third-Party

Liability retentions, for which there is no risk sharing with the entities.

In a synthetic way, the mechanism of a pool therefore consists of an excess of loss treatment relating to the aggregate of a portfolio accepted by AXA Global Re, to which it may be added quota-shares enabling AGRe to transfer residual risks to AXA Group entities. As already mentioned, a pool is made up of risks of the same nature: damage to property, general third-party liability, motor vehicle liability, transport, etc. By extension, the set of pools and Group covers forms the so-called "pooled" perimeter. The treaties entering into the specific retrocession are, on the other hand, qualified as "non-pooled". The major difference between these two perimeters consists of the transformation and mutualization of risk from a net point of view, which only exists for pooled treaties. This pooling mechanism enables AXA Global Re to create most of the value it brings to AXA Group.

2.3 AGRe segmentation

The segmentation in reserving consists in defining segments or groups of risks on which actuarial estimates are established. According to AXA Group's standards, one has to group risks profiles that show similar development patterns and characteristics, remembering that:

- The more detailed the split is, the more volatile the results are;
- The less detailed the split is, the more diversification effect is added in the results.

Therefore, segmentation is performed with a trade-off between homogeneity and representativeness, i.e. by defining groups that are composed of similar risks and that have enough observed data. AGRe's current segmentation was defined in order to fulfil this equilibrium and to meet operational constraints involved in the different types of AGRe's ceding mechanisms. Those reserving segments are called Loss Ratio scopes (or LR scopes) at AGRe. Only non obsolete and material LR scopes will be introduced in this section.

2.3.1 Property and Casualty Business

Non-Pooled business

Non-pooled business segmentation was defined after a specific study in 2013 (see [1]), in order to find the best compromise between homogeneity and representativeness of the groups of risks. Material non-pooled business segments are listed below.

SOLVENCY II BUSINESS LINE	LOSS RATIO SCOPE
Credit and Suretyship Insurance and Proportional Reinsurance	Financialrisks
Fire and Other Damage to Property Insurance and Proportional Reinsurance	LPT_Property
Fire and Other Damage to Property Insurance and Proportional Reinsurance	QSGermanyEngineering
Fire and Other Damage to Property Insurance and Proportional Reinsurance	QSGermanyProperty
Fire and Other Damage to Property Insurance and Proportional Reinsurance	QSMexicanProperty
General Liability Insurance and Proportional Reinsurance	LPT_Liability
Marine Aviation and Transport Insurance and Proportional Reinsurance	LPT_MAT
Marine Aviation and Transport Insurance and Proportional Reinsurance	QSGermanyMarine
Motor Vehicle Liability Insuance and Proportional Reinsurance	QSGermanyMotor
Non-Proportional Casualty Reinsurance	Engineering
Non-Proportional Casualty Reinsurance	LiabilityFacDesk
Non-Proportional Casualty Reinsurance	LiabilityFrance
Non-Proportional Casualty Reinsurance	LiabilityLong
Non-Proportional Casualty Reinsurance	LiabilityMedium
Non-Proportional Casualty Reinsurance	LiabilityUK+ACS
Non-Proportional Casualty Reinsurance	Miscellaneous
Non-Proportional Casualty Reinsurance	NPDecennialConstruction
Non-Proportional Casualty Reinsurance	QSJapanMotor
Non-Proportional Casualty Reinsurance	QSKoreaMotor
Non-Proportional Casualty Reinsurance	SGR_Miscellaneous
Non-Proportional Casualty Reinsurance	SGR_RCA
Non-Proportional Casualty Reinsurance	SGR_RCG
Non-Proportional Marine Aviation and Transport Reinsurance	Transport
Non-Proportional Property Reinsurance	CatNatNonAuto
Non-Proportional Property Reinsurance	Property
Non-Proportional Property Reinsurance	PropertyFacDesk
Non-Proportional Property Reinsurance	SGR_Property

Table 2: Property and Casualty non-pooled business segmentation

In 2014, Saint-Georges Re (SGR) portfolio was transferred from AGL (AXA Global Life) to AGPC (AXA Global P&C). An external valuation of the technical provisions was performed. Since then, technical provisions estimation has been internalized and is regularly reviewed. SGR portfolio corresponds to old run-off pooled and non-pooled lines of business whose underwriting years are anterior to 2006 (GTPL, MTPL, Personal Accident, Financialrisks, Property, and Miscellaneous). Current net Best Estimate claims reserves of this portfolio represents 3% of total AGRe's Best Estimate of net claims reserves.

3 new reserving lines of business were created in 2020 following the Loss Portfolio Transfer with XLICSE.

Pooled business

Property and Casualty pooled business segmentation is presented in the following table:

SOLVENCY II BUSINESS LINE	LOSS RATIO SCOPE
Non-Proportional Casualty Reinsurance	PoolCyber
Non-Proportional Casualty Reinsurance	PoolEngineering
Non-Proportional Casualty Reinsurance	PoolLiability
Non-Proportional Casualty Reinsurance	PoolMotor
Non-Proportional Casualty Reinsurance	PoolSGR_PA_WC
Non-Proportional Casualty Reinsurance	PoolSGR_RCA
Non-Proportional Casualty Reinsurance	PoolSGR_RCG
Non-Proportional Marine Aviation and Transport Reinsurance	PoolMarine
Non-Proportional Property Reinsurance	PoolPropertyLineslip
Non-Proportional Property Reinsurance	PoolPropertyPA
Non-Proportional Property Reinsurance	PoolPropertyParEvt
Non-Proportional Property Reinsurance	PoolPropertyParRisk
Non-Proportional Property Reinsurance	PoolSGR_Property
Non-Proportional Property Reinsurance	PoolTIP

Table 3: Property and Casualty pooled business segmentation

2.3.2 Life Business

Life Business is segmented into non-pooled and pooled business. Due to the non-materiality of health treaties, the entire Life/Health business is reported in Solvency II line of business "Life reinsurance".

Non-Pooled business

The Life Non-Pooled lines of business were designed in 2020 using short-term vs long-term criteria, treaty type (XS, SL, XP, or QP) and with respect to their relative materiality and underlying life products for long-term business. The following table shows the detailed non pooled life segmentation.

SOLVENCY II BUSINESS LINE	LOSS RATIO SCOPE
Non-Proportional Life Reinsurance	NP - LT - HK HV II
Non-Proportional Life Reinsurance	NP - LT - HK HV II Revamp
Non-Proportional Life Reinsurance	NP - QP - LT
Non-Proportional Life Reinsurance	NP - SL - ST
Non-Proportional Life Reinsurance	NP - XP - LT
Non-Proportional Life Reinsurance	NP - XP - ST
Non-Proportional Life Reinsurance	NP - XS - ST
Non-Proportional Life Reinsurance	NP - XS - ST CAT

Table 4: Life non-pooled business segmentation

Pooled business

AXA Global Re accepts risks from the AXA ceding entities and retrocedes around 20% to external reinsurers. The other part is pooled in the so-called "Life Pool", whose results are shared between participating entities. AGRe used to retain a share on this pool (45% in 2015-2017, 45.66% in 2018). Since underwriting year 2019, Life business is no more retroceded to AXA entities, and AGRe keeps 100% of the net result.

AGRe Life pool and retention are composed of the following elements:

- Risk attaching treaties (long-term)
- Loss occurring treaties (short-term)
- Life CAT treaties
- Group cover treaties (CAT)

On the Group cover treaties, there is no calculation of IBNR recoverables as the treaties are very unlikely to be triggered.

Life reserving lines of business were redesigned in 2019 with the aim of balancing the number of lines between P&C and Life with respect to their relative materiality and to better capture the underlying life products for long-term business. Below is displayed the segmentation for risk attaching (LT) and loss occurring treaties (ST).

SOLVENCY II BUSINESS LINE	LOSS RATIO SCOPE
Non-Proportional Life Reinsurance	CAT
Non-Proportional Life Reinsurance	LT - Germany Risk Attaching
Non-Proportional Life Reinsurance	LT - HK - Cancer Therapy I
Non-Proportional Life Reinsurance	LT - HK - Cancer Therapy II
Non-Proportional Life Reinsurance	LT - HK - Health Elite
Non-Proportional Life Reinsurance	LT - HK - Health Vital
Non-Proportional Life Reinsurance	LT - HK - Other
Non-Proportional Life Reinsurance	LT - MasterLife
Non-Proportional Life Reinsurance	LT - Other
Non-Proportional Life Reinsurance	ST - Other
Non-Proportional Life Reinsurance	ST - XP - EUR - BE Disability
Non-Proportional Life Reinsurance	ST - XP - Other
Non-Proportional Life Reinsurance	ST - XS

Table 5: Property and Casualty pooled business segmentation

In order to have a bargain between enough data for triangle projections and homogeneous risks segments, AGRe's reserving team have tested several segmentation configurations. Some specific treaties have been excluded and handled separately.

3 Non-life reinsurance reserving: Reserves estimates and methodologies

As we will soon see, future cash flows forecast is closely linked to reserving issues and estimates. This is why, the need to carry out such an actuary project was initially formalized by AGRe's reserving team. The surveys and mathematical models implementations detailed throughout this thesis were also accomplished with the support of and under the reserving team's responsibility and supervision. We will then dedicate this entire chapter to bring up non-life reinsurance reserving topics and cover thoroughly reserving methodologies, in the light of AGRe's own practices.

As the AXA Group's reinsurance company, AXA Global Re accepts risks from AXA local entities through reinsurance treaties. Therefore, AGRe has liabilities that must be properly assessed with respect to Group's standards and regulatory requirements. Namely, AGRe reports the technical provisions related to these liabilities in French Gaap, IFRS and Solvency II statements. In this context, AGRe's reserving team estimates essential components of the technical provisions through a so-called "reserving process".

3.1 Technical reserves and outstanding claims reserves

This section sets out the regulatory and actuarial principles of technical provisioning in non-life insurance as well as reinsurance. We will intentionally overlook the concept of mathematical reserves since those are negligible within AGRe's business perimeter.

The inverted production cycle of the insurance market and the claim dynamics motivate the need for reserving and the design of predictive modeling tools to estimate reserves. In insurance, the premium income precedes the costs. An insurer will charge a client a premium, before actually knowing how costly the insurance policy or contract will become. In typical manufacturing industry this is not the case and the manufacturer knows, before selling a product, what the cost of producing this product was. So, in principle, an insurer offer a guarantee for which it does not know the exact cost at the time of the sale. It will know that cost only a posteriori, when the warranty period has elapsed and all the claims that it covers will have been declared and settled (loss development time span).

That is why an insurance company has a regulatory obligation to be able to meet its commitments, which translates to the obligation to form technical provisions at the end of each financial year, corresponding to the estimated amount of commitments remaining at its expense. At a specified evaluation moment the insurer must assess correctly outstanding liabilities with respect to contracts already sold.

In non-life insurance, there are two main technical reserves: unearned premiums reserves and claims payable reserves.

Reserves for unearned premiums or "unearned premium reserves" (UPR): unearned premiums, including the proportion of the amount of insurance premiums corresponding to risks that have not passed at the balance sheet date. If an insurance policy offers a guarantee covering several fiscal years (which is for example the case of an annual policy taken out mid year), the insurer is required, at the end of the first fiscal year, to form reserves to cover its commitments for the second fiscal year. The reserves formed must correspond to the premium of the policy pro rata temporis of the duration of guarantee remaining (unless the insurer can demonstrate that the risk covered is subject to a high seasonality). Part of the premium is thus considered to be attached to the first fiscal year and the other to the second.

Reserves for claims to be paid or "outstanding claims reserves": type of technical reserve (or accounting provision) in the financial statements of an insurer, that seek to quantify and to provide for the liabilities for insurance losses which have occurred but which have not yet been settled. When the period of guarantee covered by a policy has started and even once it is over, the insurer is in fact not released from its commitments: it must cover the claims declared but not yet settled, as well as those that have not yet been declared but for which the insured will be able to prove that they occurred during the covered period. Reserves formed to deal with these commitments are then:

- Reserve for declared, but not paid losses;

- Reserve for losses that have arisen but have not been reported.

Claims (or loss) reserve, i.e. the capital necessary to settle open (or soon to be declared) claims from past exposures, is an important element on the insurer's balance sheet.

From a risk point of view, the outstanding claims reserves are provisions for the past, since they cover losses that have already occurred (whether the claims are known or not), while unearned premium reserves are put aside for the future, since they cover losses that have not yet taken place.

If we consider the case of a reinsurer and especially AXA Global Re, outstanding claims reserves are clearly more important than unearned premium reserves, since most reinsurance treaties are taken out on January 1 and cover an annual period, therefore not giving rise to the constitution of UPR. As an example, regarding AXA Global Re risk acceptance in 2021: $\frac{\text{UPR reserves}}{\text{Outstanding claims reserves}} = 2.1\%$ (= 0.28 % for the net business point of view).

Let us have a detailed look at the elements constituting the outstanding claims reserves.

An insurance policy provides, in return for the payment of a premium, acceptance of the liability to make payments to the insured person or entity on the occurrence of one or more specified events over a specific time period. The occurrence of the specified events and the amount of the payment are both usually modeled as random variables. Claims may not settle immediately due to possible delays in their reporting (time that elapses between the occurrence of an insured event leading to a claim and the reporting of that claim to the insurance company) and their settlement process (time between the reporting and the settlement of a claim). In other words, there is, in general, a delay in a claim settlement, on account of reporting and settlement procedures, because it usually takes time to declare a claim and to evaluate the size of it. The time difference between claims occurrence and claims closing (final settlement) can take days (e.g. in property insurance) but it can also take years (typically in liability insurance). It is indeed very intuitive that a material or property damage claim settles quicker than a grave body injury claim involving a complex indemnification evaluation process and in some case life pensions/annuities. Closed claims may also reopen due to new developments (for example an injury requiring extra treatment).

Put together, the development of a claim typically calls for some handling time. The existence of the above-mentioned delays in the run-off of a claim requires the insurer to hold capital in order to settle these claims in the future. Claims reserving now means that the insurance company puts sufficient provisions from the premium payments aside, so that it is able to settle all the claims that are caused by past written insurance contracts. For a given covered risk stricken with a disaster, the insurer may found itself in one of the following three situations (corresponding to three types of claims in the books of an insurance company based on the claim's run-off status):

- The insurer has exact knowledge of the amount of the loss (and has possibly already paid for it in part or in totality). There is then no need to form provisions: there is at most only one cash flow lag. Such claims for which complete development has been observed are called closed claim;

- The insurer has estimated knowledge of the amount of the claim (and has possibly already partially paid for it), in which case he reserves the estimated amount of what remains to be paid. The provision thus formed is called case reserves. IBNER ("Incurred But

Not Enough Reported") provision must also be made in the event that the loss amount is underestimated and not sufficiently provisioned by case reserves. The sum of case and IBNER reserves thus make up the capital for claims "Reported But Not fully Settled" (RBNS) at the present moment or the moment of evaluation, that is, the moment when the reserves should be calculated and booked by the insurer;

- The insurer has no knowledge of the loss (whether the insured himself is not aware of it, or has not declared it yet). In this situation the insured event took place, but the insurance company is not yet aware of the associated claim, whose complete development (from reporting to settlement) will take place in the future. It cannot therefore provision the amount of the file as in the previous case, but must all the same gather provisions related to this "Incurred But not Yet Reported" (or IBNYR) claim.

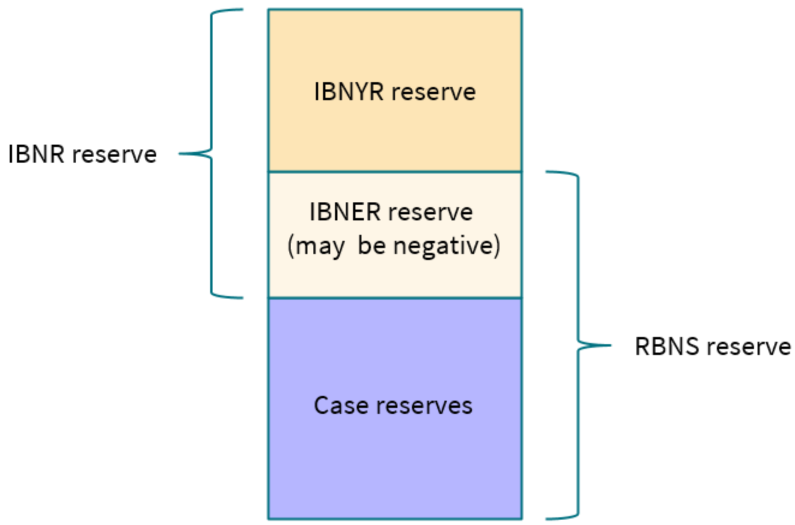


Figure 23: Different components of the claims reserve

Figure 23 provides a quick correspondence for the acronyms we have just introduced to describe unknown IBNYR claims, and RBNS known but not completely paid ones, thus dividing the claim reserve into the IBNYR reserve and the RBNS reserve. The latter being further divided into the aggregation of the individual case reserves and the IBNER reserve (actuary's adjustment to the case reserves on a collective basis). Insurance companies will reserve capital to fulfill their future liabilities with respect to both RBNS and IBNYR claims.

It is important to note that the case reserves are provisions estimated on a case-by-case or file-to-file basis for each incident that has occurred and has been declared. On the contrary, the IBNER and IBNYR are estimated by statistical methods for a set of contracts, at the scale of a branch (the branches definition or the segmentation used for provisioning therefore has an impact on their estimation). The future development of such claims is uncertain and predictive modeling techniques, calibrated on historical development data observed on similar claims, are used to calculate appropriate reserves. The object of the actuary is then to estimate IBNYR and IBNER reserves, the sum of which is referred to as the IBNR ("Incurred But Not Reported") reserves. IBNR thus corresponds to the part of the claims expense estimated statistically.

Ultimate loss

The ultimate loss is a concept of paramount importance in the determination of IBNR reserves. Let's see, how an insurer perceives the ultimate cost of claims to be borne by it, initially unknown, and how this perception breaks down and evolves over time.

As shown in Figure 24, the ultimate loss, for a given claim, is composed of three elements:

1. Settled or paid claims: how much has been paid by the company on an insurance claim so far. There is no actuarial judgment here, only accounting considerations.

2. Case reserves: the case estimate is the claims handler's expert estimate of the outstanding amount on a claim. Also known as "Case Outstanding", "Case OS", or simply "OS", this represents how much the claims adjuster believes still will be paid to settle this particular claim. Whether this amount is sufficient to cover the actual cost of the claim depends on the company's policies.

3. IBNR: represents an estimation of the claims amount that has not yet been recorded by accountants or claims handlers. It includes:
   - claims that occurred but haven't been reported to the company yet (IBNYR);
   - claims that have been reported but have too much or too little case reserves (IBNER);
   - claims closed but susceptible to be reopened;
   - salvage (sale of damaged goods for which the insured has been indemnified by the insurance company) and subrogation (collection by the insurance company of the amount of a paid claim from a negligent third party or his insurer) not included in the case reserves.

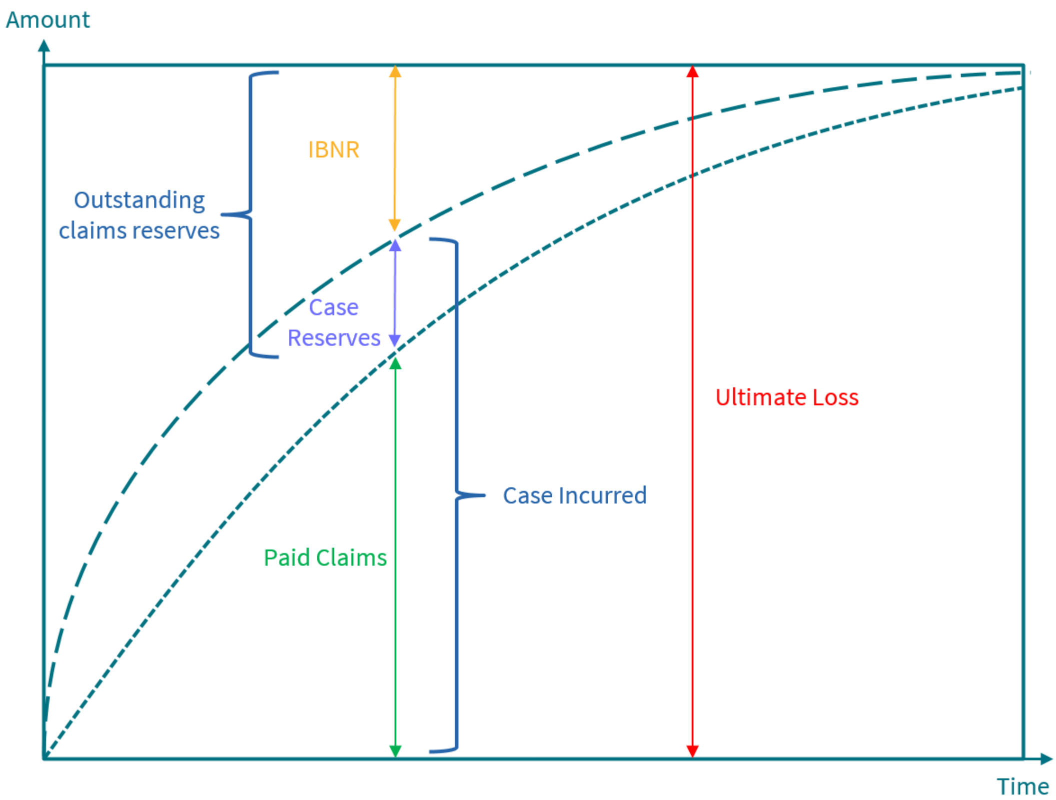


Figure 24: Loss components and reserves evolution over time

The sum of paid claims and case reserves form the claims incurred or case incurred:

$$\text{Claims Incurred} = \text{Case Incurred} = \text{Paid Claims} + \text{Case Reserves}$$

The sum of case and IBNR reserves form the outstanding claims reserves:

$$\text{Outstanding claims reserves} = \text{Case Reserves} + \text{IBNRs}$$

As illustrated in Figure 24, the real ultimate loss does not vary over time but is a priori not known, whereas the estimated ultimate loss may vary. Paid claims and case reserves increase over time, as claims are reported and settled. Case reserves do not have a monotonous behavior over time: they decrease when a payment is made and increase when new claims are declared. IBNR and oustanding claims reserves should always decrease over time, if the ultimate loss were fully known from the subscription. In practice, this is obviously not the case. We can also notice, in Figure 24 that the amount of cumulative payments does not reach the amount of case incurred at the end of the development period represented. This is typically what is observed in practice: payments have a longer development than the case incurred.

3.2 IBNR assessment methodologies

3.2.1 Loss reserving data formatting

Insurance companies typically register data on the development of an individual claim. We refer to data registered at this level as granular or micro-level data. Typically, we aggregate the information registered on the individual development of claims across all claims in a portfolio, or in a given branch, or even in a homogeneous sub-group of risks. This aggregation leads to data structured in a triangular format, of which an illustration is rendered in Figure 25 below. Such data are called aggregate or macro-level data because each cell in the triangle displays information obtained by aggregating the development of multiple claims.

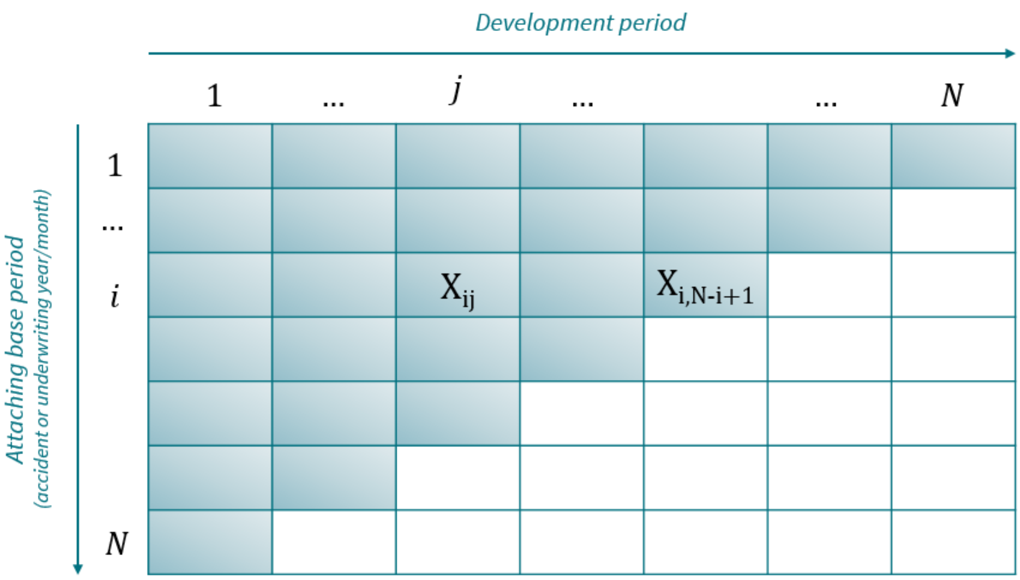


Figure 25: Generic development triangle

The triangular display used in loss reserving is called a run-off or development triangle and is represented in a simple form in the above figure.

- On the vertical axis the triangle lists the attaching base periods: it can either be the claims accident/occurrence period (year, quarter, month, etc.) or the risk/treaty underwriting period (year, quarter, month, etc.). The two conventions are referred respectively to "loss occurring" and "risk attaching". The loss payments booked for a specific claim are connected to the year during which the insured event occurred or to the year the covered risk was subscribed.
- The horizontal axis represents the development periods which indicate the payments delays since occurrence of the insured event, or since the risk/treaty subscription period, according to the chosen convention.

As far as we are concerned, the choice between the risk attaching or loss occurring conventions hardly has an impact for AXA Global Re, as for most reinsurance companies, since reinsurance treaties are almost all subscribed on January 1st for a period of one year.

Different pieces of information can be stored in run-off triangles. Depending on the kind of data stored, the triangle will be used to estimate different quantities. For example, in incremental format, the development triangle cells may display:

- the claims payments (paid claims increments);
- the changes in incurred claims amounts (incremental paid claims plus case estimates variation or reported loss increments);
- the numbers of claims that occurred in a specific year and were reported with a certain delay.

In a cumulative configuration, the cells may display:

- the cumulative paid amounts;
- the cumulative reported loss amounts (cumulative case incurred or incurred claims amounts);
- the total numbers of claims from an occurrence/underwriting year, reported up to a certain delay.

Throughout this thesis, we will analyse and manipulate risk attaching convention cumulative claims (paid or incurred) triangles: attaching years will be the underwriting years of the considered treaties, and development period will be counted in years. For the rest of this study we will therefore speak indifferently of attaching base period or underwriting years. Accordingly, our default triangle configuration will be:

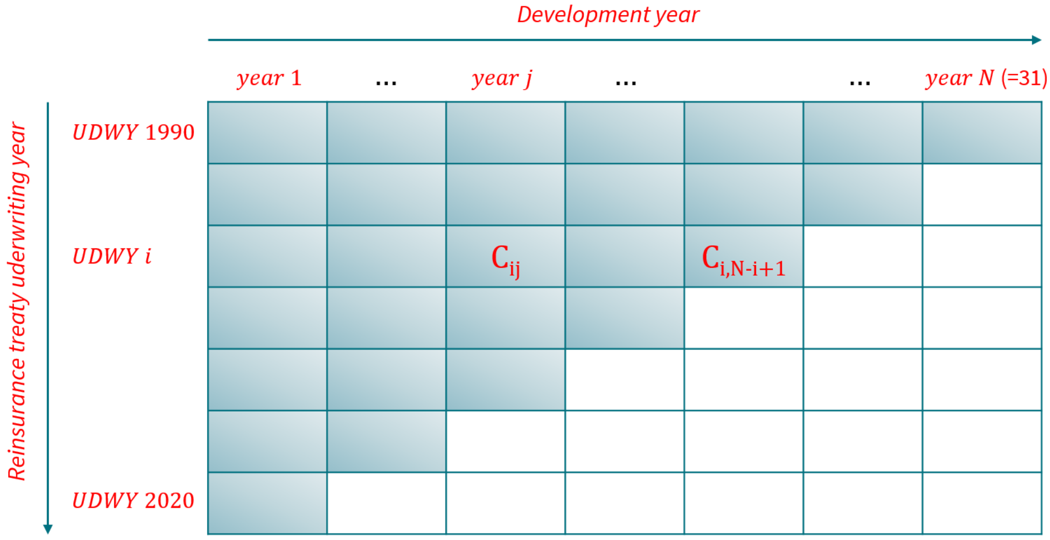


Figure 26: Specific development triangle

Let's have a look at the generic development triangle of Figure 25. We consider that we have a history of N years of subscription, for which we have observed 1 (for the last underwriting year) to N (for the first underwriting year) years of development. Since the first year of subscription, N calendar years have elapsed: the triangle does indeed have N diagonals. The underwriting years are denoted $i \in [\![1, N]\!]$ and development years $j \in [\![1, N]\!]$. Calendar years correspond to the diagonals and therefore satisfy an equation of the type: $i + j = k$ for fixed k in $[\![2, N+1]\!]$. Finally, the information available in the triangle, known in the passed, verify the condition: $i + j \leqslant N + 1$. The triangle contains numerical quantities. As we have seen, they can for example be settlements, incurred losses, claims numbers, premiums, etc. These quantities can be incremental or cumulative by attaching period. In the following, we will note in lowercase $x_{i,j}$ the incremental quantities, and in uppercase $X_{i,j}$ the cumulative quantities, so that we have,

$$\forall i \in [\![1, N]\!], \quad X_{i,1} = x_{i,1} \quad \text{and} \quad \forall j \in [\![2, N-i+1]\!], \quad X_{i,j} = X_{i,j-1} + x_{i,j}$$

For the sake of generality, we chose to denote by $x_{i,j}$ or $X_{i,j}$ the considered quantities. When we specifically analyse development triangles displaying cumulative case incurred quantities it is customary to write it down $C_{i,j}$. Finally, let us underscore that, in our case, the studied quantities are considered as a function of the underwriting year and of the development year.

When we will refer to the "ultimate" not being followed by any specification, it will imply any of the possible considered quantities in a given developement triangle (ultimate - loss, premium, paid claims, reported claims, number of claims,...). It will generally be the ultimate loss.

3.2.2 Estimation methods

In this subsection, we present the principle of several actuarial methods for estimating triangles end development points (and in particular IBNRs), each having advantages and drawbacks. The selection of a given method depends on the reserving segment and the underwriting year. This choice is mostly based on common sense rules and reserving professionals qualitative judgments. Further manual adjustments may be performed by actuaries on top of well defined methods, with the view to reach more consistent results.

The deterministic Chain-Ladder model

The most widely used and standard method to estimate outstanding loss reserves (and so IBNRs) is the so-called Chain-Ladder method. It is also the starting point for many other methods and is based on the concept of development factor. We define the individual development factors (or link-ratios) $f_{i,j}$, for $i+j \leqslant N$ by:

$$f_{i,j} = \frac{X_{i,j+1}}{X_{i,j}} \quad \text{which amounts to} \quad X_{i,j+1} = f_{i,j} \cdot X_{i,j}$$

In this form, the relevance of the underlying "chain" term in the "Chain-Ladder" method's name appears: the cumulative quantities are linked together thanks to development coefficients.
This method is based on a strong assumption: it assumes that, for $j \in [\![1, N-1]\!]$, the development factors $f_{i,j}$ are independent of the year of subscription i. We can therefore denote by f_j their common value. Making this assumption consists in assuming that developments do not depend on underwriting years, which is obviously debatable. With this hypothesis, we therefore have, for $j \in [\![1, N-1]\!]$:

$$\frac{X_{1,j+1}}{X_{1,j}} = \frac{X_{2,j+1}}{X_{2,j}} = \cdots = \frac{X_{N-j,j+1}}{X_{N-j,j}}$$

These ratios being assumed to be equal, their common value can be expressed as:

$$f_j = \frac{\sum_{i=1}^{N-j} X_{i,j+1}}{\sum_{i=1}^{N-j} X_{i,j}}$$

In practice, the ratios are not equal. We therefore retain the last formulation to define the Chain-Ladder development (or age-to-age) factors f_j. In the counterpart probabilistic model published in 1993, Thomas Mack shows that it is the only unbiased estimator of the development factor.

Link-ratios $f_{i,j}$ can also be interpreted as individual observations of the development factor between development periods j and $j+1$. An estimate of f_j is then obtained by a weighted mean of the observed individual link-ratios $f_{i,j}$, weighted by the $X_{i,j}$:

$$f_j = \frac{\sum_{i=1}^{N-j} X_{i,j+1}}{\sum_{i=1}^{N-j} X_{i,j}} = \frac{1}{\sum_{i=1}^{N-j} X_{i,j}} \sum_{i=1}^{N-j} X_{i,j} \cdot f_{i,j}$$

Other weights $w_{i,j}$ can be chosen, allowing some observed link-ratio to be excluded if considered as "outlier" and thus not representative of the development factor ($w_{i_0,j} = 0$ for some observation i_0).

Once the development coefficients are known, it is easy to complete the triangle, using the chain principle: $X_{i,j+1} = f_j \cdot X_{i,j}$, with $i+j > N+1$ this time. We thus obtain the ultimate amount for the underwriting year i (with the convention that an empty product is equal to 1), by projecting the last observed charge:

$$X_{i,N}^{CL} = X_{i,N+1-i} \cdot \prod_{j=N+1-i}^{N-1} f_j.$$

If the $X_{i,j}$'s are cumulative case incurred $C_{i,j}$, this estimated ultimate amount is the ultimate loss estimate $C_{i,N}^{CL}$, and we can deduce the amount of IBNR by subtracting the current and known cumulative case incurred.

$$IBNR_i^{CL} = C_{i,N}^{CL} - C_{i,N+1-i} = (\prod_{j=N+1-i}^{N-1} f_j - 1) \cdot C_{i,N+1-i}$$

This leads to an estimate where the IBNR reserves strongly depends on the current reported claim amount.

Finally, we define the notion of development pattern (period pattern or development rate): for $j \in [\![1, N-1]\!]$:

$$z_j = \frac{\prod_{k=1}^{j-1} f_k}{\prod_{k=1}^{N-1} f_k} = \frac{1}{\prod_{k=j}^{N-1} f_k}$$

z_j is interpreted as the ultimate quantity (ultimate loss for example) proportion known after j development periods. In the case of reported claim amount quantities, it is an estimate of the % of ultimate claim amount reported after j years of development. It is therefore natural to set $z_N = 1$, since the ultimate amount is supposed to be known after N periods (years in our case). In the Chain-Ladder model, those development patterns do not depend on the attaching period (reinsurance treaty underwriting year in our case).

Another manner to express IBNR estimates in the Chain-Ladder model is then:

$$IBNR_i^{CL} = (\frac{1}{z_{N+1-i}} - 1) \cdot C_{i,N+1-i}$$

Mack's distribution-free Chain-Ladder model

The previously introduced traditional Chain-Ladder method provides a point estimator $X_{i,N}^{CL}$ for the forecast of $X_{i,N}$, using past information. Being a purely deterministic model, it does not allow to determine how reliable that point estimator is and to what extent it is liable to be in the vicinity of the true but not currently known end point $X_{i,N}$. To answer such questions an underlying stochastic model that extends and reproduces the Chain-Ladder deterministic estimates is needed.

That is exactly what the distribution-free Chain-Ladder model, as an underlying stochastic model disclosed in 1993, does. This method allows to estimate the standard errors of the Chain-Ladder predictions. It is not parametric (does not assume a distribution) and is conditional (uses conditional expectations). Mack's model is based on three fundamental assumptions:

- Independence with respect to attaching base period, i.e. the random variables $(X_{i_1,j})_{1 \leqslant j \leqslant N}$ and $(X_{i_2,j})_{1 \leqslant j \leqslant N}$ are independent for $i_1 \neq i_2$. This strong assumption is not verified if the triangle is affected by significant changes (modified claims management, variations in inflation, etc.).

- Existence of development factors: $\forall j \in [\![1, N-1]\!], \quad \exists f_j \quad$ such that $\quad \forall i \in [\![1, N]\!]$:

  $$\mathbb{E}\left[X_{i,j+1} \mid X_{i,1}, \ldots, X_{i,j}\right] = f_j X_{i,j}, \quad \text{that is to say} \quad \mathbb{E}\left[f_{i,j} \mid X_{i,1}, \ldots, X_{i,j}\right] = f_j$$

  f_j is therefore defined as the best estimate of $f_{i,j}$, knowing $X_{i,1}, \ldots, X_{i,j}$.

- Existence of volatilities: $\forall j \in [\![1, N-1]\!], \quad \exists \sigma_j \quad \text{such that} \quad \forall i \in [\![1, N]\!]$:

$$\mathrm{Var}\left[X_{i,j+1} \mid X_{i,1}, \ldots, X_{i,j}\right] = \sigma_j^2 X_{i,j}, \quad \text{that is to say} \quad \mathrm{Var}\left[f_{i,j} \mid X_{i,1}, \ldots, X_{i,j}\right] = \frac{\sigma_j^2}{X_{i,j}}$$

Under these model assumptions, the Chain-Ladder development factors are unbiased estimators of Mack's factors, and are uncorrelated with each other (see [2]). The ultimate estimator is the same as in the Chain-Ladder method. The essential contribution of Mack's model, in addition to the stochastic generalization of the Chain-Ladder method, is to offer the possibility of evaluating the error made by the estimator of the ultimate, by year of subscription or globally. This gives an indicator of the precision of the estimator and makes it possible to study the distribution of reserves by making an assumption on its shape, knowing its first two moments.

Munich Chain-Ladder method

The Munich Chain-Ladder method rests on the same principle as the model of Mack, but instead of development triangles of one only quantity, it studies quantities ratios development triangles, for example paid claims to case incurred ratio. By projecting both paid claims and case incurred we assess the same thing: the ultimate loss. However, practice shows that the ultimate losses obtained through the Chain-Ladder method on cumulative payments triangles and on case incurred ones, do not converge. Thus, the Munich Chain-Ladder method makes it possible to reduce this gap, by projecting quantities of two development triangles at the same time rather than separately. The idea behind this methodology is to use more information in order to make the estimate more robust.

Loss ratio

Some insurance markets have existed for a long enough time and are sufficiently stable for us to have a good idea of their profitability. The most widely used profitability indicator in insurance is the loss ratio. Thus, the insurer or reinsurer, knowing the premiums it has received for a given year of subscription, is able to estimate the ultimate loss for the underwriting year in question, by simply multiplying the premiums by the expected loss ratio.

The loss ratio method is the simplest method to compute IBNR reserves. It uses a so-called "a priori loss ratio" to estimate the "a priori" ultimate loss of some underwriting year i, from which the IBNR reserve is estimated:

$$IBNR_i = C_{i,N}^{a \text{ priori}} - C_{i,N-i+1}$$

where $C_{i,N}^{\text{a priori}} = LR_i \cdot P_i$ is the a priori ultimate claim, with LR_i being the a priori loss ratio and P_i being the amount of premiums, both related to underwriting year i.

A priori loss ratio can be derived from a pricing model or from historically observed loss ratios. This method, like all others, can be carried out at different aggregation levels (portfolio, branches, sub-branches, etc.).

Bornhuetter–Ferguson method

Like other loss reserving techniques, the Bornhuetter–Ferguson method aims to estimate incurred but not reported insurance claim amounts. Generally considered a blend of the Chain-Ladder and expected claims loss reserving methods, the Bornhuetter–Ferguson method uses both reported (or paid) losses as well as an a priori expected loss ratio to arrive at an ultimate loss estimate.

The idea behing this combination of loss ratio and Chain-Ladder methods is simple and full of common sense: the Bornhuetter-Ferguson ultimate is a weighted average between the a priori ultimate and the Chain-Ladder ultimate. The weighting changes according to the degree of confidence attributed to the Chain-Ladder estimator: older is the attaching base period, the more information is available for the Chain-Ladder method, the more its estimator is reliable and the more its weight in the Bornhuetter-Ferguson method is important. More formally, considering loss quantities for a given subscription year $i \in [\![1, N-1]\!]$, the Bornhuetter-Ferguson estimator of the ultimate $C_{i,N}^{BF}$ is expressed as follows:

$$C_{i,N}^{BF} = z_{N+1-i} C_{i,N}^{CL} + (1 - z_{N+1-i})\, C_{i,N}^{\text{a priori}}$$

As announced, the Bornhuetter-Ferguson ultimate is indeed a weighted average of the Chain-Ladder ultimate $C_{i,N}^{CL}$ and the a priori ultimate loss $C_{i,N}^{\text{a priori}} = LR_i \cdot P_i$. Let us simply note that the weighting coefficients are themselves estimated by the Chain-Ladder method and are therefore, as desired, not constant. The following writing is equivalent to the previous one but allows another interpretation of the Bornhuetter-Ferguson method:

$$C_{i,N}^{BF} = C_{i,N+1-i} + (1 - z_{N+1-i})\, LR_i \cdot P_i$$

$C_{i,N+1-i}$ representing the observed historical part, to which we add an estimated part. As the Chain-Ladder patterns indicate that we are currently observing a share z_{N+1-i} of the ultimate, there logically remains a share $1 - z_{N+1-i}$ involved in future development. We therefore take this part of the a priori ultimate loss as the estimated part.

The Bornhuetter-Ferguson method can be generalized by any combined methods, considered as a mixed between loss ratio and link ratio methods. The ultimate claim amount is once again estimated by a weighted mean between the ultimate estimates derived from the two methods:

$$C_{i,N}^{\text{combined}} = (1 - \alpha_i) \cdot C_{i,N}^{\text{loss ratio}} + \alpha_i \cdot C_{i,N}^{\text{link ratio}}$$

The weight α_i often depends on the maturity of the underwriting year i. When the underwriting year i is recent, there is few observations on claims and one should give more weights to the a priori ultimate obtained with the loss ratio method, thus α_i is set close to 0. On the contrary, when the underwriting year i is old, more credit is given in the ultimate obtained with the link-ratio method, and α_i is taken close to 1.

De Vylder least squares IBNR method

De Vylder's least squares method comes within the scope of factor models, and is based on the hypothesis that the incremental quantities $x_{i,j}$ are written in the form of a product of three parameters, each corresponding to a particular direction in the development triangle, namely, for $(i, j) \in [\![1, N]\!]^2$:

$$x_{i,j} = x_i y_j \lambda_{i+j}$$

De Vylder's model further assumes that the parameter λ_{i+j}, which follows the evolution of calendar time and can be interpreted as an inflation factor, is constant. It can then be "merged" with the other parameters and the model is written: $x_{i,j} = x_i y_j$. By adding the condition $\sum_{j=1}^{N} y_j = 1$ (without it the parameters are not identifiable), we can then interpret x_i as the ultimate corresponding to the subscription year i and y_j as the (incremental) rate of the development year j. The parameters are then estimated by the least squares method (possibly weighted by $w_{i,j}$), minimizing the quantity:

$$\sum_{i+j\leqslant N+1} w_{i,j}\left(x_{i,j}-x_i y_j\right)^2$$

This can be written in the form of a system by differentiating according to the x_i's and y_j's:

$$x_i = \sum_{j=1}^{N+1-i} \frac{w_{i,j}x_{i,j}y_j}{w_{i,j}y_j^2} \quad \text{and} \quad y_j = \sum_{i=1}^{N+1-j} \frac{w_{i,j}x_{i,j}x_i}{w_{i,j}x_i^2}, \quad \forall (i,j) \in [\![1,N]\!]^2$$

whose solutions can be obtained numerically, step by step, through successive iterations. De Vylder's least squares method is quite simple and mathematically sound. It is however less easy to interpret, and less flexible than the Chain-Ladder method, especially if one does not introduce weights $w_{i,j}$.

GLM models

GLM models (for generalized linear models) are very widespread in the field of statistics in general and are based on a parametric stochastic structure. They have been used for provisioning since the 1990s, and suppose that the incremental triangle values are the realization of random variables following the same law from the exponential family, of which only the parameters change according to the underwriting and development years. A commonly used model in practice is that of the over-dispersed Poisson law. Introducing a parameter ϕ, the model is then written:

$$\mathbb{E}\left[x_{i,j}\right] = x_i y_j \text{ and } \operatorname{Var}\left[x_{i,j}\right] = \phi x_i y_j$$

It is ultimately a question of estimating the parameters x_i, y_j and ϕ, which is allowed by the generalized linear model, through regression. The GLM model approach therefore provides a stochastic structure for the whole incremental development triangle. In return, it imposes to make an assumption on the law of increments. In particular, if the chosen law has a positive support (as is the case of the over-dispersed Poisson distribution), then the model is only viable for triangles whose increments are always positive, which represents a strong constraint.

Bootstrap (applied on top of another method like CL to get a estimated distribution of reserves)

We will not delve here into the mathematical intricacies of the bootstrap method for reserving. Let's just have an overview. First of all, it is a non-parametric method of simulation based on the principle of resampling. It can be applied in any context, as long as the studied data can be considered as realizations of independent and identically distributed random variables.

Bootstrapping is a statistical method for estimating the sampling distribution of an estimator, by sampling with replacement from the original sample (observed data). It has been called the plug-in principle, as it is the method of estimation of functionals of a population distribution by evaluating the same functionals at the empirical distribution based on a sample.

Within the framework of development triangles, the studied quantities, whether they are cumulative or incremental, do not come from the same law: the influence of the subscription and development years is not negligible. Thus, we have to consider the Pearson residuals (normalized residuals) of the observations contained in the triangle, reported to a theoretical expected values stemming from a GLM model. The method therefore loses its non-parametric advantage here. Resampling the residuals triangle gives rise to a new triangle and an ultimate value per underwriting year. By repeating this operation the desired number of times, we obtain sufficient realizations of the ultimate to study its bootstrap distribution, as well as its mean, variance, quantiles, etc. Although it is poorly adapted to the development triangles' studied quantities, the bootstrap method leads to a distribution of the ultimate with very few observations. It nevertheless imposes the use of a GLM type model.

AGRe's reserving team resorts to the Group standards' statistical method "Mack-Bootstrap of claims triangles", presented in Appendix A, to generate moderate and stressed reserves scenarios and compute corresponding levels of reserves (respectively 60%-quantile and the 90%-quantile of the reserves distribution). The Risk Margin Moderate (RMM) and the Risk Margin Stress (RMS) are obtained by difference with the Best Estimate level.

IBNR/OS ratio

This method consists in applying a ratio to the amount of outstanding reserves (case reserves) in order to estimate the IBNR reserves:

$$\text{IBNR Reserves} = \text{ratio}_{\text{IBNR/OS}} \cdot \text{Case Reserves}$$

At AGRe, such method were used in particular for the reserving of asbestos risks, where new and unanticipated claims developments occurred and methods based on triangles were not reliable. One can resort to benchmark in such cases.

Survival ratio

This method consists in applying a ratio to an amount of annual claims paid considered to be an average of the rhythm of payment for the years to come.

$$\text{IBNR Reserves} = \text{Survival ratio} \cdot \text{Average yearly payment}$$

The amount of reserves obtained is then sufficient to pay the claims for x more years, where x is equal to the survival ratio, if the rhythm of yearly payments does not change significantly. Such method can be used if the recent paid claims are believed to be the only reliable information. It uses an indicator, the survival ratio, than can be compared with the market.

Tail factors

If we consider that the available history is not sufficient to correctly describe the development pattern of the studied branch, we can add development coefficients f_j for $j \geqslant N$, called tail factors. In practice, the case where f_{N-1} is not close to 1 means that there is not enough observed development periods in the triangle, and a tail factor should be derived to take into account further late developments (thus considering a longer development than the historical and observed one).

Tail factors cannot be estimated by the Chain-Ladder method, due to lack of data, and must therefore come from another source of information. When they are used, they are generally estimated by expert opinion or resulting from a market research. As mentioned, a tail factor should be used when the duration of the claim development is likely to be higher than the length of the observed period. In addition to expert opinion and market research, other methods exist to derive a tail factor, further denoted F_N. A few of them are presented in Appendix B.

3.2.3 Expected recoverables on IBNR (ceded IBNR reserves)

Let us be more specific, and estimate IBNR in a non-life reinsurance and retrocession context. Until now, we indeed described several general IBNR assessment methods, applicable both on an arbitrary insurance portfolio and on a portfolio of reinsured risks without consideration of retrocession, as is the case of AGRe's portfolio of local entities' ceded risks (accepted by AGRe) gross of retrocession. We now present how reserves are estimated and handled in the special case of a partly retroceded portfolio, in view of AGRe's specificities and practices.

As we saw, at AGRe the most part of accepted risks from ceding companies is retroceded, either through back-to-back retrocession treaties, or via the pools, where « Group covers » are applied and the results of the pools are ceded to the pools participants with respect to their shares.

At Full Year (FY) 2020, 90% of the accepted AGRe's reserves were retroceded.

Case of back-to-back retrocession

This type of cession applies to the non-pooled business where the local treaties are mirrored. The ceded treaty is linked to the accepted local treaty and has the same characteristics (QS/XL, priority, limit...). In general, several reinsurers sign a share of the treaty. The difference between the accepted signed share and the sum of reinsurers' shares is the net share for AGRE. Over time, the reinsurers can decide to commute their share, in such case AGRE's net share increases.

The ceded reserves are calculated by applying the sum of all reinsurers' shares to the gross reserves. The contractual databases are used to obtain the effective ceding rate for each treaty and section.

- In the acceptation (accepted risks) contractual database, there exist a field indicating the accepted signed share for each "mirror" local treaty between AGRe and AXA entities, that we will denote p_a (mostly equal to 1).
- In the retrocession (retroceded risks) contractual database, the signed share field designates either the part of each reinsurer, denoted p_r, or the commuted part with the cedent, denoted p_c.

On the gross side, the file to file reserve is reduced by the commuted part with the cedents. If X is the file to file reserve as informed in the acceptation database:

- $\frac{X}{p_a}$ denotes the file to file reserve at 100%;
- $p_c \cdot \frac{X}{p_a}$ corresponds to the commuted part;
- $X_{\text{accepted}} = X - p_c \cdot \frac{X}{p_a}$ corresponds to the file to file reserve after commutation (the amounts stated in AGRE's accepted liabilities).

The ceded file to file reserve is then calculated as follows:

$$X_{ceded} = \frac{X}{p_a} \cdot \sum_r p_r = \frac{\sum_r p_r}{p_a - p_c} \cdot X_{\text{accepted}}$$

Note that if $p_c = p_a, \quad X_{\text{accepted}} = 0$ and $X_{\text{ceded}} = 0$.

Concerning the gross IBNR reserves, calculated by line of business and underwriting year, they must be first allocated treaty by treaty. The allocation key is the earned premiums for treaties whose maturity is inferior or equal to 8 years, and the case reserves for older treaties. Once the IBNR reserves are allocated, ceded IBNR are calculated the same way as for the file to file reserves.

Case of Group covers

The pools are protected with "Group covers" corresponding to reinsurance treaties underwritten with external reinsurers. Thus, each pool can be seen as a specific entity, with a risk acceptance (accepted premiums) and a retrocession (premiums brought to reinsurers). The characteristics of Group covers treaties are not always in line with the accepted local treaties: the difference corresponds to AGRE's net retention in accordance with its own risk appetite.

1. Loss ratio method:

Pool models give a gross and net of reinsurance result of a given pool and for a given underwriting year. More precisely, pool models integrate the scenarios generators of the set of risks accepted by a given pool and takes into account the underwritten Group covers to simulate the results of that pool. The simulations average result gives the Best Estimate of the “a priori” ultimate loss, gross and net of Group covers. Thus, net IBNR reserves can be calculated the same way they are calculated on the gross side applying the loss ratio method with the net loss ratio:

$$IBNR_{net} = max(LR_{net} \cdot P_{net} - C_{net}, 0)$$

where LR_{net} is the net loss ratio, P_{net} the net premiums (difference between the accepted premiums and the premiums paid for Group covers), and C_{net} the net reported claim amount (including Group covers recoveries on claims paid and Group covers recoverables on file to file reserves). The condition $IBNR_{net} \geq 0$ is added for sake of prudence, that is, we do not anticipate a decrease of reported claims. The ceded IBNR are then calculated by difference:

$$IBNR_{ceded} = IBNR_{accepted} - IBNR_{net}$$

2. Bornhuetter-Ferguson method:

The Bornhuetter-Ferguson method can be applied on a net basis the same way it is applied on gross amounts, provided that one can use a claims development pattern. Claims development patterns are usually calibrated on gross triangles. It can be assumed that net claims developments follows the same rhythm as the gross ones. With this method, the net IBNR reserves is calculated as follows:

$$IBNR_{net} = (LR_{net} \cdot P_{net}) \cdot (1 - \alpha)$$

where $\alpha \in [0, 1]$ is the % of reported claim depending on the maturity of the UWY (the older the closer to 1). The ceded IBNR proceeds as previously expressed, with the constraint $IBNR_{accepted} \geq IBNR_{net} \geq 0$. The Bornhuetter-Ferguson method allows to release IBNR from an “a priori estimate” little by little, following a claims development curve.

3. Ratio methods:

Two ratios can be considered for the calculation of IBNR reserves' ceding rates:

- the premiums ratio: $R = \frac{P_{ceded}}{P_{accepted}}$

- the case reserve ratio: $R = \frac{\text{Case Reserve}_{\text{ceded}}}{\text{Case Reserve}_{\text{accepted}}}$

Ceded IBNR reserves is then expressed as follows:

$$IBNR_{ceded} = IBNR_{\text{accepted}} \cdot R$$

Case of pool quota-shares

The pools results are shared with the entities according to their quota-share by underwriting year.

3.3 Conclusion

We have covered several loss reserves estimation methodologies in a non-life reinsurance setting. Each method has its strengths and weaknesses according to the situations and loss data in/on which they are implemented. It is precisely where the experience and good practices of the reserving actuary is tho most valued. He is indeed expected to know exactly when to use one or another method and to be able to justify that choice.

IBNR is of course a component of paramount importance within AGRe's technical provisions, and the calculation of its best estimate for each reserving segment is the core mission of AGRe's reserving team. Nevertheless, other essential technical provisions components are to be assessed. For more details regarding those complementary reserves and how they are estimated at AGRe, the reader can refer to Appendix C.

4 Non-life reinsurance FX risk and the importance of a good representation of future cash flows in currencies

As the reinsurance platform of AXA Group, AGRe generates and receives a lot of cash flows in numerous currencies, referring to its various internationally based counterparties (AXA entities worldwide, international reinsurers and brokers, ...). Consequently, AGRe undergoes foreign exchange risk and seeks to consolidate and refine its roughly estimated knowledge on the future cash flows in currencies that will go through its business. Especially, AGRe looks for a more reliable representation of near future monthly cash flows for anticipating treasury requirements and thus be able to manage more efficiently its available cash. In addition, AGRe is also after a better representation of the amounts and durations of its liabilities in currencies, in order to be able to hedge and invest for the long-term more effectively.

We will detail in this section the emergence of and the forms taken by FX risk in non-life reinsurance, and how the different aspects of this risk can be mitigated or even removed in ideal situations. We will see why an accurate prediction of future cash flows and a trusty description of current liabilities in currencies is a powerful incentive for reinsurers' endeavors.

4.1 Introduction

Assuming the foreign exchange market is an efficient market, foreign exchange rates will always reflect the totality of the information available and will fluctuate in a random fashion as a function of the arrival of new information. The direction in which foreign currency rates will vary is in principle impossible to forecast.

Erratic fluctuation of exchange rates and the distortions they induce with regard to technical results and profit, as well as to loss accounts, is one of the main concerns for reinsurance companies. Reinsurance companies are particularly exposed to FX risk, mainly because of the geographical dispersion of covered risks on an international scale, which involves that they deal in a large number of currencies. AXA Global Re's portfolio is comprised of risks labelled into about 40 distinct currencies, five of which being dominant (EUR, GBP, USD, CHF, HKD) in terms of net liabilities. Thus, the total net assets of AGRe expressed in national currency (EUR) on the basis of assets and liabilities held in those various currencies are subject to significant fluctuations due to corresponding exchange rate fluctuations.

Below are the net reserves (IBNR + case reserves excluding Germany Quota-Shares) proportions in each material currencies at half year 2021.

Currency	IBNR proportion	Case reserves proportion	Reserves proportion
EUR	46.4%	84.5%	78.5%
GBP	14.6%	7.0%	8.2%
CHF	9.6%	4.7%	5.5%
USD	11.7%	2.8%	4.2%
HKD	11.7%	0.0%	1.9%
TOTAL	**94.0%**	**99.1%**	**98.3%**

Table 6: AGRe's net reserves proportion for the 5 top currencies

We can see that almost all AGRE's claims reserves are identified in those five currencies.

For reinsurance companies, as for all firms operating at the international level, foreign exchange risk is considered of paramount importance and is constantly under close scrutiny. It has given way to a careful management in the international financial market operators, but compared with other financial institutions and international firms, AGRe, as well as all reinsurers, encounters a certain number of specific problems in the management of foreign exchange risk. Essentially, these problems are linked with the random nature of the liabilities accepted and borne.

4.2 The foreign exchange risk of reinsurers

4.2.1 Accounting exposure and economic exposure

Foreign exchange exposure in a given currency is defined as the difference between the assets and the liabilities held in that currency. This difference (positive or negative) represents a net asset, the value of which, expressed in national currency, will fluctuate with the foreign exchange rates.

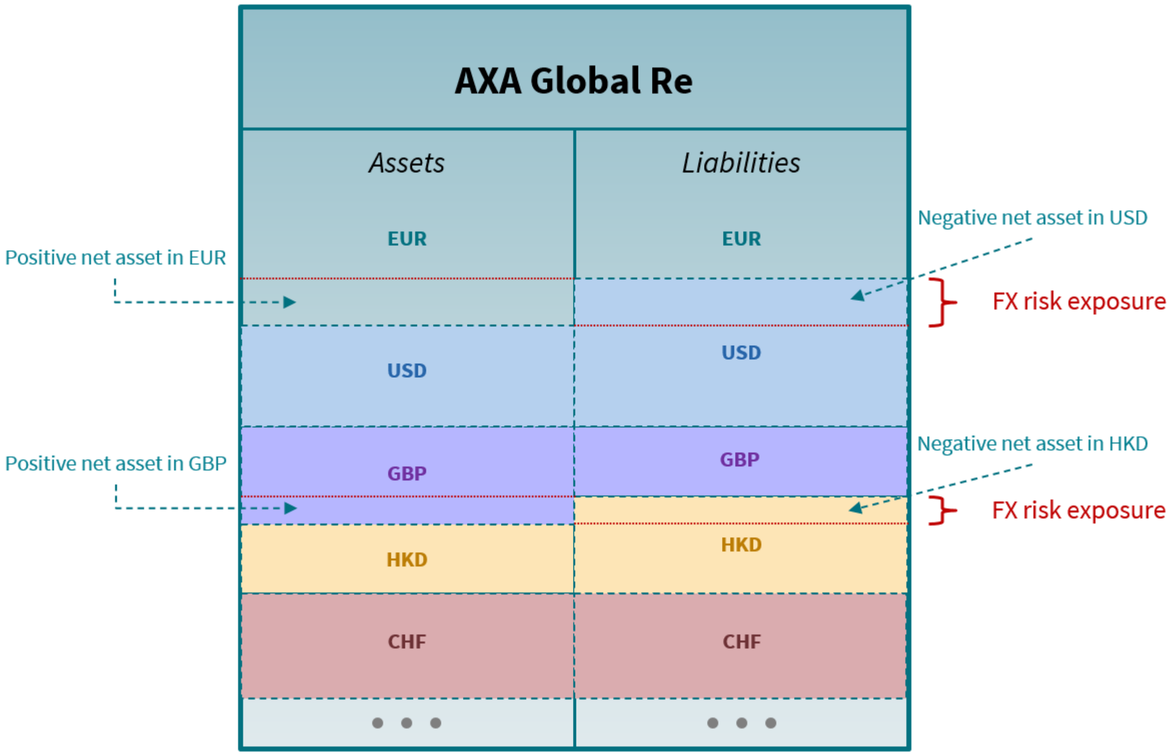


Figure 27: FX risk exposure and net assets in currencies

Converting those net values in original currencies into national currency is necessary for two reasons:

- first, because the assets expressed in national currency constitute the point of reference for an international based company which spreads its activities outside its own country with the aim of increasing its future profits (labeled in national currency). Foreign exchange rate fluctuations will alter the prospective value of future profits, thus resulting in an economic risk.
- second, because in order to reflect its activities and record the state of its assets, a firm with international operations must periodically exhibit the books of account and papers, whose values are all expressed in a single currency of account, generally being the firm's national currency. The accounting figures (representing the firm's image, performance, and soundness in the eyes of investors, creditors, shareholders, and competitors) are then liable to be affected by foreign exchange rate fluctuations, thus leading to an accounting risk.

The accounting risk thus relates to the exposure of the firm at a given moment (when balance sheet are drawn up), whereas the economic risk relies on the whole foreseeable future (developments in time of foreign exchange exposure and random distribution of future profits). Both foreign exchange fluctuations' direct effects (on the books of account) and indirect effect must be considered. If, because of foreign exchange profits and losses, trading results and reports published appear volatile, the firm will seem a risky investment to risk-averse investors. This is why, one cannot consider the accounting and the economic aspects of FX risk separately.

The European directive Solvency 2 offers an economic framework to take into consideration both risk notions together with their indirect and interrelated effects, within a time horizon of one year. This regulatory framework dedicates a special FX risk assessment sub-module within the broader Market risk module. FX risk measure is thus based on a shock applied to the Net Asset Value of the entities in their respective reporting currencies. Technically speaking, stochastic models simulate a PL on both assets and liabilities, derived from percentage variations on correlated currency values (with respect to the domestic currency). Those stressed conditions are also integrated into pricing formulas and models in order to reevaluate owned derivative products' values.

4.2.2 Accounting risk in a reinsurance company

The accounting risk is constituted by both the risks

- of distortion in the pattern of trading results of the firm (profits, turnover);
- that foreign exchange profits or losses will emerge when, for balance-sheet purposes, the various accounts in foreign currencies are converted into national currency.

In principle, the total foreign exchange loss or profit recorded by any firm at the end of its financial year will depend upon two factors:

- the foreign exchange exposure at the beginning of the financial year, multiplied by the percentage fluctuation of the foreign exchange rate during the year;
- the variations in this exposure, multiplied by the percentage fluctuation of the foreign exchange rate between the date of each variation and the date of the balance-sheet.

As an example, for a given currency (X) whose exchange rate vis-à-vis of the national currency (Y) at the end of a financial year is lower than all the levels reached during that year (say X/Y = 2 at the beginning and X/Y = 1 at the end with a steady decrease of X/Y along the year):

- a corresponding short exposure (negative net asset of say -1M X) at the beginning of the financial year and a decrease in that exposure during the year (say -1.5M X at the end of the year) can bring about a foreign exchange profit (here $(-1.5-(-1))\cdot(1-2)=0.5$M Y).
- a corresponding long exposure (positive net asset of say 1M X) at the beginning of the year and an increase in the exposure during the year (say 1.5M X at the end of the year) will give rise to a foreign exchange loss (here $(1.5-1)\cdot(1-2)=-0.5$M Y).

4.2.3 Economic risk in a reinsurance company

Exposure and liabilities

Before even seeking to manage FX exposures, one has to evaluate that exposure currency by currency. AGRe, as any reinsurer, cannot simply assess the assets and liabilities in currencies linked to each of the reinsurance treaties it has accepted (and potentially retroceded), and estimate the contribution of each of those treaties separately to its global foreign exchange exposure and risk.

A given proportional reinsurance treaty will bring about payments of reinsurance balances at a certain predetermined frequency (quarterly, half-yearly, or yearly) but the date of the payments, the balance amount and sign (credit or debit), is not known in advance. Timing of payments (if there are any) for a non-proportional treaty is even more unpredictable.

The same way insurance payments are generally performed after several time lags (reporting and settlement), reinsurance balances are reported by the ceding insurer with a time lag of possibly several months from the date of the closing of the quarterly - or half-yearly - account. A further time lag can then occur between the forwarding of the balance and the date of payment.

A reinsurer should then endeavor to achieve a reliable assessment of the total assets and liabilities emerging from the overall acceptances and retrocessions for each foreign currency within the scope of relevant homogeneous grouping of treaties, whose claims developments are estimated historically with sufficient confidence.

- "total liabilities" corresponding to commitments shown in the accounts (premium and claims reserves, debts payable) and more globally all the foreseeable payments to be made in a foreign currency.
- "total assets" refering to the assets in foreign currencies which have already been entered in the accounts (financial investments, technical deposits, and credit balances), but also all foreign exchange receivable in the foreseeable future.

Thus, if a given branch in a particular country is experiencing a sudden rise in its total losses, one's foreign exchange exposure should be reevaluated by revising upwards the estimate of one's commitments expressed in the currency concerned.

As we saw, at AGRe those groups of considered homogenous risks are called "Loss Ratio scopes", and liabilities (expressed only in euros) for each one of those LR scopes are carefully evaluated quarterly by AGRe's dedicated reserving team. This team is led to take into account new information about individual treaty's case reserves and adjust IBNR reserves, in accordance with latest loss experiences among ceding AXA entities.

All the issue relies exactly in the fact that a reinsurer has great difficulty in evaluating, even in a relatively global way, his economic foreign exchange exposure. It is indeed not always possible to build up sufficiently broad and well-diversified segments in each currency to be able to assess commitments in a precise manner. Similarly, it is challenging to produce reliable liabilities estimates (expressed only in national currency) at the aggregate level of homogeneous segments of risk, for then allocating those commitments down to currencies. In other words, one could assess liabilities rather accurately at a macro level (LR scope, i.e reserving segment for example) without consideration of their currency dimension, and share out those liabilities afterward by currencies in an reasonable way.

Another source of risk comes from the temporality of information availability, that is, the time lag that remains between the real liabilities situation and their evaluation by the reinsurer. In a general case, the direct insurer is kept continuously informed of all developments in his business, of the underwriting of new policies, of renewals of old contracts, of claims which arise and of reports of experts, and is therefore able to adjust his total liabilities in keeping with all this data and thus reduce the extent of this source of uncertainty. As for the reinsurer, by accepting the reinsurance treaties offered by ceding insurers, it participates in their commitments and thus assumes derived liabilities whose development patterns are not directly known to him. The reinsurer can only rely on the information furnished by the cedents, or on its own historical experience with similar liabilities. Moreover, the ceding entities perform assessments of their liabilities but often share only the information related to known reported claims subject to reinsurance recoverables in accordance with the treaty contractual clauses and within some agreed timeframe. Hence, the reinsurer must rely on less and delayed information for estimating its liabilities.

As far as AXA Global Re is concerned, the clauses of the treaties oblige the ceding companies to notify claims and send accounts (quarterly, semi-annually or annually) with specific deadlines (example: Q1 accounts must be sent no later than May 15). Regarding XS treaties, declaration thresholds are defined (generally 50% or 75% of the priority), i.e. the ceding company must notify AGRe (which notifies the retrocessionaires) as soon as it becomes aware of a loss exceeding this threshold.

A sound and regular flow of data from ceding entities to AGRe, concerning the positions linked to technical reserves and more generally reinsurance balances, is of crucial importance and is guaranteed by several technical accounting tools. Hence, AGRe can be fed systematically with useful indications regarding the developments in the business of its cedents. In this manner, it can have access to a precise depiction of its own commitments and how those commitments will evolve and develop.

A good knowledge of one's liabilities, materializing mainly through claims reserves, in each currency is therefore essential to quantify one's FX exposure and study one's FX risk. We just saw indeed that to better identify AGRe's liabilities with a view to address the issue of FX exposure and corresponding risk management, it is imperative to go beyond current practices, that is estimating IBNR at the aggregate level of LR scopes (AGRe's reserving segments). To ensure this indispensable prerequisite, we will present in next chapter a procedure providing a reliable representation of AGRe's reserves labeled in foreign currencies.

Currency configurations of reinsurance contracts and corresponding FX risk sources

Transactions (premiums and payments of claims) under a given reinsurance treaty can be either:

- in the same currency as the direct insurance transactions to which they relate. This situation is referred as reinsurance in original currency;
- in the currency of a third country in order to group together under one reinsurance treaty, direct insurance transactions of the same kind, but expressed in different currencies. A reinsurer can resort to this modality to cover in one single currency (its national currency or an international currency such as the dollar) a portfolio of risks localized in different parts of the world. The circumstance where a reinsurer offers a reinsurance treaty whose accounts are labeled in various original currencies but for which payments are realized in a single currency, is also considered reinsurance in the currency of a third country. In both cases, the reinsurer accepts commitments expressed in a currency different from the original currency, and the profits and losses under those reinsurance treaties will be affected by the fluctuations of the foreign exchange rate.

Figures 28 and 29 below, illustrate those two reinsurance configurations.

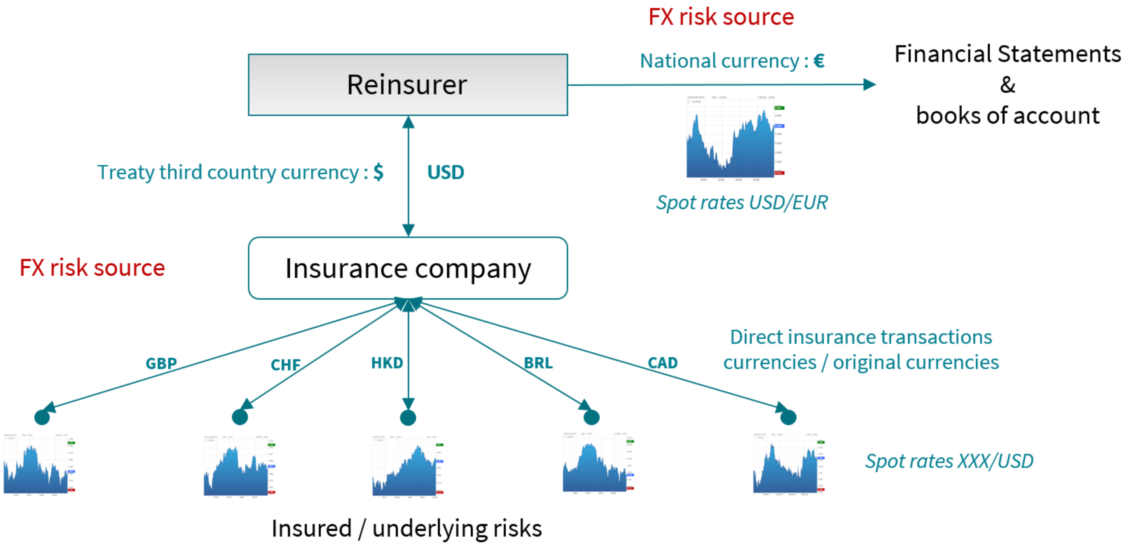


Figure 28: Reinsurance in a third country currency

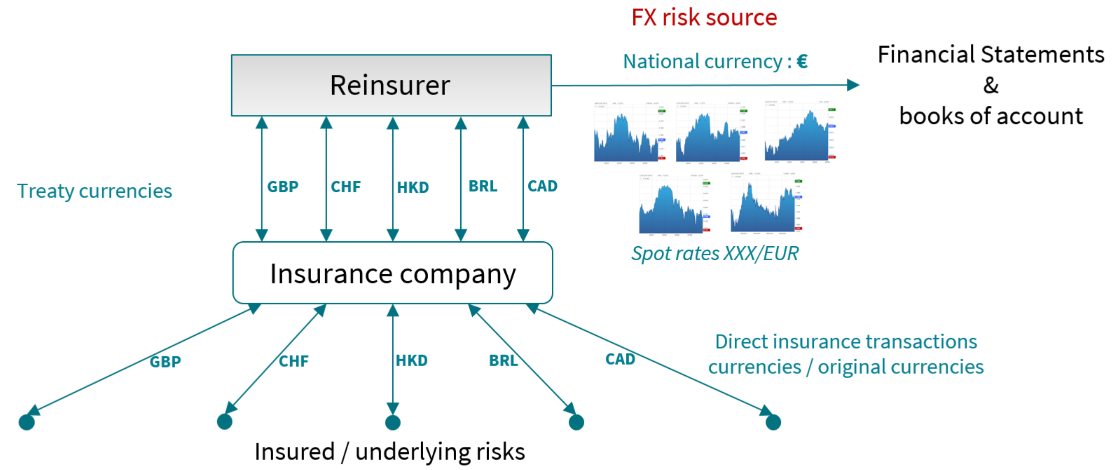


Figure 29: Reinsurance in original currency

As already displayed in a didactic manner in above figures, FX risk sources for reinsurance in original currency and reinsurance in the currency of a third country are described respectively by the two situations below:

- A reinsurer, with a national currency X, who accepts treaties expressed in currency Y concerning direct insurance transactions labeled in Y will undergo a risk stemming from fluctuations in the X/Y parity. When the Y rate changes, the value in X of all the assets and liabilities labeled in Y will vary accordingly.
- The same reinsurer who take out treaties in Y, concerning direct insurance transactions expressed in a variety of currencies, will equally be subject to the same FX risk, to which is added up the risk of fluctuations in the exchange rate between Y and the original currencies. The rises or falls of one of these original currencies in terms of Y engender a change in the nominal value of the assets and liabilities expressed in Y and thus in X.

Let us focus a bit on the third country currency reinsurance configuration. In non-proportional reinsurance, without a proper hedging of foreign currencies commitments through investments of received premiums in instruments tied to the corresponding FX rates, a reinsurer may undergo considerable foreign exchange losses and distortions in the technical results. Indeed, the payment of claims valued on the basis of original currencies occurs at a posterior and undetermined date in relation to the collection of premium in the treaty currency, and an appreciation of underlying currencies over that period can lead to higher payments than expected.

The covering in national currency, for instance, of commitments assumed in foreign currencies is equivalent to resorting to forward assets in national currency to cover liabilities in foreign currencies, which generates a position of foreign exchange exposure. To put it differently, reinsuring foreign currencies risks in national currency would mean for AGRe to sold foreign currencies spot for euros in order to buy them back later at an uncertain rate (receive premiums in euros and pay the euro counterpart of claims in foreign currencies: 2 different rates due to time lag between premiums and claims payments).

Figure 30 synthesizes where FX risks emanate according to the chosen reinsurance currency system.

	National currency	Treaty currency	Original currency(ies)
Reinsurance in original currency	X	Y	Y
Reinsurance in national currency	X	X	Z1, Z2, Z3
Reinsurance in the currency of a third country	X	Y	Z1, Z2, Z3

foreign exchange risk source

Figure 30: Reinsurance currency system and FX risk source

It is then clear that reinsurance in the currency of a third country increases the foreign exchange risk, because the profitability of accepted risks under this type of reinsurance treaty will be affected by the fluctuations in the exchange rates of both the treaty currency (in relation with the national currency) and the original currencies (in relation with the treaty 3nd country's currency). This reinsurance configuration can furthermore generate distortions in the technical results, since purely FX originating technical profit may be wrongly interpreted as underlying business technical profit.

Most inward treaties between AGRe and AXA entities are of the first type (reinsurance in original currency), but two exceptions arise for which all amounts are expressed, all accounts are made up and all payments are due in a stated currency (reinsurance in the currency of a third country):

- Local treaties (i.e. specific retrocession) with entities assuming portfolios in countries whose currency is too volatile. This is the case with AXA Turkey entity, for which the third country currency is the EUR, and Mexico one for which the USD is the only transaction currency.

- Group cover treaties, between AGRe and the reinsurance market, that mutualize different geolocalized portfolios made up of risks evaluated in many different local currencies. For those reinsurance agreements, the denominational, transactional and settlement currency is mainly the EUR (exception for Motor and Workers Compensation businesses' Group cover that is in GBP, and Hong-Kong portfolios' Group covers protections).

Figure 31 summarizes those currency flows in the light of AGRe risk acceptance and retrocession mechanisms.

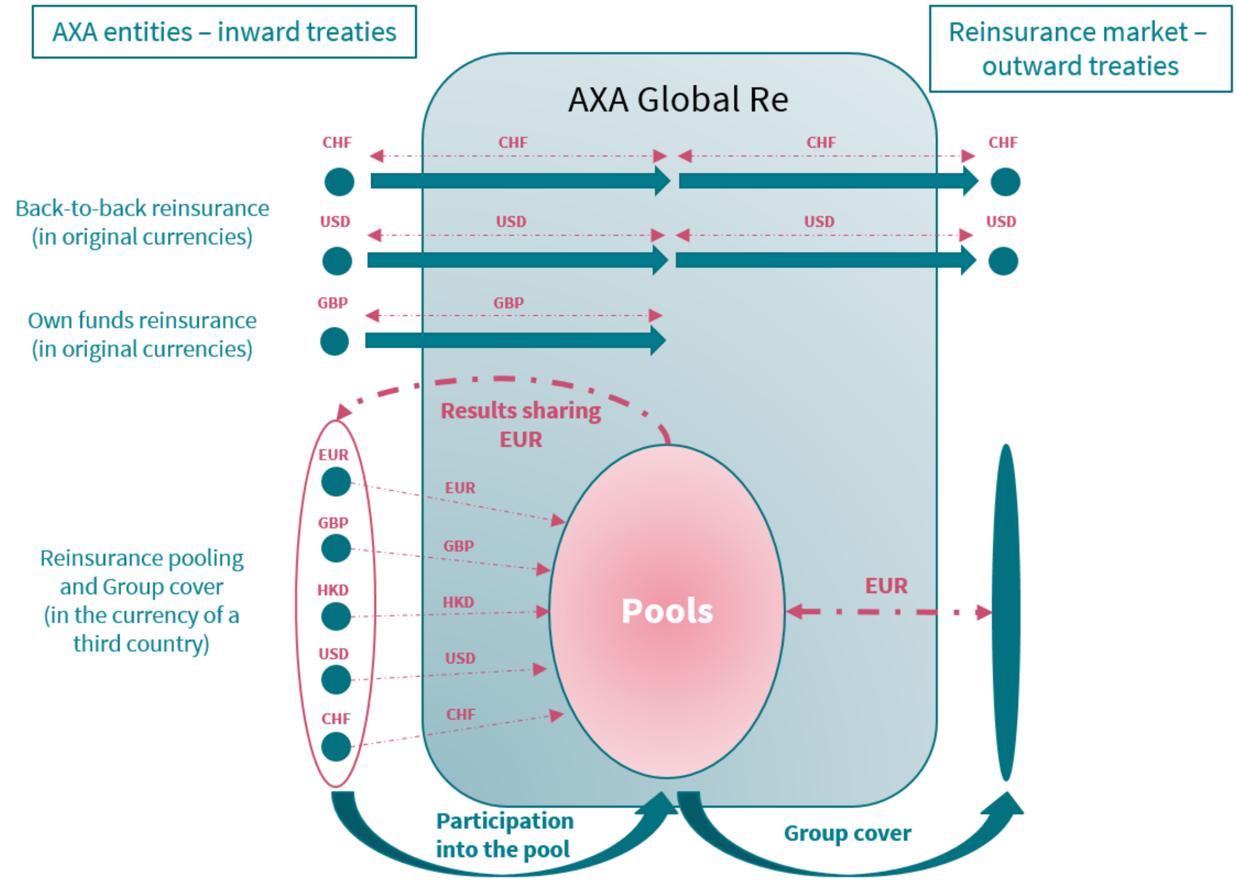


Figure 31: Currency configurations of AGRe's reinsurance contracts

A third source of distortion concerns non-proportional reinsurance. Let us take the example of an excess of loss treaty expressed in US dollars (in which the priority has been fixed in dollars), covering original insurance transactions in Mexican pesos (MXN). If a MXN claim is applicable to that USD treaty, the recovery in USD is then the excess equivalent value of the claim in USD (at the settlement date) over the treaty priority expressed in USD. An appreciation (or depreciation) of MXN against USD will give rise to a downward (or an upward) reevaluation of that priority valued in equivalent Mexican pesos. As a result, a rise (or a fall) in the parity MXD/USD will bring about a less (more) favorable reinsurance treaty from the point of view of the reinsurer.

To give a concrete example, AGRe catastrophe excess of loss reinsurance master agreement, covering losses occurring on risks written in Colombia, Greece, Saudi Arabia, Oman, United Arab Emirates, Bahrain, Kuwait, Qatar, Lebanon, Mexico, and Turkey, stipulates that:

- "The Deductible and the Limit shall be converted into the currency concerned at the rate of exchange as used by the Adherent and ruling on the date of occurrence of the loss or as defined in this reinsurance agreement;
- The balance of any loss payment in excess of the Deductible shall be converted from the currency in which the loss was settled into the currency in which this Reinsurance Agreement is written, in respect of the Adherent concerned, at the rate of exchange as used in the accounts books of the Adherent on the date or dates of settlement of the loss by the Adherent."

One last point regarding this time proportional reinsurance. As already mentioned, there exists a time-lag between the date of closure of a proportional reinsurance account and the date on which this account is furnished to the reinsurer. During this period, FX rates fluctuation is liable to affect the technical balance value. Since the cedent establishes the reinsurance balances and can decide when to report it to the reinsurer, within a determined time interval, it is then able to choose, to a certain extent, the exchange rates for the conversion of the original currencies into the currency of payment. The reinsurer has the possibility to protect itself against the possible distortions caused by this one-sided advantage of the ceding company emerging in third country's currency reinsurance. At the signing of the reinsurance treaty, it can be agreed that the conversion rates used will correspond to the parities prevailing on a specified date, whether the date on which the accounts are closed, or the claims occurrence or even settlement dates.

As for retrocession, a reinsurer must strive to maintain the parallelism between inward treaties and its outward counterparts, and mirror its retrocessions transactions' FX rates practices on their corresponding acceptance transactions ones.

4.3 FX risk management in a reinsurance company

4.3.1 Procedures of foreign exchange risk management available to the reinsurer

From a general point of view, a risk cartography englobing all potential risk sources liable to impact one's business is outlined. For each of the risk identified, a corresponding risk appetite is defined and a quantification of said risk is then carried out. If a given risk measure goes out of the bounds specified by the associated risk appetite, efforts are made to mitigate it. So, after pinpointing and quantifying its accounting and economic foreign exchange exposures, a reinsurer must first define its FX risk appetite, that is, the level of FX risk it is prepared to accept in pursuit of its objectives, before action is deemed necessary to reduce that risk. In other words, it should determine to what extent it wishes to bear FX risk. Only once its risk appetite is clearly outlined, should the reinsurer become interested in the available procedures for the reduction and transfer of FX risk. It can then choose within a wide range of possible instruments and methodologies of foreign exchange risk management in order to mitigate or eliminate FX risk according to its maximum risk acceptance and at the lowest possible cost.

AGRe's FX risk appetite is defined on the basis of an indicator proceeding from a 10% FX rates variation shock (appreciation or depreciation) applied to all held net asset in currencies. As a recall, the calculation of the FX risk is performed through an internal model providing the level of losses on AGRe's balance sheet linked to FX fluctuations, and in line with AGRe's portfolio of assets and liabilities at a given moment as well as the positive and negative correlations existing between the fluctuations of the different currencies flowing to and out of AGRe.

First of all, it must be pointed out that there exist some limits to reducing FX risk in a reinsurance business. To prevent a fall of acceptance (original) currencies from having a negative impact on its results, a reinsurer cannot simply confine his operations to transactions in national (or a unique stable) currency, that is, either in limiting the geographical range of accepted risks or in expressing all reinsurance treaties in national (or a unique stable) currency. Both the vital geographical dispersal of risks for a reinsurer and the increased FX risk compounded by the system of third country's currency reinsurance, get in the way of those FX risk mitigating options. Even if the reinsurer's assets and liabilities were to be expressed only in national currency (and not in original currencies), it would not prevent them from fluctuating in line with the exchange rates of the underlying insurance transactions original currencies and would have the same effects (distortion of technical results and increased foreign exchange risk) as reinsurance in the currency of a third country.

To summarize, a reinsurer has no means to forestall the influence of foreign exchange rate fluctuations on movements in its business turnover and book profits, introducing distortions as a component of the accounting risk. It can only seek to reduce the economic impact through appropriate comments made at the time of its annual report publication.

However, other procedures of foreign exchange risk management can be used. To reduce its FX risk, a reinsurer can:

- rise premiums to offset possible foreign exchange losses (not a common practice);
- tune its outward payments' maturity dates;
- adapt its acceptance policy including FX risk constraints;
- share the burden of FX risk among both parties (insurer vs reinsurer or reinsurer vs retrocessionaire) through the choice of the treaty currency during the negotiations preceding the conclusion of a contract;
- operate in the spot or futures markets to adjust his foreign exchange exposures;
- buy optional currency futures contracts, which comport a choice between two pre-established dates for the delivery of a given currency under the underlying futures contract (can serve as a hedge for uncertain maturity date debts receivable or debts payable expressed in foreign currency);
- more generally, enter into relevant financial investments.

4.3.2 Congruence

The simplest and most straightforward rule of FX risk management is that of congruence. Statutory congruence corresponds to the situation where all current foreign exchange exposures are nil, and can be achieved through a policy of financial investments such that, for each currency, the assets will cover exactly the liabilities (statutory values must coincide). Reaching congruence is a process that removes both FX profits and losses, in addition to presenting the following positive aspects:

- it is not very costly to implement for the reinsurance business naturally achieves it in part, since technical deposits ensure that the reinsurer maintain in the currency of acceptance all assets invested as technical reserves;
- it reduces FX transaction costs.

Nevertheless, a number of limits and hindrances in the application of that method is to be noted.

- First, congruence does not address risks of distortion of technical results and risks associated with reinsurance in the currency of a third country.
- Second, some difficulties can emerge along a congruence procedure, especially when it is not applied to the currencies of developed countries which have a wide range of financial instruments actively traded. Attempt at congruence is indeed not recommended for currencies offering only low liquidity instruments, and thus exposing the reinsurer to a risk of capital loss when selling rapidly its assets to settle large claims.

- In addition, it is sometimes impossible to follow a policy of congruence in certain countries, since regulations in force oblige reinsurers to hold an amount in the foreign currency concerned representing over 100% of the value of the accepted liabilities.
- More importantly, if the reinsurer hasn't an accurate knowledge of the extent of its liabilities currency by currency, which is generally the case, it neither will have exact knowledge of its FX exposure nor will it know whether congruence is achieved or not. Only the accounting risk is manageable through the principle of congruence by hedging systematically the opening balances, since the accounting value of the liabilities is known. Regarding the economic risk, the reinsurer can only tend towards congruence on the basis of a subjective estimate of its liabilities amounts and future developments (i.e. timings of future cashflows issued from those liabilities). Therefore, foreign exchange losses can still arise if the liabilities (and its cash flows developments) have been overestimated (or underestimated) in a currency which is falling (or rising) in value.

Because of the random nature of its liabilities, it is extremely difficult for a reinsurer to forecast exactly the dates and amounts involved in the transformation of those liabilities into effective settlements, thus allowing it to resort to perfect hedging through congruence. However, it can always achieve an approximate hedging to cover the overall foreign exchange risk arising from its exposure in foreign currencies at a given date t. In the case that the net asset held in a particular currency is too substantially positive (negative) at t, the foreign exchange exposure can be adjusted by selling (buying) that currency forward with a predetermined maturity that does not necessarily reflects the various, a priori unknown, maturities of the underlying assets and liabilities and their interdependences.

This is exactly what is done by AGRe: it currently follows a congruence procedure which aims to neutralize any FX impact springing from the accounting risk from one year closing to the next by counterbalancing any asset/liabilities mismatch, valued in French GAAP (Generally Accepted Accounting Principles) standard at year Y. Hedging against FX risk, performed only on that local French GAAP basis, used to be a requirement of the Code des Assurances (R331). AGRe indeed resorts each year to a 1-year-horizon forward hedging in order to match global net assets valued in each currency via French GAAP. Every pools and local treaties are considered in the computation of those global net assets by currency. For each currency held, a global congruence is then performed at year Y+1 opening on the basis of the aggregated net asset denominated in that currency, entering into a 1-year forward contract to sell the net asset equivalent amount (positive or negative) of that currency at the forward rate at that time. Figure 32 shows that forward hedging process executed yearly by AGRe.

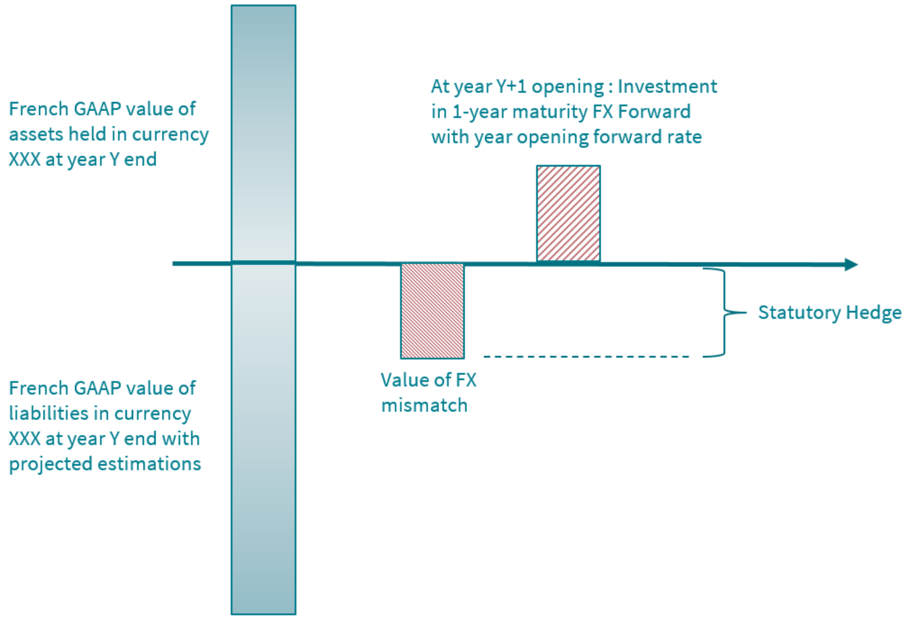


Figure 32: FX Forward hedging process at AGRe

At AGRe, mismatches below 1 equivalent million euros are tolerated. Expressed differently, any currency showing a French GAAP valued absolute net asset of more than this amount in equivalent euros is handled through the above forward procedure, which constitutes a conservative and risk averse approach that does not take advantage of that currency favorable movements.

Once again, only liabilities expressed in currencies on the basis of French accounting principles were considered for this hedging procedure. Outside this framework, we will not question the confidence of AGRe in its liabilities' estimates and will assume that those liabilities amounts are perfectly known within reserving segment (at the LR scope level). What remains to be done, is to set up and carry out a reliable scheme for the evaluation of liabilities at the currency level within each LR scope, as a requirement for the application of congruence.

To sum up, usually inaccurate representations of liabilities in currencies and low liquidity financial instruments in certain minor currencies, get in the way of an efficient congruence. Those drawbacks motivate the consideration of the diversification of foreign exchange exposures (and risk) as an alternative solution. But if a reinsurer were able to predict (to a certain extent) the durations and corresponding amounts of its liabilities labeled in currencies, it would become possible to hedge accurately its futures exposures in foreign currencies by means of congruence, together with the possibility to invest its available cash in those currencies.

An ideal configuration would be an assets/liabilities close-to-perfect matching (in amounts, durations and currencies), which would ensure the exact liquidity needed in each currency at the time the future cash flows take place, along with a cautiously chosen series of investments of the net assets in excess. Let us take a look at the following diagram, presenting that ideal situation for an arbitrary medium-tail line of reinsurance and a given currency. In the case that current commitments materialize through settlements only lately, the assets needed to face those payments won't be required immediately and can be dedicated to adapted investments, such as different maturities bonds in the currency considered, that will yield a return on those otherwise immobilized assets.

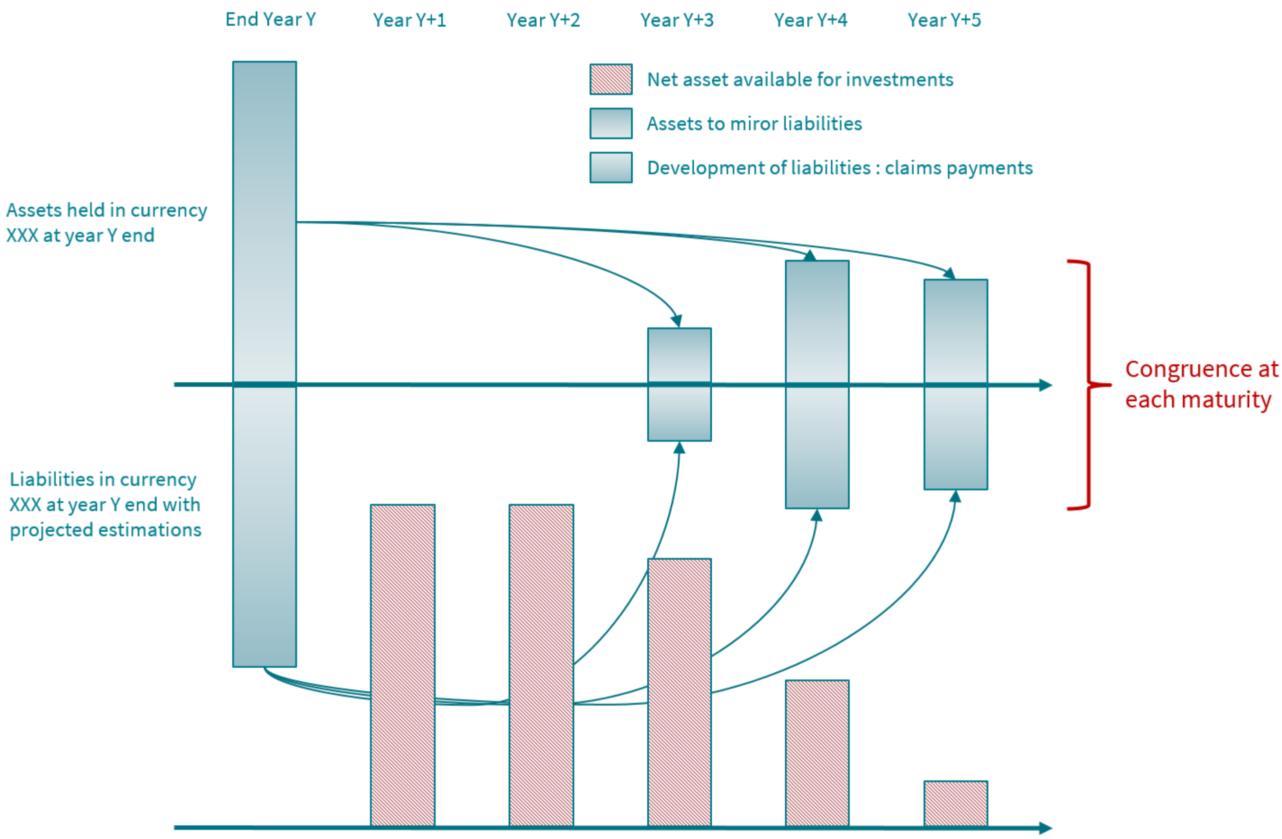


Figure 33: Perfect congruence situation and available cash for investments oportunities

Longer the claims development of a given segment, longer the availability of cash in currencies, higher the number of possible investments and higher the final yield. This way, one could clearly avail oneself of the ability to get a proper insight into one's liabilities value and pace of materialization. Such a capacity would directly translate into a perfect congruence FX cover as well as profits.

A mix of the 1-year forward hedging method (for low-return currencies or short-tail reinsurance lines) together with available assets investments (for high-return currencies and long-tail reinsurance segments) could thus constitute a reasonable and fruitful FX strategy for a reinsurer such as AGRe.

4.3.3 The diversification of foreign exchange exposures

This strategy comes from the simple observation that if one is not absolutely confident about its evaluation of liabilities in currencies, trying to get rid of FX risk matching exactly uncertain assets and even more uncertain liabilities in currencies, is an illusory goal. It would amount to give up the benefit of foreign exchange profits not balancing it by an elimination of the possible losses. One could instead attempt at compensating the foreign exchange losses on certain currencies by means of the profits on others. In fact, the total elimination of the FX risk, even if it were feasible, does not necessarily correspond to an optimal goal from the point of view of the theory of choices under uncertainty. In an efficient foreign exchange market, a risk premium proportional to the risk resulting from a diversified foreign exchange exposure is served. In other words, the acceptance of a FX risk is to be remunerated.

Positive and negative correlations between the fluctuations of various parities make possible a significant reduction of the risk arising from a combination of foreign exchange exposures by comparison with the sum of the risks of individual exposures. For example, a positive exposure on the dollar is liable to be offset by a negative exposure on a strongly correlated currency. The implementation of a policy of FX exposures diversification can both reduce the overall risk and generate a stable profit, trough the identification of an optimal combination of foreign exchange exposures (with random returns) reaching a maximum of expected utility subject to specific constraints, and the application of a course of action achieving this combination. For instance, if the estimated liabilities in dollars finally turn out to be less than expected, the share allotted to the dollar in the reinsurance portfolio will in fact prove higher than it would have been in an optimal portfolio.

The congruence recipe is not automatically ruled out by a policy of diversification of foreign exchange exposures. It rather constitutes a particular case among a multitude of other feasible solutions.

AGRe accepts indeed to bear a foreign exchange risk to a certain extent, as we saw defined by its risk appetite. However, until recently AGRe had not been explicitly on the lookout for yield springing from a portfolio optimization dynamic, and efforts toward this direction had not been formalized.

A restriction to a diversification strategy: technical deposits

Technical deposits requirements get indeed in the way of any strategy of FX exposure diversification, since most of the reinsurer's assets are blocked in deposit accounts in favor of the cedents and thus cannot be used at will to fulfil this strategy. At AGRe, technical deposits are made on a case reserves basis. That is, if a cedent entity reports a claim valued at $50m and subject to recovery on an XS $10m treaty, $40m will be payable to the cedant upon settlement of the claim. Those $40 millions should be "deposited" to ensure that they will be available when the claim is settled. AGRe business is in reality a bit more complex: technical deposits that are to mirror approved file-to-file reserves with respect to ceding companies, are actually mutualized by currency up to a certain limit (not necessarily equal to the actual cases reserves in those currencies). Those "pooled technical deposits" grant unexploited flexibility to carry out and optimize investments. It can moreover dispose of reserves which are not subject to a deposit requirement and benefits from the various procedures for reducing and transferring the foreign exchange risk as well as choose carefully how to calibrate its own portfolio of FX exposures.

This apparent hindrance can be overcome by mean of letters of guarantee (the issuing bank replaces the reinsurer and assumes its liabilities in the event of a default by the reinsurer up to the amount of the letter). Supplying the ceding insurers with this form of warranty issued by first-class banking institution, enables a reinsurer to get around technical deposits requirements and facilitates its FX risk management policy since it is then free to dispose of and invest the amounts corresponding to the reserves.

At AGRe the mechanism of letters of guarantee is only used on the retrocession side of its business (from retrocessionaire to AGRe). Along with the posting of collaterals, it contributes to the reduction of counterparty risk and reserves.

4.4 Conclusion

We have just seen at length what drives the willingness of AGRe, and potentially of all reinsurers, to build a reliable representation of its own liabilities in currencies, and beyond that, to devise a sound and systematic methodology to predict to a certain extent future inflows and outflows of cash in currencies. This keenness takes root in the conceivably substantial rewards resulting from a perfect (and utopian) knowledge of future transactions in currencies flowing through ones business, namely:

- the ability to reach perfect congruence and thus protect oneself totally against FX risk,
- while being able to invest optimally the available cash;
- or the acquisition of a clear view ahead, crucial to achieve an optimal combination of foreign exchange exposures reaching a maximum of expected utility;

In this chapter, we laid emphasis mainly on yearly time units, but the usefulness and possible far reaching spin-offs of a quality estimation of future cash flows are unaffected for smaller time scales and time units, such as monthly prediction over a one-year horizon.

Throughout this thesis first part, we strove to lay down solid enough foundations to finally be able to get down to the nitty-gritty. We now possess a clear sight on the environment surrounding - and the direction taken by - our research project, of which we will present in next two parts the developments and implementations as endeavors towards previously brought up objectives.

Part II
Technical cash flows short-term prediction

We have divided the ambitious objective of forecasting future technical cash flows into two sub-objectives:

- a 12-month horizon monthly projection;
- a run-off yearly projection on the basis of current reserves.

This thesis' second part will be dedicated to the first goal, namely the pursuit of predicting cash flows related to technical accounts materializing within the next twelve months, on both the acceptance and retrocession sides of AGRe's business. Let's begin with a prelude in order to introduce the types of cash flows that circulate through AXA Global Re and define the scope of what we seek to achieve here.

As displayed by Table 7 as well as Figures 34 and 35, inflows and outflows of cash at AGRe (made up of different currencies) are classified into five budget groups ("GENERAL EXPENSE ALL", "TAX", "TECHNICAL ACCOUNT", "FINANCIAL" and "INVESTMENTS"), themselves composed of budget sub-categories.

BUDGET CATEGORIES	BUDGET GROUP	FLOW	DESCRIPTION
DIVIDEND RECEIVED	GENERAL EXPENSE ALL	Cash in	Dividend coming from AXA Global Broker (in charge of placing retroceded risk in countries where AGRe do not have certification)
OTHER INCOME	GENERAL EXPENSE ALL	Cash in	
GENERAL EXPENSE	GENERAL EXPENSE ALL	Cash out	
GROUP INSURANCE	GENERAL EXPENSE ALL	Intra-group	
STAFF COST	GENERAL EXPENSE ALL	Cash out	Wages and employee benefits
OTHER TAX	GENERAL EXPENSE ALL	Cash out	TVA, C3S, CVAE and CFE
CORPORATE TAX	TAX	Cash out	Corporate income tax calculated on the basis of AGRe's accounting profit at year Y-1
DERIVATIVE	TECHNICAL ACCOUNT	Cash out	Linked to CAT BOND issued by AGRe
MDP GROUP COVER	TECHNICAL ACCOUNT	Cash out	Minimum and deposit premium related to Group covers taken out at year Y renewal
MDP -	TECHNICAL ACCOUNT	Cash out	Minimum and deposit premium related to retrocession contracts taken out at year Y renewal
MDP +	TECHNICAL ACCOUNT	Cash in	Minimum and deposit premium related to acceptation contracts taken out at year Y renewal
POOL ACCOUNT	TECHNICAL ACCOUNT	Intra-group	Premiums recorded - claims recorded - commissions recorded + stock of year Y-1 deposits - stock of year Y deposits (deposit = claims reserves to be paid and unearned premium) paid to (or received from) each participant in the pool
QUOTA SHARE	TECHNICAL ACCOUNT	Cash in	QS Germany + QS Mexico
TECHNICAL ACCOUNT -	TECHNICAL ACCOUNT	Cash out	Claims, adjustment premiums, commissions and cash calls
TECHNICAL ACCOUNT +	TECHNICAL ACCOUNT	Cash in	Claims, adjustment premiums, commissions and cash calls
BANKING FEES	FINANCIAL	Cash out	
BANK INTEREST	FINANCIAL	Cash out	
CAPITAL INCREASE	FINANCIAL	Cash in	Increase AGRe's capital to face its liabilities or to increase its underwriting capacity
CAPITAL DECREASE	FINANCIAL	Cash out	
DIVIDEND	FINANCIAL	Cash out	Dividend distribution to AGRe's shareholders
TREASURY FX TRANSFER	FINANCIAL	Intra-group	
TREASURY TRANSFER	FINANCIAL	Intra-group	
COUPONS / NOTIONAL	INVESTMENTS	Cash in	
VENTE	INVESTMENTS	Cash in	Before maturity financial asset selling
INVESTMENT	INVESTMENTS	Cash out	Investments according to year Y planning

Table 7: Types of cash flows at AGRe and description

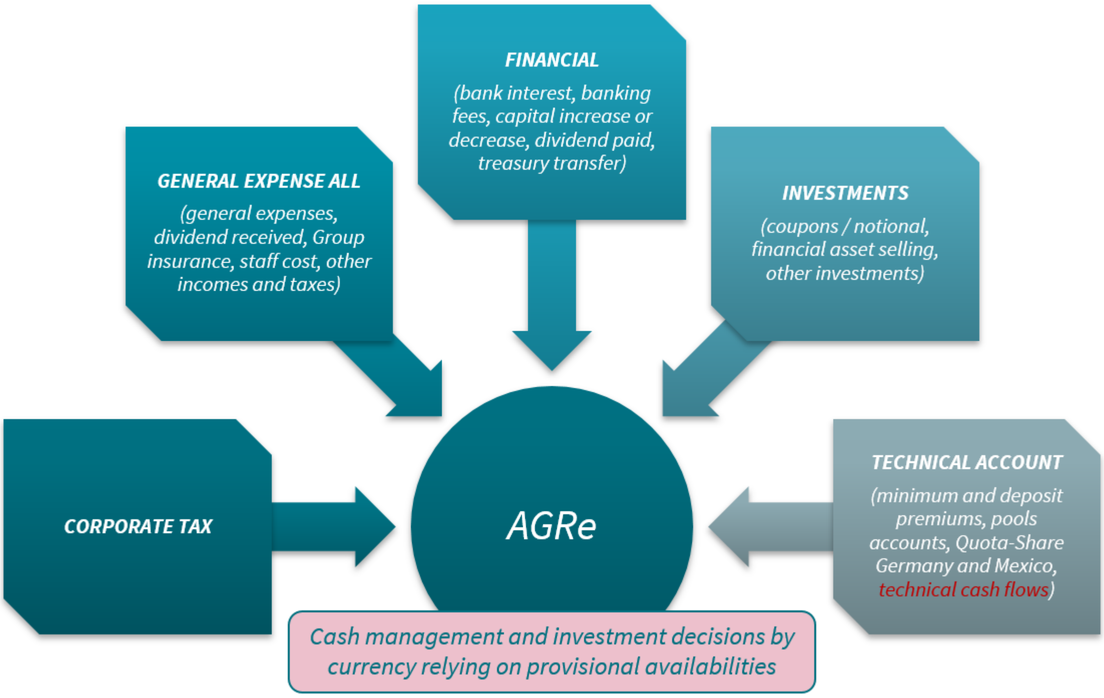


Figure 34: Types of cash flows at AGRe

AGRe treasurers are rather at ease with transaction timing and amount values regarding cash flows comprising all account headings but the "TECHNICAL ACCOUNT" one.

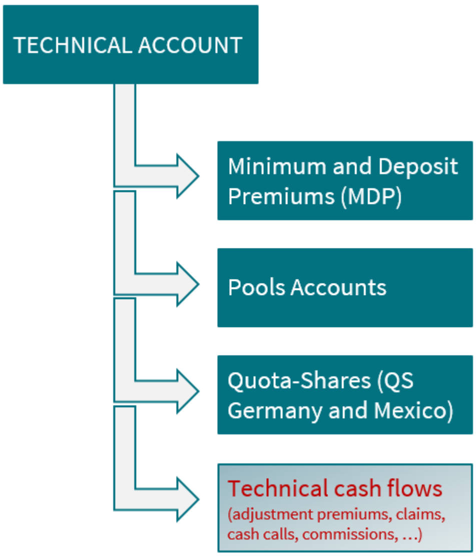


Figure 35: Technical account lines

"TECHNICAL ACCOUNT" budget group splits into the following lines of budget:

Minimum and Deposit Premiums (MDP)

When calculating, at the beginning of a given reinsurance period, the premium to be paid by a cedent entity to a reinsurer for a non-proportional cover, a reinsurance premium rate determined by the reinsurer is applied to the Estimated Gross Net Premium Income (EGNPI) provided by the cedent, thus obtaining the associated MDP. It corresponds to a minimum premium because its the least premium that the reinsurer will accept in return for the protection bought by the ceding company. It is a deposit, since it is expected to be paid to the reinsurer at the start of the reinsurance period.

At the expiry of the considered reinsurance agreement, the reinsured advises the reinsurer on the final premium income i.e. the total premium amount written by the cedent and subject to the considered reinsurance treaty (Gross Net Premium Income or GNPI). Since the MDP, collected at the beginning of the reinsurance agreement period, is based on estimates, the actual reinsurance premiums to be collected from the ceding company needs to be ascertained, applying the rate on the GNPI, thus leading to the adjustment premium defined as the difference between this actual reinsurance premium and the MDP. In the case that the actual reinsurance premium is greater than the MDP, the cedent will pay the additional adjustment premium to the reinsurer. On the contrary, the reinsurer will generally not refund the difference, as the MDP agreed and charged at the beginning of the reinsurance period coincide with the minimum premium.

Pools Accounts

Retrocession balances linked to the pool quota-share system on pools (mainly the pool Property) run by AGRe and retroceded in part to all participating AXA entities. Those cash flows are related to both reserves (case reserves and IBNR) deposits made by AXA ceding entities, which share the total result of the pool via quota-share according to their participation in that pool.

Quota-Shares (QS Germany and QS Mexico)

The German Quota-Share treaties are specific treaties underwritten with AXA Germany, which aim at reducing its equalization reserves. The equalization reserves are long-term reserves that an insurance company keeps to prevent cash flow depletion in case of significant unforeseen catastrophes and aims at covering the large losses resulting from a disastrous event, such as a flood, an earthquake, a massive storm or fire hitting an area where the insurer has insured several properties.

This set of treaties are managed as per clean cut accounting practice. Clean-cut Quota-Share reinsurance treaties are proportional treaties that cede business on a financial year basis and are associated with incoming and outgoing premium and claim portfolios. That is, liability is withdrawn from AGRe at the end of the expiring treaty year by premium and loss withdrawals and the same liability is transferred to the next treaty year by premium and loss entry. Concretely, this clean cut materializes in a single balance cash flow between AGRe and AXA Germany at each year end.

Likewise, the Mexican Quota-Share is a new reinsurance program put in place in 2019 which provides a protection together with a substantial reduction of reserves for natural catastrophe (mandatory within AXA Mexico according to the local regulation). This QS is not handled through clean cut.

"Technical cash flows"

What we will designate as "technical cash flows" throughout this entire short-term prediction study corresponds to all cash flows classified into the TECHNICAL ACCOUNT that are not part of MDP, pool accounts and Quota-Shares treaty. In complement to that exclusion definition, those cash flows are mainly related to payments of claims, adjustment premiums, commissions and "cash calls". The concept of "cash call" refers to a reinsurance contract provision, common in proportional contracts, which allows a reinsured company to make claim and receive immediate payment for a large loss without waiting for the usual periodic payment procedures to occur.

Minimum and Deposit Premiums (either paid by AXA entities to AGRe or paid by AGRe to external reinsurers), Pools Accounts as well as Quota-Share balances take the form of cash flows, the size, currency denomination and schedule of which are well anticipated by AGRe treasury teams. Striving to predict those inflows and outflows of cash, through elaborate models, would then have little added value. That is why we will rule them out of our study, and only focus on the attempt to determine, to some extent, both the rhythms and sizes of near future "Technical cash flows" in currencies.

This amounts to an effort to seek to improve current short-term projections of cash and allow treasury professionals to better foresee and prepare for cash needs in the distinct currencies composing AGRe business transactions. As we saw in the introduction, a more reliable vision of future cash requirements in currencies adds up to more cash investment optimized opportunities and a higher profitability from AGRe's assets.

One last point before entering into the technical considerations of the resolution of above issue: let us review succinctly AGRe's claims treatment process. At AGRe, claims are handled according to the following steps:

1. claims notice / loss advice (by claims department);
2. technical report (by claims department);
3. technical account writing (claims accounting figures + technical annex) performed by accounting department which generate automatically an entry in AGRe's accounting and financial systems (the date of which we will call the writing date);
4. balance global settlement with respect to a third party (possible netting with commissions and adjustment premiums).

5 Technical cash flows prediction from existing balances

A first angle of attack to address the problem of forecasting next twelve months' cash flows, consists in striving to predict settlements delays from already existing, recorded and not yet settled account balances, that is estimate the time intervals between today's date and actual inflows and outflows stemming from opened account balances as of date.

5.1 Data description and descriptive analysis

5.1.1 Financial database

To answer the above stated problematic, that is, to project existing recorded financial balances into cash settlements, we will have at our disposal a "financial database" that encompasses all settled and not yet settled financial balances between AGRe and its counterparties. This database provides indeed historical AGRe's signed balances data, with respect to ceding AXA companies as well as with respect to external reinsurers to which AGRe retrocedes part of its portfolio. This historical database includes all accounting balances going back to 2008 and is extracted from AGRe's systems every two weeks. It is comprised of approximately 350 000 observations (200 000 for the scope we are interested in).

Each line of this database thus represents a financial transaction (already carried out and closed or soon to be settled or a long time opened and without prospect of being reconciled and solved out), in a determined currency, between AGRe and a given counterpart participating in that transaction. Those balances can be made up of many components and represent a netting of possibly several sub-balances (claims payments along with adjustment premiums and commissions). In this chapter we will analyse and endeavor to make the most of those available financial data.

Lets have a rapid overview of the nature of the variables making up that database. It contains contractual information associated to each single balance, such as the AGRe's juridical entity involved (basically AGL or AGPC), the third party's identification code (namely the "BUID" standing for the business ID), name and place in regard to AGRe's business (cedent or retrocessionaire), to which that balance is allotted, together with the reinsurance treaty and section concerned (their ID numbers as well as the treaty's perimeter, i.e. Group cover or Local, and underwriting year).

VARIABLE NAME	**DESCRIPTION**
type_compte	Balance account types: "Appel au comptant" (cash call), "Définitif / réel", "PB / PP" (participation benefits), Primes provisionnelles (PMD), "Rachat / Commutation" (novation, commutation), "Spécifique"
accept_retro	Whether the considered balance is connected to an acceptance ("Accept") or a retrocession ("Retro") kind of risk transfer from the perpective of AGRe.
date_debut_periode_compte	Beginning date of the accounting period
date_fin_periode_compte	Closing date of the accounting period.
compte	Account ID number to which the balance is assigned (218 different accounts on the acceptance side and 639 on the retrocesion side as for AGPC and AGL).
libelle_compte	Account name to which the balance is assigned (for example "C/C AXA CAMEROUN", "C/C AXA CS GERMANY", "C/C SCOR / FR", "C/C SCOR GLOBAL LIFE ZURICH /CH", "C/C SWISS RE /US", "C/C XL RE BRASIL /BR", ...).
notation	Subjective score given by the accounting teams to external reinsurers (111, 222, 333 or 444 principally) reflecting settlements process spead and simplicity between AGRe and its external retrocessionaires. Informed if the balance is a retrocession one (i.e. related to a retrocessionaire).
devise	Currency in which the balance is settled (in line with the corresponding contract's section).
date_comptable	Balance's account record date.
date_ecriture	Balance's account writing date.
date_reglement	Balance's settlement date.
montant_initial	Balance's amount expressed in original currency.
montant_initial_euro	Balance's amount expressed in euro.
montant_non_solde	Not yet settled balance's amount expressed in original currency. Either equal to the initial amount or not informed if already settled.
montant_non_solde_euro	Not yet settled balance's amount expressed in euro. Either equal to the initial amount or not informed if already settled.

Table 8: Financial historical database main variables

Let's recall that we made the decision to filter out financial balances associated to MPD, pools accounts and specific Quota-Shares treaties (Germany and Mexico).

5.1.2 General accounting and contractual databases

In addition to financial historical data, let's now introduce another rich source of data, part of which we will explore and harness throughout this thesis: the accounting and contractual databases. Three of them exist:

- one picturing AGRe's risk acceptance point of view (inward treaties);
- one dedicated to AGRe's risk retrocession point of view (outward treaties);
- one issued from the two previous databases and corresponding to AGRe's net risk point of view (acceptance + retrocession).

Their names' prefixes are respectively ACC_ (acceptation), CES_ (cession) and NET_ (net of retrocession). Those prefixes are followed by the accounting closure type (HY for half-year or FY for full year) and exercise year. That way, at the closing of half-year 2021 accounting period, the technical databases describing AGRe's situation, relatively to its accepted, retroceded and net business contracts and accounts, were named ACC_HY_2021, CES_HY_2021 and NET_HY_2021 respectively.

Those three databases contain mostly the same variables, thus rendering the same information outlines from three different perspectives. Those variables comprise three main categories: general variables and labels, accounting and estimated ones. One little divergence between the three databases relates to how variables issued from accounting sources are named. Those accounting category's variables are prefixed whether by "ACC_", "CES_" or nothing, in accordance with the database considered.

The table below offers a brief description of some relevant variables contained in the accepted business technical database.

VARIABLE NAME	DESCRIPTION
GENERAL VARIABLES	
PERIMETER	Legal entity within AGRe accepting the risk (mainly AXA Global Re P&C - AGPC or AXA Global Re Life - AGL))
TRY_NUM	Risk acceptance reinsurance treaty's commercial reference number / Risk acceptance contract number
TRY_NAME	Treaty's name
TRY_TYPE	Treaty's type (Local, Financial, QS Germany, QS Mexico,...)
SEC_NUM	Risk acceptance section's number (one treaty can be composed of several sections)
SEC_ID	Risk acceptance section's unique ID (treaty's sections are given a unique ID among all reinsurance treaties in AGRe portfolio)
SEC_CURRENCY	Section's currency
SEC_COVER_FORM	Reinsurance cover or treaty type (QP: quota-share, XP: surplus, XS: excess of loss, SL: stop loss
SEC_EFFECT_DATE	Section's date of effect
SEC_END_DATE	Section's end date
CEDENT (or REINSURER)_BUID	Cedent (or Reinsurer) entity's business ID
CEDENT (or REINSURER)_NAME	Cedent (or Reinsurer) entity's name
CEDENT (or REINSURER)_COUNTRY_ ...	Cedent (or Reinsurer) entity's country (name, ID, ISO 2 code)
POOL_NAME	Pool's name if the treaty is linked to a pool
NONPOOLED_POOLED	Identify whether the treaty is linked to a pool or not
SOLVENCY_II_BULINE	Treaty's Solvency II buisiness line, i.e. technical provisions segmentation defined by S2 directives
LR SCOPE	Loss Ratio scope of the treaty x section. Reserving segmentation on the basis of wich IBNR reserves are estimated.
ULAE_CR_IBNR_TRY_SEC	Unallocated Loss Adjustment Expenses Reserves at the level treaty x section
ACCOUNTING VARIABLES	**Amounts expressed in Euro at the rates of current accounting closure**
ACC_PRIMES_EMISES	Flow of written premiums recorded during the current accounting period
ACC_PORTEFEUILLE_PRIMES	Flow of premiums from portfolio entries (in case of clean cut for example) recorded during the current accounting period
ACC_SINISTRES_PAYES	Flow of paid claims recorded during the current accounting period
ACC_PROV_PNA_CLO	Unearned premium reserves stock recorded at the closing of current accounting period
ACC_PROV_SAP_CLO	File-to-file (or case) reserves stock recorded at the closing of current accounting period (information transmitted by ceding companies)
ACC_PROV_IBNR_CLO	IBNR reserves stock recorded at the closing of current accounting period (information transmitted by ceding companies)
ACC_PROV_MATH_CLO	Mathematical reserves stock recorded at the closing of current accounting period
ACC_COMM_REASS	Flow of reinsurance commissions recorded during the current accounting period
ACC_COMM_NOVATION	Flow of novation commissions recorded during the current accounting period
ACC_COMM_FRONTING	Flow of fronting commissions recorded during the current accounting period
ACC_COMM_INTERMEDIAIRE	Flow of intermediary commissions recorded during the current accounting period
ACC_COMM_COURT	Flow of brokerage commissions recorded during the current accounting period
ACC_COURT_RECU	Flow of received brokerage commissions recorded during the current accounting period
ACC_COTISATION_REC	Flow of reinstatement premiums recorded during the current accounting period
ESTIMATED VARIABLES	
PRIMES_A_EMETTRE	Estimate at current accounting closure of premiums to be received
PRIMES_REC_A_EMETTRE	Estimate at current accounting closure of reinstatement premiums to receive
PRIMES_ACQUISES_PAE	Contractual earned premiums recalculated from the previously calculated premiums to be received
PRIMES_NON_ACQUISES_PAE	Unearned contractual premiums recalculated from the previously calculated premiums to be received
PROV_DD	Additional case estimate provisions (in complement to the accounted case reserves already recorded at the balance sheet date) corresponding to the difference between the case reserves estimate from loss advices and that from the accounts. It allows to take into account last claims information available from the technical loss database.
PROV_IBNR_BE	Additional IBNR Best Estimate provisions (in complement to the accounted IBNR reserves already recorded at the balance sheet date).

Table 9: Accounting and contractual main variables

Let us make the following observations (valid for the three databases):

One reinsurance treaty can be divided into several sections. A given section of a given reinsurance treaty covers a clearly defined perimeter of underlying risks, and is related to a specified currency of reinsurance, a determined set of guarantees, an explicit reinsurance configuration and specific contract's parameters (retention limit, capacity, layers bounds,...).

The data granularity corresponds to a unique combination of one AGRe's third party (cedent entity or external reinsurer) and one reinsurance contract's section (treaty x section). In other words, quantitative values and qualitative information at those databases' finest grain level are associated with a given third party entity along with a given reinsurance contract's section.

Technically speaking, the granularity is either defined by CEDENT_BUID x TRY_NUM x SEC_NUM or CEDENT_BUID x SEC_ID.

Each variable linked to provisions has two other corresponding variables, suffixed by either "_OUV" (provision stock at current accounting period opening valued with rates prevailing at previous year-end closure) or "_OUV_DER" (provision stock at current accounting period opening valued with rates prevailing at current account closing). For example, "PROV_DD_OUV" stands for the account opening's additional case reserves evaluated in euro on the basis of the exchange rate at the time of previous full year account closing, whereas "PROV_DD_OUV_DER" provides the same quantities but evaluated on the basis of the exchange rate at the time of current account closing.

Similarly, all accounting variables that are not provisions (i.e. related to flows of premiums, commissions, claims payments,...) have corresponding cumulative variables with either suffix "_ENR" (cumulative values of accounting years previous to current exercise translated in euro through the exchange rates at the time of past accounting exercises), or suffix "_ENR_DER" (same quantities but valued in euro through the rate prevailing at current closing).

From the variables exhibited in the table 9 we are able to retrieve quantities of interest such as total earned premiums, cumulative paid claims (ACC_SINISTRES_PAYES + ACC_SINISTRES_PAYES_ENR_DER), IBNR reserves (ACC_PROV_IBNR_CLO + PROV_IBNR_BE) as well as case reserves (ACC_PROV_SAP_CLO + PROV_DD) for a given cedent and a given reinsurance treaty x section, or at different other levels of aggregation.

The considered databases are comprised of many other variables we chose not to display for having less relevance to our studied topic and soon to be introduced implemented models.

We will avail ourselves of those databases at hand to enrich the historical registered balances observations, adding some technical and contractual variables through merge on the basis of both the third party business ID and the contract's section ID.

5.1.3 Descriptive analysis and limits of currently applied procedure

Before building models on those data, it is essential to understand what those data can tell us without resorting to intricate methodologies. Indeed, a great deal of information can be retrieved via some minor and direct manipulations of the available historical financial balances.

We will see that this basic investigation alone will invalidate the procedure currently implemented by the treasury team to foresee near future cash flows. This descriptive and visual analysis will furthermore lead us on the way to devise a relevant model to forecast next-12-month monthly cash flows in currencies originating from already constituted but not yet settled balances. It will equally manifest clearly the limits of such a model to address the broader issue of predicting all near future cash flows coming either from already existing balances or not.

Currently applied procedure ("dummy model")

What has been done until now at AGRe, with the purpose of forecasting currency cash requirement within the year, was to project all opened balances amounts (recorded and not yet settled at the time of projection) homogeneously over the next twelve months. That is, aggregate those amounts by currency and then divide them by 12 to obtain the supposed monthly cash flows in currencies in the year to come.

To fix ideas, let us consider that at the date d the sum of all not yet settled balances' signed amounts in USD is, say, minus one million. The predicted cash flows for a year period beginning at d, following the current procedure, will then be minus 83 thousands USD monthly for the next 12 months.

This method, which consists in assuming that the net balances by currency at date d will generate constant monthly cash flows over the next 12 months, is based on two strong and not so realistic hypothesis:

(H1) All opened balances at d will be fully settled over the next 12 months.

(H2) All cash flows occurring over the 12 months following the date of projection d, are solely issued from unsettled net balances existing at date d. In other words, possible settlements from balances recorded after d but settled before $d+12$ months are not taken into account.

Significant differences have been observed:

- between actual and projected cash flows timing;
- between the amounts actually collected / disbursed after one year and those projected according to current basic methodology.

For its simplicity and lack of fidelity to AGRe's real world cash flows behavior, we will call this projection procedure the "dummy" model. This dummy model, used so far for lack of a better one, will serve as a reference point to compare and appraise any new implemented model.

Historical financial balances' perspectives and settlement's velocity rate statistics

First, we designate as a "balance's lifetime" the time interval (in days or months) between the balance's financial writing (or registering) date and its settlement date (when the transaction actually takes place). It corresponds to the time before a recorded financial balance gives rise to an incoming or outgoing payment.

If we place ourselves at an arbitrary passed or present date d (as of date d), and observe historical financial balances data with respect to that considered date d, three possible configurations arise for a given balance. As illustrated in Figure 36, with reference to d, we will either find already settled balances (that "lived" in the past in regard to d), or ones that has been written before d and are not yet settled (as of d), or ones that has been written at d exactly. A fourth option would of course concern not yet existing (registered) financial balances at d.

We will make use afterwards of the acronym "**EBNYS**" to describe the Existing But Not Yet Settled balances at a time d.

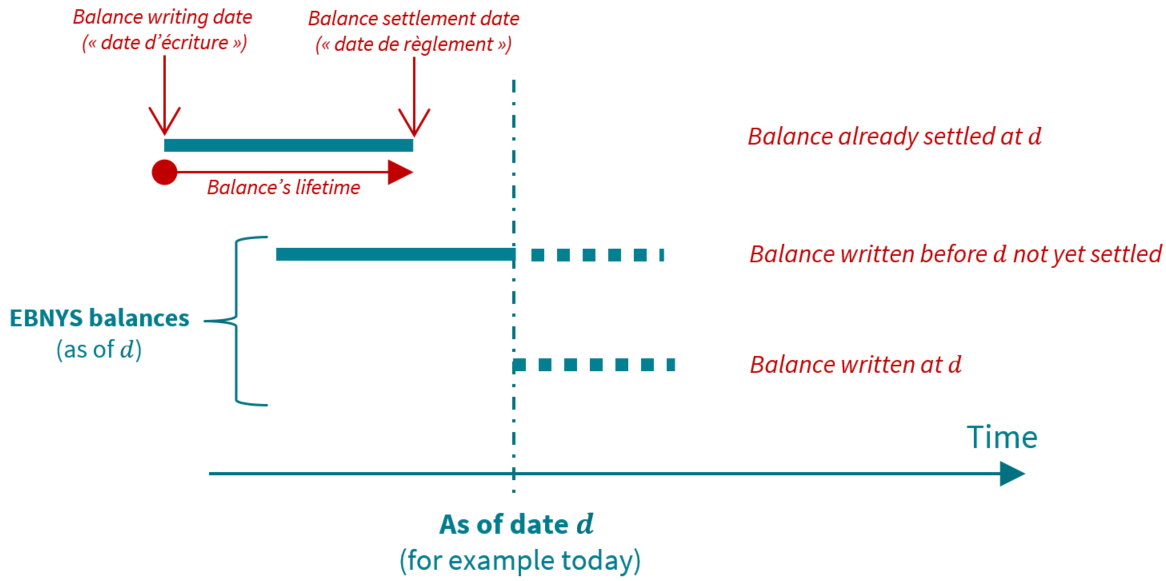


Figure 36: Financial balances' lifetime configurations with respect to a date d

In order to probe the celerity aspect of balances' settlements, several perspectives can be considered for a given date d of reference (as if d were the present time):

- First (Perspective 1), one can take into account in that study all existing balances at d not yet settled. Those balances were opened before d and have each experienced a specific time up to d. Each of them will go on for another proper lapse of time before being settled. From this analysis viewpoint, those remaining (or survival) times are the contemplated observations on which descriptive statistics are based.
- Second (Perspective 2), one can analyses settlements cadences from the same point of view, except without considering "degenerate" balances that have long since been written (as of d) but that have not yet engendered any cash transaction (and probably never will) up to the database extracting date (present time not equal to d). The same observations are considered (i.e. remaining time before settlement from d), but extreme ones are filtered out.
- Third (Perspective 3), one can look at all the balances constituted at d and examine the time periods from that common starting date until the distinct settlements event times for each of those balances. Since restraining the written date, i.e. the shared starting point, to a single date would bring about a too small subset of balances for reasonable statistics to be built on, the choice have been made to consider instead all balances beginning at a given quarter.

Figures 37 represents the 3 different analysis' frames of reference, displayed in the same order as previously described. Obviously, those 3 perspectives can be applied to any subset of financial balances, defined by combinations of categorical variables' classes for instance.

Figure 37: Financial balances' lifetime analysis from different perspectives

Independently of the chosen perspective, several dimensions can be considered to get a grip on settlements rates:

- one can look into the number of settled balances, relatively to the overall number of initial balances considered, as a function of time. That is, for a date of reference d, study the ratios $\left(\frac{\#\text{balances settled until d+t}}{\#\text{open balances at d}}\right)_t$;
- or one can base its survey on the cumulative absolute amounts of settled balances as a function of time with respect to the overall absolute amount of initial balances considered. That is, observe the as-of-date d ratios $\left(\frac{\sum |\text{amounts (in euro) of EBNYS balances at d settled until d+t}|}{\sum |\text{amounts (in euro) of EBNYS balances at d}|}\right)_t$.

The second option have been selected to analyse historical financial balances' settlements pace, since we are more interested in inflows or outflows of currency amounts rather than projection of settlements numbers.

To be able to put up average and standard deviation statistics on settlements rates, or even pseudo weighed empirical cumulative distribution function (more properly named "settled cumulative amount proportion as a time function"), we need to place ourselves at many different times of reference and see how (not yet settled) balances at each of those times evolve. That way, considering various as-of-dates d, we get a series of historical time slices (or samples) that will help to fully harness part of the information contained in the financial historical database and linked to settlements rhythms. The following figure displays that time sampling procedure for two distinct analysis perspectives.

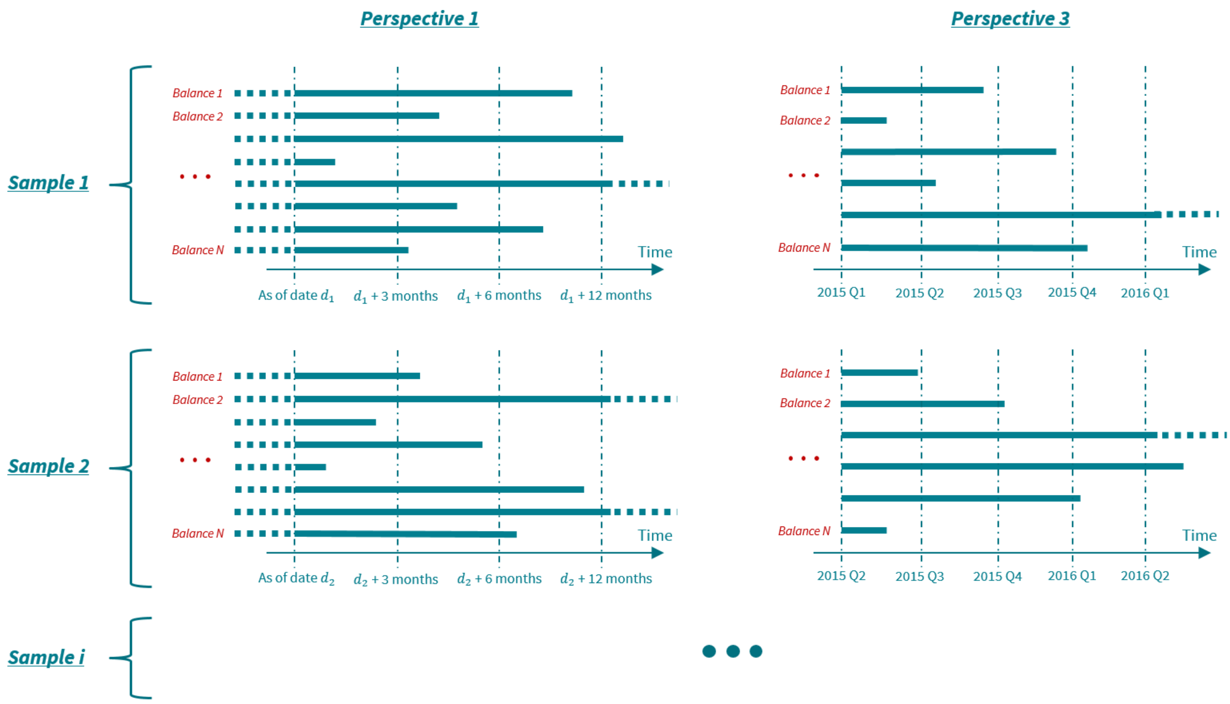


Figure 38: Financial balances' lifetime analysis sampling

The above procedure has been implemented for monthly as-of-dates from January 1st 2013 to January 1st 2020, for the perspectives 1 and 2. It has also been carried out according to perspective 3 on the same time range, but for quarterly starting balances (i.e. first sample encompassing all balances written in the first quarter of 2013, and last sample comprised of balances registered in the first quarter of 2020).

Part of the results are presented in the tables below, which shows the averages and standard deviations of settled financial balance amount proportions 3, 6 and 12 months after all the reference dates d chosen between 2013 and 2020 (for perspectives 1, 2 and 3 respectively). The results are split on the basis of whether the balances stem from a risk's acceptation contract or a risk's retrocession one.

Accept / Retro	**Balance amount proportion settled to 3 months**		**Balance amount proportion settled to 6 months**		**Balance amount proportion settled to 12 months**	
	average	standard deviation	average	standard deviation	average	standard deviation
Accept	18.7%	9.0%	25.4%	9.2%	30.7%	9.7%
Retro	47.7%	18.3%	64.5%	15.1%	78.6%	8.6%

Table 10: Average (over as-of-date months) cumulative amount proportion settled as a function of time (with NA)

Accept / Retro	**Balance amount proportion settled to 3 months**		**Balance amount proportion settled to 6 months**		**Balance amount proportion settled to 12 months**	
	average	standard deviation	average	standard deviation	average	standard deviation
Accept	51.1%	18.7%	69.3%	14.8%	84.1%	9.8%
Retro	51.1%	18.8%	69.2%	15.0%	84.6%	8.3%

Table 11: Average (over as-of-date months) cumulative amount proportion settled as a function of time (without NA)

Accept / Retro	Balance amount proportion settled to 3 months		Balance amount proportion settled to 6 months		Balance amount proportion settled to 12 months	
	average	standard deviation	average	standard deviation	average	standard deviation
Accept	69.2%	18.8%	84.8%	11.0%	94.5%	5.4%
Retro	58.6%	20.5%	81.4%	14.9%	92.7%	5.9%

Table 12: Average (over attaching quarters) cumulative amount proportion settled as a function of time

It is interesting to notice that the average proportions of amount settled vary significantly from one point of view to the other.

We can read on Table 10, corresponding to the first manner to analyse balances' lifetime (perspective 1), that on average, 12 months after a date of reference, approximately 31% of acceptance balances and 79% of retrocession balances are settled (in terms of amount). To rephrase it, of the total acceptance (retrocession) side and not yet settled balances' amount at a date d, on average 31% (79%) will transform into actual cash transactions within 12 months.

Since the perspective currently adopted by AGRe treasurers to anticipate near future cash flows is the first one, we can clearly refute hypothesis (H1) "All opened balances at d will be fully settled over the next 12 months" under perspective 1. To put it another way, the assertion that all existing and not yet settled financial balances at a date d, will be settled within one year after d, is not valid in the light of Table 10 historical statistics.

Table 10 calls attention to the fact that an important part of existing and not yet settled balances at an arbitrary date d, are what we allow ourselves to call "inert balances". Those dormant lines may have been recorded many years before d and may never give origin to real cash flows after d (at least they have not till this survey present time). This is the reason why, to remove that bias, we rule out those motionless balances and consider an alternative perspective. Looking at financial balances under perspective 2, we observe the average proportions exhibited in Table 11. Even taking out those problematic balances, we remark that about 84% of EBNYS balances' total amount at a time d is settled before 12 months.

Now let's have a look at Table 12, and see what we can learn. On average, 69% (resp. 59%) (of the total aggregate amount) of acceptance (resp. retrocession) balances are settled within the 3 months following their entry date. Thus, a considerable part (in terms of absolute amounts) of balances written after a reference date d will be settled before $d + 12$ months, which invalidates (H2) (i.e. "All cash flows occurring over the 12 months following the date of projection d, are solely issued from unsettled net balances existing at time d"). To express this in a different way, taking as an analysis reference the writing period of balances (perspective 3), we note that a consequential part of next-12-month cash flows proceed from yet non-existing balances at the projection time d.

Historical financial balances' settlement cadence

We assume here the perspective 3, and construct development triangles with the following structure:

- aggregation level: to play with, and defined on the basis of any (or combination of) existing or created categorical variables' classes (ex: Acceptance vs Retrocession balances, balance's currency, balance's amount class in euro,...);
- attaching base period: quarter corresponding to balance's writing date (ex: 2020-Q1);

- development period: number of months between the balance's writing date and its settlement date;
- value: cumulative amount settled proportion.

For the two specific aggregation levels, that are the balances' acceptance and retrocession types, we then obtain a development at maximum over 12 months, for each considered starting quarter (here from 2013-Q1 to 2021-Q1) as displayed in the two following figures.

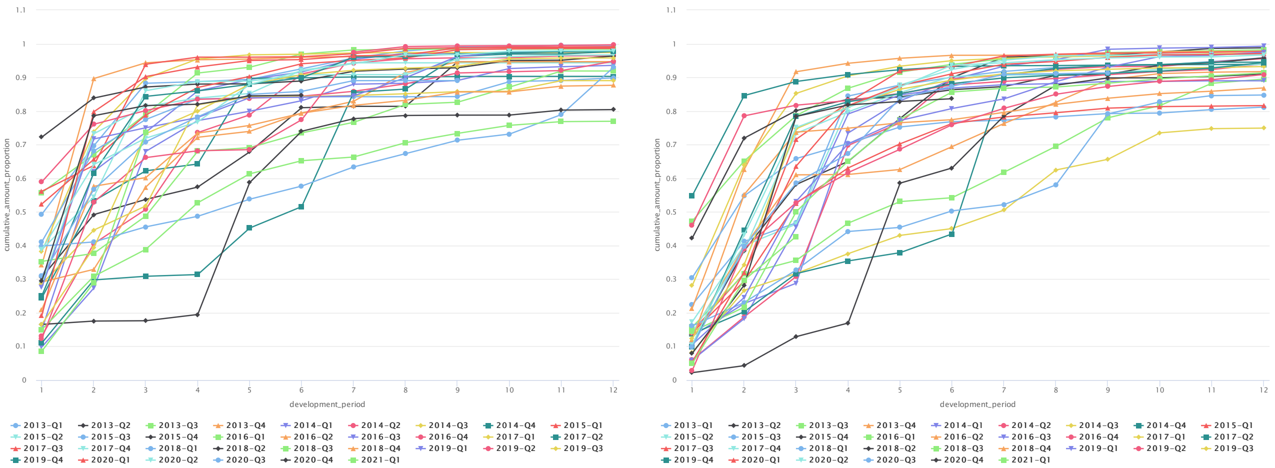


Figure 39: Cumulative settled amount proportion as a function of time (month) for each writing quarters - Acceptance linked balances (left) and Retrocession ones (right)

From those individual development curves we are then able to compute mean curves, Chain-Ladder development pattern curves, and confidence intervals. The two figures below summarize this way the two previous figures' set of individual development curves.

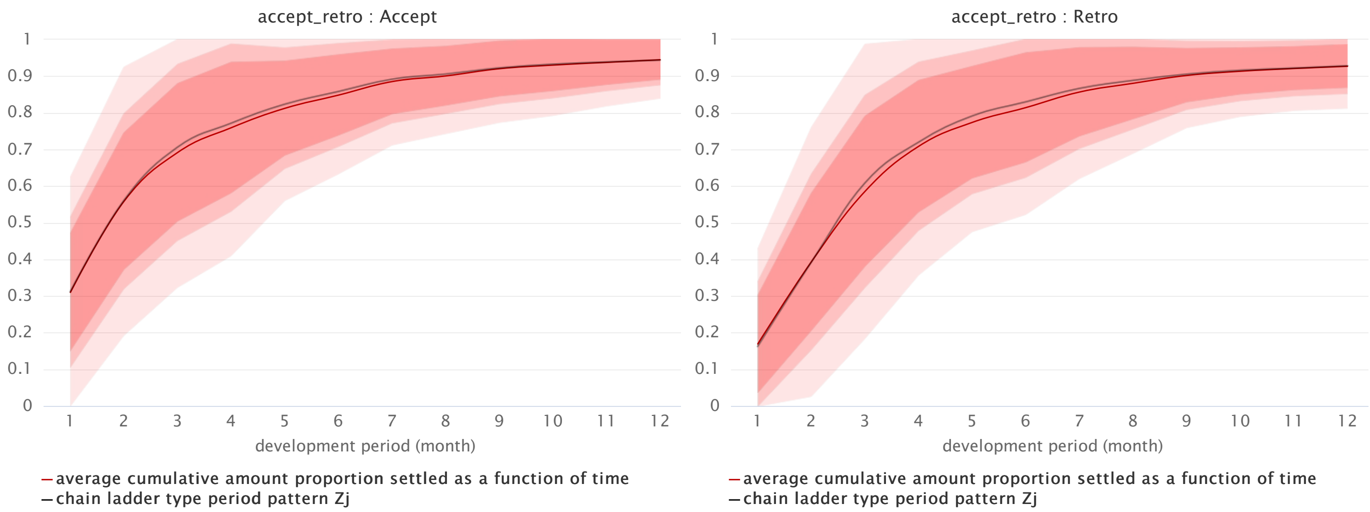


Figure 40: Average (over attaching quarter) cumulative settled amount proportion as a function of time (month) with confidence intervals (68%, 80% and 95% CI) - Acceptance linked balances (left) and Retrocession ones (right)

We can retrieve straightforwardly the average proportions informed in Table 12 from a single look at the average curves shown in red.

Other variables and characteristics can be considered to define separate groups of registered balances and confront their settlements cadence.

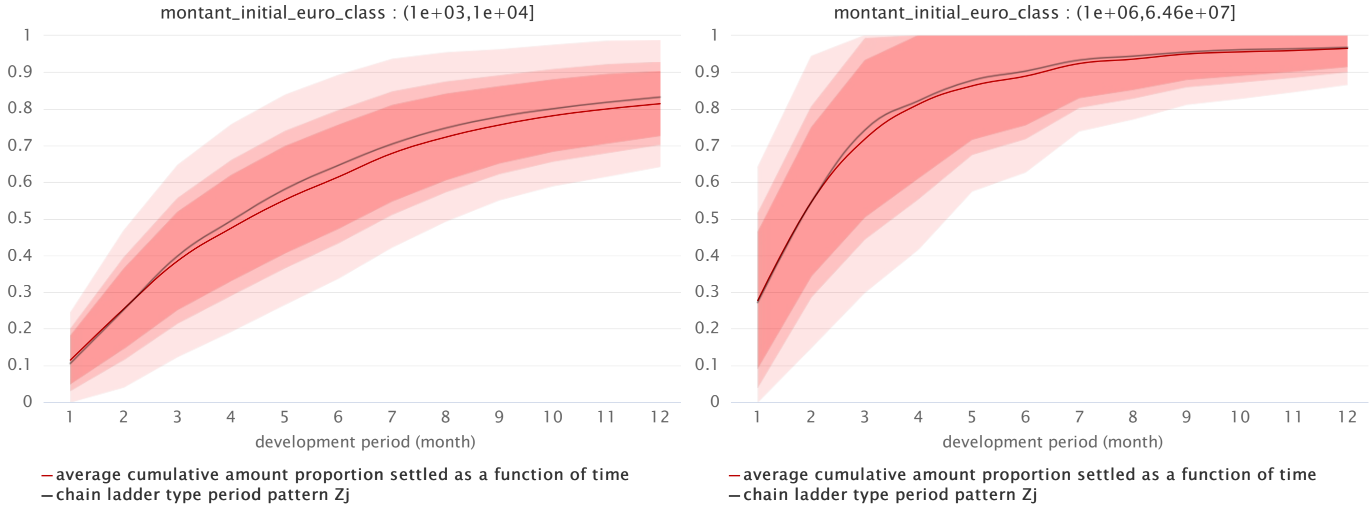


Figure 41: Average (over attaching quarter) cumulative settled amount proportion as a function of time (month) with confidence intervals (68%, 80% and 95% CI) - from 1k to 10k euros absolute amount balances (left) and more than 1M euros absolute amount balances (right)

From Figure 41 we can clearly notice that high amounts tend to be settled more rapidly than small ones. The observation of other amount (in euro) classes' curves confirm the fact that higher the balance's amount in equivalent euros, the sooner after the writing date are the transactions completed, i.e. the higher the speed of settlement. This certainly comes from the fact that since balances with high amounts have more impact on AGRe business and weigh more on financial indicators, accounting teams must dedicate more time and effort to recover (or pay) those substantial amounts.

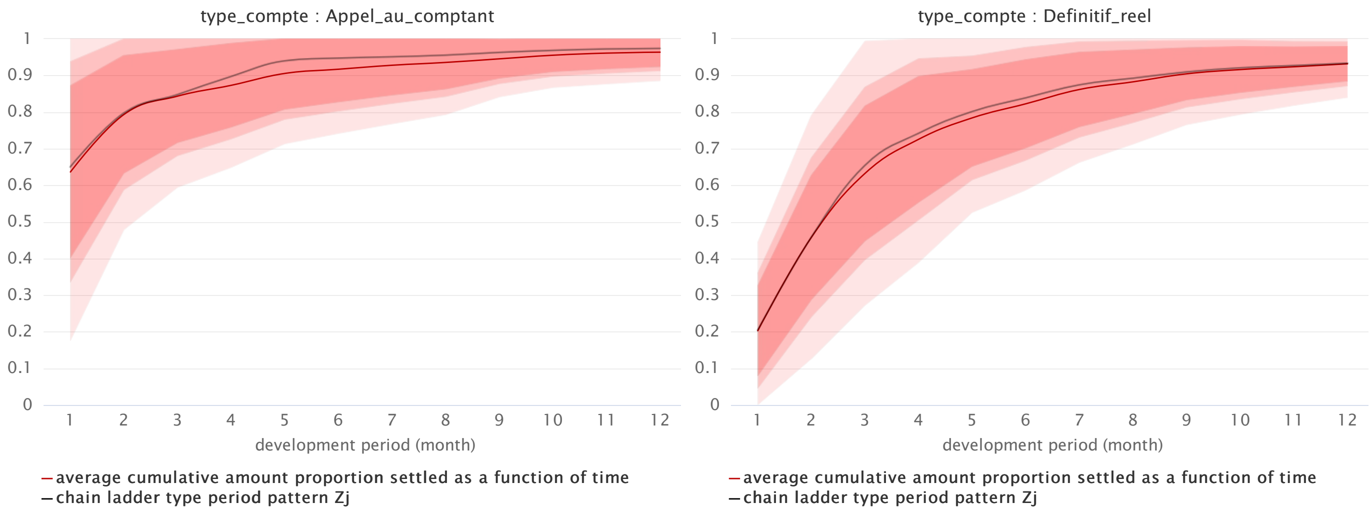


Figure 42: Average (over attaching quarter) cumulative settled amount proportion as a function of time (month) with confidence intervals (68%, 80% and 95% CI) - Cash calls type balances (left) and "Definitive / real" ones (right)

As illustrated in Figure 42 and as expected, cash calls settlements seem to develop much more quickly (in average) than other ones. Cash call type balances materialize almost instantaneously after their entry into effective cash transactions.

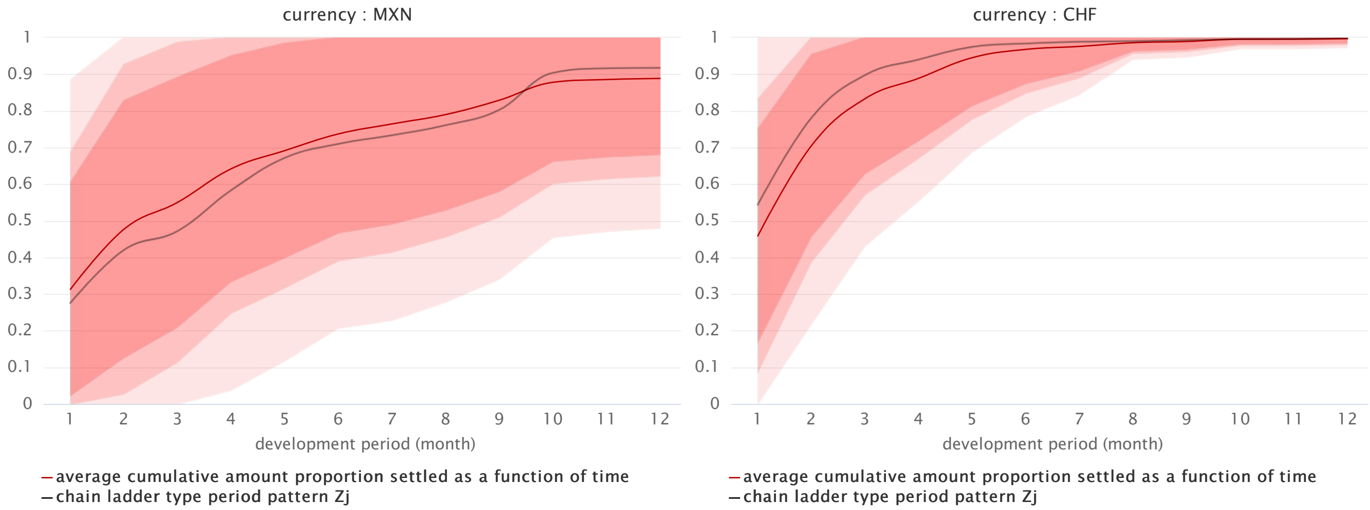


Figure 43: Average (over attaching quarter) cumulative settled amount proportion as a function of time (month) with confidence intervals (68%, 80% and 95% CI) - Mexican pesos (MXN) balances (left) and Swiss Franc (CHF) ones (right)

The two average curves in above figures reveal a considerable difference in settlement progress and volatility between balances expressed in Mexican pesos and the ones expressed in Swiss Franc. CHF balances seem to give rise to settlement cash flows way more swiftly than MXN ones. Furthermore, CHF balances settlement cadences appear more stable and steady than seemingly more volatile MXN ones.

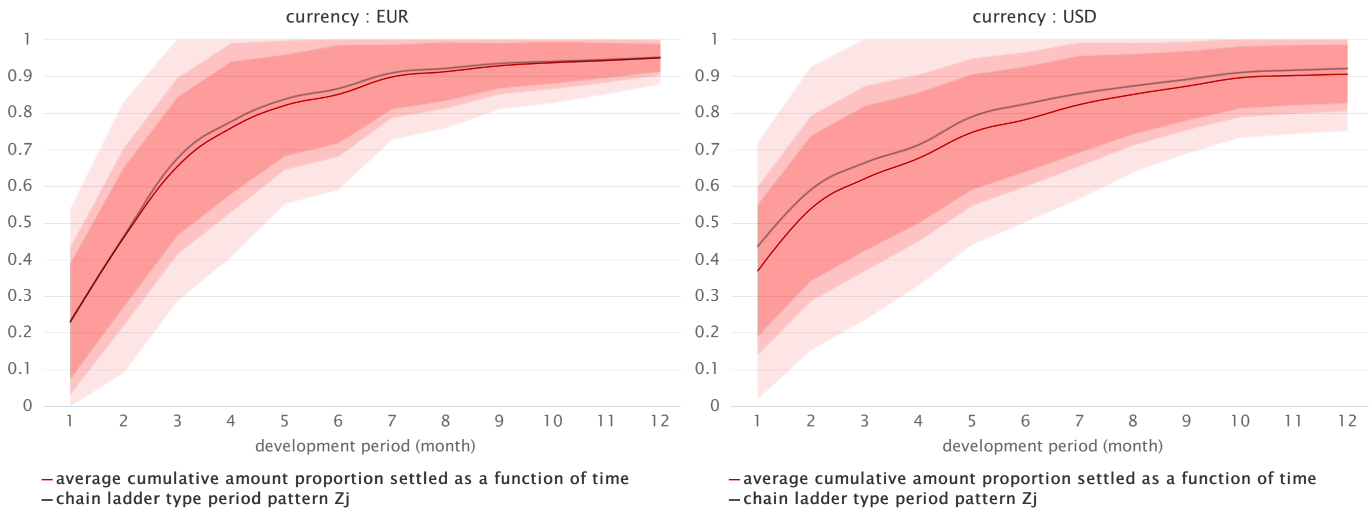


Figure 44: Average (over attaching quarter) cumulative settled amount proportion as a function of time (month) with confidence intervals (68%, 80% and 95% CI) - Euro (EUR) balances (left) and US dollar (USD) ones (right)

In Figure 44 we discern two distinct behaviors: USD balances show (with respect to EUR balances) a higher average proportion of settled amount in the first two months after balance writing but a lower average velocity rate of settlement from 1 to 12 months.

Historical financial balances' settlement time interval histograms

We stressed previously that we would only consider settlement velocity's indicators based on amounts (in equivalent euro) settled rather than on number of balances settled. We allow to break that rule momentarily in order to introduce the following interesting histograms, which represent the frequency of balances classified into categories of settlement time intervals. 6 classes have been defined: [0,1], (1,3], (3,6], (6,12], (12,24] and (+2years].

- $[0, 1]$ class encompasses all balances leading to a cash settlement before 2 months exactly after their entry date.
- $(1, 3]$ class encompasses all balances leading to a cash settlement after 2 months and before 4 months exactly after their entry date.
- And so on.

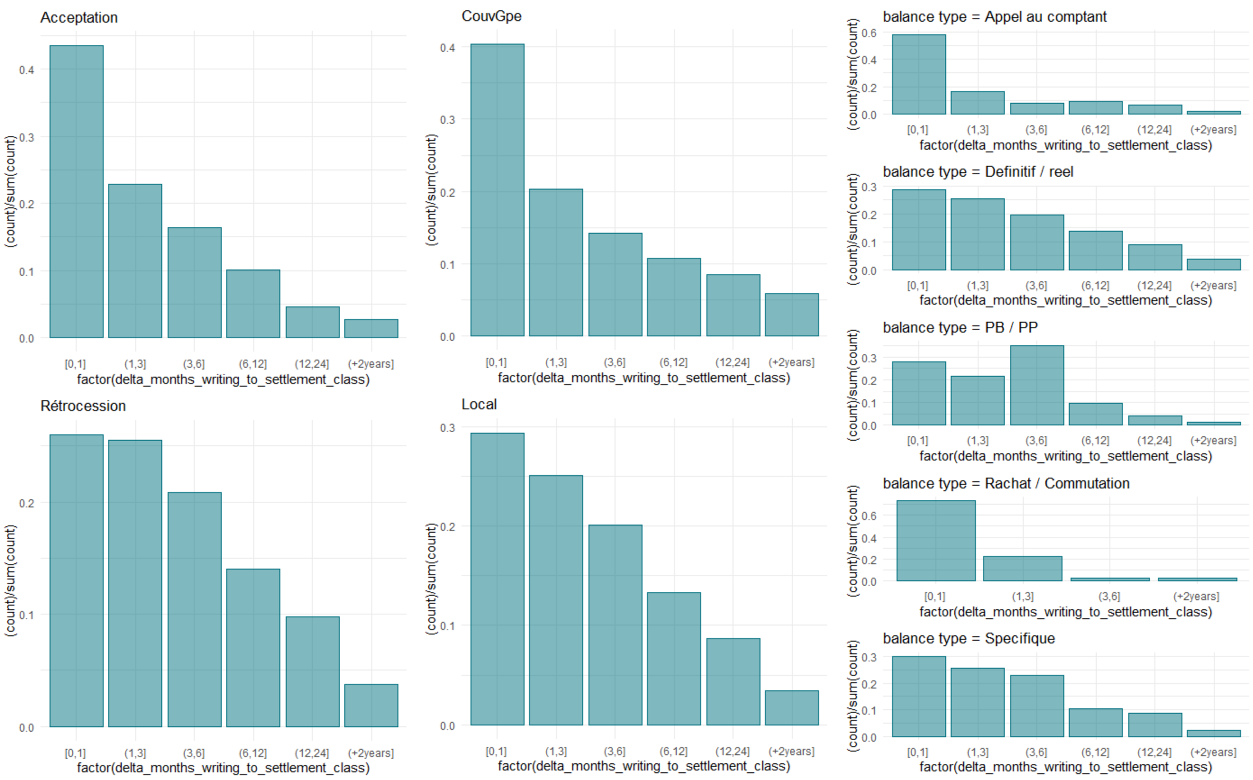


Figure 45: Frequency histogram for different classes of some categorical variables – horizontal axis: categories of time interval (in months) between balance writing and settlement – vertical axis: frequency observed historically for already settled financial balances

Figure 46: Frequency histogram for different classes of some categorical variables – horizontal axis: categories of time interval (in months) between balance writing and settlement – vertical axis: frequency observed historically for already settled financial balances

Once again, different categorical variables have been selected to picture dissimilarities of settlement cadence along the several classes composing those variables. We won't describe at length each and every variations and apparent sources of differentiation between the above displayed settlement rhythms. Whether on the basis of cumulative settled amount proportion curves or from above frequency histograms, we can clearly assert that there exist evident divergences between certain currencies, between amounts levels, reinsurer scores (for retrocession balances) or even entry types (cash calls, ...).

Let us now try to avail ourselves, through more formal and general methodologies, of the variables at disposal (or easily created) whose classes' values influence strongly the pace of corresponding balances settlement.

5.2 Framework and search for an appropriate model

5.2.1 Response and possible explanatory variables, Sketch of a procedure

Response variable

One could presume that the definition of the target variable in this context is straightforward and assume it obviously to be the timelife of a given balance. That is, the time interval (in a given chosen unit: days, months, years, ...) between the balance's written date (date of entry or registration in AGRe financial and accounting systems) and its payment date (whether from AGRe to a counterparty or the other way around). This is not the case as there is a catch that has been a central consideration for the choice and the implementation of a proper model.

As we cautiously pointed up, 3 different perspectives can be considered when evaluating balances time dimension. When trying, at a specific date d, to work out how EBNYS (at d) balances will develop into actual transactions over next months, one is thus more interested into predicting the remaining lifetimes of those balances rather than their whole lifetimes. Nonetheless, one cannot totally discard the information at disposal which is the time from the EBNYS balance inception. In other words, the time the balance has already lived till d (from writing to projection date).

We will then adopt perspective 1 (as is done for currently applied dummy model) and let a model handle freely extremely old or tricky balances. It follows that the response variable should be the remaining lifetime of EBNYS financial balances with respect to a reference date of projection d.

We decided to measure the length of time of this target variable in months. It represents a natural choice of unit since the stated goal is to quantify future monthly incoming or outcoming cash flows.

We will see that this response variable can take either the form of an integer number (of months from d till settlement) or of an interval class not necessarily of the same length, gathering several possible months.

Possible explanatory variables

In the search of a model enabling us to predict near future cash flows, we first strived to single out relevant variables explaining the variations in the settlement patterns of financial balances. Thanks to a descriptive analysis of historical balances we already identified a set of variables liable to act on the balances' time to settlement.

To formalize a bit, we saw that there seems to be a cause-and-effect relationship or a least a correlation between variables linked to each balance such as its currency, entry type, amount, or acceptance vs retrocession nature, the reinsurer scores or the reinsurance treaty perimeter, and the time interval needed to settle that balance after its creation. As depicted by preceding graphs (Figures 39 to 46) one can expect changes in the lifetime of a balance to happen after changes in one or various of those listed explanatory variables. The same must be true as for balances' remaining lifetime.

Some of those categorical variables are made up of many levels (49 for the currency), most of which are not significantly populated. In such cases, the number of levels can be trimmed down discriminating on the basis of observed balances aggregated absolute amount. For example, we could look at the distribution of aggregate absolute amount per class and keep only those grouping balances whose absolute amounts sum is higher than a certain quantile of that distribution. That is, retain only prevailing levels in terms of total aggregate amount and group minor ones inside a class named "other" for example.

The process of reducing the number of levels within a given categorical variables can be applied in a different manner. We indeed grant ourselves the flexibility to redefine existing classes mainly trough merging of similar levels manifesting analogous settlement pattern. For instance, the reinsurers' notations given by AGRe's accounting teams could be divided into only two classifications, namely "good notation" grouping notations 111, 222, and 333 and "bad notation" grouping remaining notations. One could equally boil the balance entry type categorical variable's levels down to 2 or 3 relevant levels instead of 5. "Specific" balances being in nature much alike "Definitive / real" ones, we could group them together, leave "cash call" ones standing out and, and classify the leftover into a class "other".

Other categorical variables with original or redefined levels may be prone to have a direct or indirect causal relationship with the time to settlement. This is why we also included and tested in our models other potential explanatory variables associated with a given balance, such as the reserving segment, the reinsurance treaty type and proportional or non proportional style, the country, the third party, the principal cedent, or even the amount sign.

Sketch of a procedure

Before selecting a model, let us agree on a "universal" procedure, independent of any selected model, to project cash flows arising from date d EBNYS balances over the following twelves months. Those cash flows will be referred as "run-off" cash flows.

1. Chose a date of reference d (must be anterior to the historical database extracting date minus 12 months) interpreted as the new date of projection.
2. Perform an "as if" transformation of the financial database (containing historical financial balances) in accordance with that date d. In other words, rearrange historical balances data in order to be coherent with a picture as of date d, that is:
   - remove balances written after d;
   - get rid of settlement dates (set to NA) taking place after d for balances written before d. Those balances are now considered EBNYS balances as of d.
3. Distinguish "data as of date not settled" corresponding to EBNYS balances as of d, and "data as of date" corresponding to the whole financial historical data base remoulded as if extracted and seen at d (including EBNYS and already settled balances as of d).
4. Train a model or devise a methodology on the basis of data as of date and apply it to data as of date not settled i.e. EBNYS balances (at d) with the objective to predict when each of those EBNYS balances will produce cash flows in their currency denomination.
5. Aggregate thus predicted cash flows by currency and future period of projection up to 12 months to get a map of near future run-off cash flows in currencies.
6. Compare the chosen model predictions with the dummy model ones and actually observed run-off cash flows within the 12 months after d.

5.2.2 Three attempts to find a suitable model

In addition to the finally retained model, of which we will speak in next subsection, three algorithms implementing three alternative and very different models have been created. Let's have a succinct overview over those 3 candidate models that didn't went through.

Triangle model

This model is based on the settlement development triangles whose structure was introduced in previous subsection 5.1.3 (part "Historical financial balances' settlement cadence"). As a reminder, those triangles show cumulative settled amount proportion as a function of the writing quarter and the development month. We already saw that a triangle is an aggregation of data at a level to be defined. One can for example have a specific triangle for each currency, or for each balance's euro amount class limited to the retrocession scope. According to the variables and classes selected the triangle will change and so will the corresponding development pattern shape.

Omitting the steps already detailed in the "Sketch of a procedure" part, the main steps of the Triangle model algorithm can be described as follow:

1. Select one or two categorical variables and build all the triangles associated to every combinations of levels;
2. Compute corresponding Chain-Ladder link-ratios and development patterns (credibility theory can once again be applied here);
3. Each EBNYS balance (as of date d) is linked to a particular development pattern according to the chosen categorical variables;
4. On the basis of the time already lived by each EBNYS balance, the attributed development patterns are adjusted (cut and rescaled) giving way to an individual development pattern for each balance taking into account both the values of considered categorical variables and the already elapsed time since their writing (till d). Development patterns adjustment (cuting and rescaling): if for example a given EBNYS balance has been written say 9 months before d, the first 8 periods of its assigned development pattern are removed. The remaining development pattern, now beginning at the 9th period, is rescaled so as to evolve from 0 to 1.
5. Each EBNYS balance amount is then multiplied by the increments of its related and adjusted development pattern, thus leading to the materialization of future monthly cash flows.

GLM model

A generalized linear model (GLM) is a flexible generalization of ordinary linear regression. Ordinary linear regression predicts the expected value of a given unknown quantity i.e. a random variable called the response variable (or dependent variable), as a linear combination of a set of observed values (predictors). The GLM generalizes linear regression by allowing for response variables to have arbitrary distributions, and for an arbitrary function of the response variable (the link function) to vary linearly with the predictors (rather than assuming that the response itself must vary linearly).

In a GLM, each outcome Y of the dependent variables is assumed to be generated from a particular distribution in the exponential family, a large class of probability distributions that includes the normal, binomial, Poisson and gamma distributions, among others. The mean, μ, of the distribution depends on the independent variables, $\mathbf{X}$, through:

$$\mathrm{E}(\mathbf{Y} \mid \mathbf{X}) = \boldsymbol{\mu} = g^{-1}(\boldsymbol{\beta}^{\intercal}\mathbf{X})$$

where $E(Y \mid X)$ is the expected value of Y conditional on $\mathbf{X}$, $\boldsymbol{\beta}^{\intercal}\mathbf{X}$ the linear predictor i.e. a linear combination of unknown parameters $\boldsymbol{\beta}$, and g the link function.

In this framework, the variance is typically a function, $\mathbf{V}$, of the mean:

$$\operatorname{Var}(\mathbf{Y} \mid \mathbf{X}) = \mathrm{V}(\boldsymbol{\mu}) = \mathrm{V}\left(g^{-1}(\boldsymbol{\beta}^{\intercal}\mathbf{X})\right)$$

The unknown parameters, $\boldsymbol{\beta}$, are typically estimated with maximum likelihood, maximum quasi-likelihood, or Bayesian techniques.

Keeping in mind the general procedure exposed previously, we would then have under this particular model architecture the following set-up for a given date of reference d:

- response variable $\mathbf{Y}$: *delta_months_as_of_date_to_settlement* i.e. number of months between d and the balance settlement;
- response distribution: Poisson, Negative Binomial, etc;
- predictors $\mathbf{X}$: to be picked out among previously listed potential explanatory variables;
- link function g: to be defined.

In this configuration, the training set would correspond to available settled lines as of date d, and the out of bag prediction set to EBNYS lines as of date. The problem is that the quantity to predict *delta_months_as_of_date_to_settlement* does not exist for the training set. Settled lines as of date being already settled at d, one can not define a remaining time before settlement for those balances.

Instead of the time between as-of-date d and the settlement, one could think of the entire balance lifetime as a response variable, that is take $\mathbf{Y}$ = *delta_months_writing_to_settlement* in place of $\mathbf{Y}$ = *delta_months_as_of_date_to_settlement*. But then how to take into account the already elapsed time of EBNYS observations? Considering the whole lifetime as the target variable, we are indeed able to train a GLM model on already settled balances observed as of d. We are furthermore able to predict that response on the testing set, i.e. to forecast a lifetime for EBNYS balances at d. From those predictions we can then induce predicted remaining lifetimes on the basis of already elapsed months observed for those EBNYS balances. Having said that, a serious issue still pops up: predicted entire lifetimes at d may be shorter than the time intervals already lived by EBNYS balances up to d. How then to reconcile the lifetime predictions and the observed number of months since the EBNYS balances' inceptions? For the same reason we could not use the response variable *delta_months_as_of_date_to_settlement*, we can neither integrate the information of already elapsed time inside the model training since this would-be predictor makes no sense for already settled balances.

There is one last hindrance, and not the least, with the application of a GLM model for the prediction of time interval lengths till balances settlements. For a reference date d, data used for training the model being already settled lines as to d, and data employed to test the model being still opened ones with respect to d, a bias obviously exists in both the training and the testing sets. The training set is not representative of the testing one, since there is a sub-representation of problematic balances in the training set, or equivalently, an over-representation of problematic balances in the testing data set. EBNYS lines incorporate a number of balances that have been written a long time ago as to d and may have difficulty to settle. The closer is d to today's date (extraction date), the higher the proportion of troublesome financial balances in the EBNYS testing data, i.e. the higher is the pollution within balances whose amount is to be projected.

Random Forest model

Random forest is a supervised learning algorithm. The "forest" it builds is an ensemble of decision trees, usually trained with the "bagging" method. The general idea of the bagging method is that a combination of learning models increases the overall accuracy and stability of predictions.

The goal of a Classification and Regression Trees (CART) is to build homogeneous cells of individuals obtained by the intersection of several half-spaces.

If we consider a data set $\mathcal{D}_n = \{(X_i, Y_i)\}_{i=1}^n$, with $X_i \in \mathbb{R}^d$ the features (or explanatory variables) and the label $Y_i \in \mathbb{R}$, or $Y_i \in \{0,1\}$, or a categorical quantity (takes a finite number of values). The predicted class is given by the majority rule in every final cell J:

$$k_J^* = \arg \max_{k \in \{1,\ldots,K\}} \hat{p}_k, \hat{p}_k := \frac{1}{|J|} \sum_{i \in J} \mathbb{1}\left(Y_i = k\right)$$

An impurity measure $Q(J)$ is defined on any subgroup J of individuals:

- misclassification error rate $Q(J) = |J|^{-1} \sum_{i \in J} \mathbb{1}\left(Y_i \neq k_J^*\right) = 1 - \hat{p}_{k_J^*}$;
- Gini index $Q(J) = \sum_{k \neq \ell} \hat{p}_k \hat{p}_\ell = \sum_{k=1}^K \hat{p}_k \left(1 - \hat{p}_k\right)$;
- entropy $Q(J) = -\sum_{k=1}^K \hat{p}_k \ln \hat{p}_k$.

If the tree $\mathcal{T}$ has M leaves $J_1, \ldots, J_M$ with cardinalities $c_i := |J_i|$, then the (penalized) loss function is

$$L(\mathcal{T}) := \sum_{i=1}^M c_i Q\left(J_i\right) + \lambda M$$

λ measures the trade-off between tree size and goodness-of-fit and can be chosen through cross-validation.

The CART procedure is a usual "greedy" algorithm:

1. for every $(j, s) \in \{1, \ldots, p\} \times \mathbb{R}$ split the subsample as $R_{1j} := \{i \mid X_{ij} \leq s\}$ and $R_{2j} := \{i \mid X_{ij} > s\}$; calculate $Q\left(R_{1j}\right)$ and $Q\left(R_{2j}\right)$;
2. find the pair (j^*, s^*) that minimizes $\left|R_{1j}\right| Q\left(R_{1j}\right) + \left|R_{2j}\right| Q\left(R_{2j}\right)$ (the "splitting rule");
3. repeat over all leaves and increase the size of the tree until an arbitrarily given level is reached;
4. reduce the size of the tree by "pruning": find a subtree that minimizes $L(\mathcal{T})$. For example, the weakest link pruning successively collapse the internal node that produces the smallest per-node increase of $\sum_{i=1}^M c_i Q\left(J_i\right)$.

The random forests is a model averaging method: it averages many noisy but approximately unbiased models, and hence reduce the variance ("bagging"). It amouts to train B classification trees, based on diverse sub-samples and subsets of $p' < p$ explanatory variables. The final predictor is the average of all the latter predicted probabilities and/or given by the majority vote among all the trees.

$$\mathbb{P}\left(\hat{Y}_{RF} = k \mid X = x\right) = B^{-1} \sum_{b=1}^B \mathbb{P}\left(\hat{Y}_b = k \mid X = x\right), \forall k$$

$$\text{or } \hat{Y}_{RF}(x) = B^{-1} \sum_{b=1}^{B} \hat{Y}_b(x), \text{ when } p = 2$$

$$\text{or } \hat{Y}_{RF}(x) = \text{the most frequently met class among } \hat{Y}_b(x)$$

A random forest algorithm have been implemented, under which the issue of forecasting time till cash transaction was viewed as a classification problem. In this model framework we indeed transformed the target variable, measured initially in months, into categories encompassing one or several months. The levels of the response variable could be for instance [0,0], [1,1], [2,2], (2,4], (4,6], (6,12], (12,24], (+2years], or [0,3], (3,6], (6,12], (12,24], (+2years].

However, exactly the same complications as in a GLM model attempt emerged. Either we chose the number of months between the as-of-date d and the inward or outward payment date as underlying target variable, and the model was impossible to train, or we resolved to retain the number of months between the entry date and the effective cash flow date, and it appeared incompatibilities between predictions and already observed lapses of time since balance writings, as well as an insurmountable source of problematic balances bias in the testing set (EBNYS balances).

Specificities, drawbacks and limitations

Model	Specificities	Limitations and drawbacks
Triangle	• Can take into account elapsed time since balance writing (i.e. already lived), and handle problematic inert lines; • Does not project a balance amount in one block to a particular unique future period, instead distributes this amount over several future periods; • Intuitive, visual and plain model.	• Procedure based on developement patterns calibrated ex post with all data available at the date of database extraction (i.e. not with as-of-date data): theoretically flawed; • Little flexibility in terms of explanatory variables (limited to two): one could design a method to include more than two features incorporating hierarchical credibility theory, but that would create more complexity and bias as well as lead to arbitrary choices as to the hierarchical structure retained.
GLM	• Can set an a priori distribution for the response variable; • Projects a balance amount in one block to a single predicted future month; • Possibility to apply a Monte Carlo procedure to get a distribution of projected amounts on future months.	• Hard to tune; • Not suited for duration issues (censoring,...).
Random Forest	• Does not assume a distribution for the response variable; • Projects a balance amount in one block to a single predicted future months interval class; The projected amount can then be shared out evenly over the months composing that class; • Possibility to apply a Monte Carlo procedure to get a distribution of projected amounts on future months.	• Not suited for duration issues (censoring,...).

5.3 A brief overview of survival analysis theory in light of current application

From previously reported model implementation attempts and the emphasis put on their limitations in regard to the issue at hand, it will not be a surprise to introduce now duration models. The inability of above algorithms to fully meet the constraints and particularities exposed earlier of a time-lapses-predicting problem is to be expected, and have been answered on the basis of a sound mathematical background by a branch of statistics: duration modelling (or survival analysis).

5.3.1 Generalities

One resorts to duration models when one is interested in the expected duration of time until an event occurs, such as death in biological organisms, failure in mechanical systems, or, as is relevant for our current topic, settlement of a given financial balance.

To get knowledge of the proportion of a population surviving past a certain time, their death (or failing) rate, as well as to what extent particular circumstances or characteristics affect (increase or decrease) their probability of survival, it is first necessary to define the notion of "lifetime". In the case of biological survival for example, death is unambiguous, as it is for AGRe's balances incoming or outcoming payments.

Those concepts of lifetime and events of death, in our survey specific case, were already well-defined in subsections 5.1.3 and 5.2.1. To put it clearly once and for all, and by analogy with biological survival analysis:

- **Population**: "Existing But Not Yet Settled" (EBNYS) as well as already settled financial balances as of a date of reference d.
- **Event (of death)**: balance settlement.
- **Lifetime**: time interval (in months) between the balance entry (or writing) date and its related cash transaction date.
- **Survival time (as of d)**: time interval (in months) between d and the balance's settlement date.

5.3.2 Mathematical formulations linked to a time variable

Now that we have well-defined terms, let us deal with a bit of the mathematics underlying the theory of survival analysis and the modelling of time to event.

Let T be a continuous random variable with cumulative distribution function (lifetime distribution function) $F(t)$ defined on $[0, +\infty[$. Its survival function $S(t) = P(\{T > t\}) = 1 - F(t)$ corresponds to the probability that the time of death happens after some specified time t.

If F is differentiable, the density function of the lifetime distribution (or event density), $f(t) = F'(t) = \frac{d}{dt}F(t)$ represents the rate of death or failure event per unit time. The survival function can then be expressed as follow

$$S(t) = P(T > t) = \int_t^{\infty} f(u)du = 1 - F(t)$$

Similarly, a survival event density function can be defined as

$$s(t) = S'(t) = \frac{d}{dt}S(t) = \frac{d}{dt}\int_t^{\infty} f(u)du = \frac{d}{dt}[1 - F(t)] = -f(t)$$

The hazard rate function λ designates the event rate at time t conditional on survival until time t, i.e the probability of not surviving for an additional infinitesimal instant of time dt knowing $T \geq t$.

$$\lambda(t) = \lim_{dt \to 0^+} \frac{P(T \in [t, t + dt] \mid T \geq t)}{dt} = \lim_{dt \to 0^+} \frac{P(t \leq T < t + dt)}{dt \cdot S(t)} = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)}$$

It then follows that the lifetime distribution function $F(t)$, its density $f(t)$, the survival function $S(t)$, the hazard function $\lambda(t)$, and the cumulative hazard function $\Lambda(t) = \int_0^t \lambda(u)du$, are connected through

$$S(t) = \exp(-\Lambda(t)) = \exp\left(-\int_0^t \lambda(u)du\right) = \frac{f(t)}{\lambda(t)} = 1 - F(t), \forall t > 0$$

We can derive quantities such as the probability of death at or before age $t_0 + t$, given survival until age t_0:

$$P(T \leq t_0 + t \mid T > t_0) = \frac{P(t_0 < T \leq t_0 + t)}{P(T > t_0)} = \frac{F(t_0 + t) - F(t_0)}{S(t_0)}$$

Therefore, the probability density of future lifetime is

$$\frac{d}{dt}\frac{F(t_0 + t) - F(t_0)}{S(t_0)} = \frac{f(t_0 + t)}{S(t_0)}$$

and the expected future lifetime is

$$\frac{1}{S(t_0)}\int_0^\infty tf(t_0 + t)\,dt = \frac{1}{S(t_0)}\int_{t_0}^\infty S(t)dt$$

where the second expression is obtained using integration by parts.

We can also define the hazard rate function in the discrete case, where T is a random variable taking values in a discrete set (that is countable but not necessarily finite) $\mathcal{T} = \{t_1, \ldots, t_n, \ldots\}$. The hazard rate function of T is then $\lambda(t) = \mathbb{P}(T = t \mid T \geq t) = \frac{\mathbb{P}(T=t)}{\mathbb{P}(T \geq t)}$ with the convention $0/0 = 0$ and $\lambda(t) = 0$ if $t \notin \mathcal{T}$. The hazard rate function uniquely determines the distribution of T and $S(t) = \prod_{t_i < t}(1 - \lambda(t_i))$.

5.3.3 Censoring phenomenon

Another important concept at the core of survival analysis is the notion of censoring. Censoring amounts to an issue of missing data. Time to event may indeed not be observed when the observation time length does not include all observed subjects undergoing the event of interest, or when said subjects disappear from observation range prior to experiencing this event.

As far as we are concerned, we face what is called right-censoring. Financial balance historical data are as a matter of fact right-censored, since the observation of EBNYS lines as of a date d provide information only on the time they already spent open (at d) but not on their entire lifetime. To rephrase it, at an arbitrary time of reference d, a proportion of observed historical balance data does not carry direct information on the balances' lifetime: we don't know at d what will be the lifetime of EBNYS balances (as of d). However, for each of those balances, we have access to the lower limit l for the true event time T such that T > l.

An additional phenomenon liable to spring up in survival analysis is truncation. While censoring refers to incomplete observation due to a random cause, truncation effect appears when the incomplete nature of the observation is due to a systematic selection process inherent to the study design. Although phenomenons of left censoring and truncation are common in actuarial data and studies (for life insurance and pensions for instance), they do not concern the current observed data.

As we saw, performing statistical inference on duration variables introduces some specific problems since "usual" variables can be measured instantaneously, whereas it takes time to collect duration ones. While gathering information on a duration variable, many circumstances are liable to pollute available observations (censoring, truncation). Individual balances with incomplete (respectively complete) observation of their lifetime are said to be censored (respectively uncensored). The information about right-censored balances is incomplete, but is not to be discarded. This constitutes indeed a valuable piece of information still usable within duration model frameworks.

Since big values of T are more likely to be censored (i.e. problematic balances are more likely to be among the EBNYS ones, whose cash flows are to be forecast), small values are then overrepresented in the sample of uncensored observations (historical settled balances used to fit a model). Not taking into account this right-censoring would lead to an underestimation of the values taken by T.

5.3.4 Basic duration model mathematical framework

- Let $(T_1, \ldots, T_n)$ be i.i.d. replications of a r.v. T representing the lifetime of financial balances: $T = d_{settlement} - d_{writing}$;
- Let $(C_1, \ldots, C_n)$ be i.i.d. copies of a r.v. C (the censoring variable): $C = d_{reference} - d_{writing}$ for EBNYS observations with respect to $d_{reference}$.

Right-censored observations are made of $(Y_1, \delta_1, \ldots, Y_n, \delta_n)$, where $Y_i = \inf(T_i, C_i)$, and $\delta_i = \mathbb{1}_{T_i \leq C_i}$. For the current application:

- C_i is the age at which balance i quits observation for any other cause than settlement;
- Y_i is the age at which balance i quits observation for any reason (settlement or not).

We therefore have for an arbitrary EBNYS (as of d) line i, $Y_i = C_i$ and $\delta_i = 0$, and for a settled (as of d) one j, $Y_j = T_j$ and $\delta_j = 1$.

We assume that there is no truncation (in our case all balances are supposed to be reported / registered) and that T is independent from C (no way to test H0: "T independent from C" against H1: "T not independent from C", see [7]). If we denote $G(t) = \mathbb{P}(C \leq t)$ and $\tau_\Lambda = \inf\{t : \Lambda(t) = 1\}$ then for any function ϕ such that $E[|\phi(T)|] < \infty$, and $\phi(t) = 0$ for $t \geq \tau_H$, then

$$E[\delta\phi(Y)] = E[(1 - G(T^-))\phi(T)]$$

5.3.5 Kaplan-Meier estimator

In the case where C and T are absolutely continuous (we assume so, even if operationally it is not the case), the Kaplan-Meier estimator of the cumulative distribution function of T, i.e. $F(t)$, is defined as,

$$\hat{F}(t) = 1 - \prod_{i:Y_i \leq t} \left(1 - \frac{\delta_i}{\sum_{j=1}^{n} \mathbf{1}_{Y_i \leq Y_j}}\right)$$

It is a piecewise constant function $\hat{F}(t) = \sum_{i=1}^{n} W_{i,n} \mathbf{1}_{i \leq t}$, where $W_{i,n}$ is the mass attributed to the i -th observation.

If $\left(Y_{(1)}, \ldots Y_{(n)}\right)$ is the vector of ordered observations values, that is $Y_{(1)} \leq Y_{(2)} < \ldots < Y_{(n)}$, and $\delta_{(i)}$ the value of δ associated with $Y_{(i)}$, then in absence of truncation, we have the following analytic expression for the jumps.

$$W_{(i),n} = \frac{\delta_i}{n-i+1} \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)$$

If $\mathbb{P}(T = C) = 0$, and $\hat{G}(t)$ is the Kaplan-Meier estimator of G, one has another expression for the weights in absence of truncation:

$$W_{i,n} = \frac{1}{n} \frac{\delta_i}{1 - \hat{G}\left(Y_i^-\right)}$$

This shows that Kaplan-Meier procedure allocates mass only to observations that are uncensored (the only ones with full information on T). Nevertheless, the censored observations influence the computation of $W_{i,n}$ through $\hat{G}$. In addition, we can observe that the larger Y_i with $\delta_i = 1$, the larger $W_{i,n}$. In other words, Kaplan-Meier estimator compensates the lack of large uncensored observations by attributing more weight to them.

Finally, if ϕ denote a function with $E[|\phi(T)|] < \infty$ and $\phi(t) = 0$ for $t \geq \tau_H$, one can estimate $\theta = E[\phi(T)] = \int \phi(t) dF(t)$:

$$\hat{\theta} = \int \phi(t) d\hat{F}(t) = \sum_{i=1}^{n} W_{i,n} \phi\left(Y_i\right) = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i \phi\left(Y_i\right)}{1 - \hat{G}\left(Y_i^-\right)}.$$

Under mild conditions, from the strong law of large numbers and the fact that $E[\delta\phi(Y)] = E[(1 - G(T^-))\phi(T)]$:

$$\hat{\theta} = E[\hat{\phi(T)}] = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i \phi\left(Y_i\right)}{1 - G\left(Y_i^-\right)} \to E[\phi(T)] \text{ a.s.}$$

5.3.6 Parametric modeling

Parametric modeling means that the distribution of T is assumed to belong to a collection of probability distributions $\{\mathbb{P}_\theta : \theta \in \Theta\}$ where $\Theta \subset \mathbb{R}^k$. The true parameter (we seek to approximate) is denoted θ_0. Restraining ourselves to continuous models in which $\mathbb{P}_\theta$ admits a density f_θ and an associated survival function $S_\theta = 1 - F_\theta$, and assuming once again that T is independent from C, we can apply the two classical methods of estimation: the method of moments and the method of maximum likelihood.

The method of moments requires to have an expression of θ is terms of first moments of T, i.e. $\theta_0 = h\left(E[T], E\left[T^2\right], \ldots, E\left[T^k\right]\right)$. Under censoring and truncation, we can resort to Kaplan-Meier method to estimate $E\left[T^j\right]$ for $j = 1, \ldots, k$, by $\int t^j d\hat{F}(t)$, and define

$$\hat{\theta} = h\left(\int t d\hat{F}(t), \ldots, \int t^k d\hat{F}(t)\right)$$

This method of moments suffers shortcomings since it usually is not efficient (as in the absence of censoring and truncation) and missing observed data in the right-tail of the distribution may deteriorate the estimation of $E\left[T^j\right]$ via Kaplan-Meier.

As for the method of maximum likelihood, the likelihood function for censured observations (replications of (Y, δ)) is more complex that in the usual case. Since, the observations also involve a discrete observed quantity δ, the dominating measure is $\lambda \otimes \left\{\delta_{\{0\}} + \delta_{\{1\}}\right\}$, where $\delta_{\{i\}}$ is the Dirac measure on $\{i\}$ and λ the Lebesgue measure on $\mathbb{R}$. Thus, keeping the same assumptions and notations as previously, the likelihood writes:

$$L_n(\theta) = \left\{\prod_{i=1}^{n} f_\theta\left(Y_i\right)^{\delta_i} S_\theta\left(Y_i\right)^{1-\delta_i}\right\}\left\{\prod_{i=1}^{n} g\left(Y_i\right)^{1-\delta_i}\left(1 - G\left(Y_i\right)\right)^{\delta_i}\right\}.$$

We can clearly see a separation in the likelihood between functions that depend on θ, and the distribution of the censoring variable C (with cdf G). To maximize this likelihood function, it is only required to maximize the first bracket which does not depend on the knowledge of the distribution of C. This approach presents nonetheless several drawbacks. First, the likelihood has usually a complex form from which a closed formulas for $\hat{\theta} = \arg\max L_n(\theta)$ cannot be extracted. Second, additionally to being computationally to intensive for large values of n, the convergence of numerical techniques may be slow and erratic. And last, such an algorithm would require to be initialized by appropriate initial values. Therefore, the method of moments could be combined with the maximum likelihood approach.

5.3.7 Regression

The previously described Kaplan-Meier approach is a univariate analysis that does not investigate and take into account the associations between the survival time of an observed subject and potential predictor variables. A regression framework allows to take convariates into consideration in the description of survival time. Thus, if we wish to include potential explanatory variables within a regression framework type duration model, the presence of an additional vector variable $\mathbf{X} \in \mathbb{R}^p$ has to be considered.

In this regression set up, the observed values are then $(Y_i, \delta_i, \mathbf{X}_i)_{1 \le i \le n}$, assuming once again that there is no truncation. In absence of covariates, T independent of C was assumed. We now must extend this hypothesis in order for the model to be indentifiable. Identifiability is a property which a model must satisfy for precise inference to be possible, that is for the true values of this model's underlying parameters to be theoretically accessible on the basis of an infinite number of observations. It is mathematically equivalent to the fact that different parameters' values must generate different probability distributions of the observable variables. Certain technical restrictions, called the identification conditions, may be required for a model to satisfy identifiability. In this particular case, we will presume either that $(T, \mathbf{X})$ is independent of C, or that T is independent of C conditionally to $\mathbf{X}$ (in which case C can depend on $\mathbf{X}$).

One can perform a regression resorting to fully parametric modelling in assuming that $T \mid \mathbf{X} \sim \mathbb{P}_{\theta(\mathbf{X})}$, where $\theta(\mathbf{X}) = h(\alpha, \mathbf{X}), h$ being a known function (in a parametric framework), and α being an unknown parameter. For example, $T \mid \mathbf{X}$ can be supposed to follow a Weibull distribution with unknown parameter $\theta(\mathbf{X})$ depending on $\mathbf{X}$. h is ordinarily set as a linear function of $\mathbf{X}$ (or a simple transformation of a linear function).

Another strategy to effectuate a regression is to model $E[T \mid \mathbf{X}] = \alpha + \beta^\intercal \mathbf{X}$ (mean regression). One could also wish to carry out quantile regression or other any type of regression, and assuming that the regression function is linear is not compulsory. If consistent estimator $\hat{F}(t, \mathbf{x})$ of $F(t, \mathbf{x}) = \mathbb{P}(T \le t, \mathbf{X} \le \mathbf{x})$ can be computed, since

$$(\alpha, \beta) = \arg\min_{a,b} E\left[(T - a - b^\intercal \mathbf{X})^2\right]$$

then one can estimate these parameters by

$$\left(\hat{\alpha}, \hat{\beta}\right) = \arg\min_{a,b} \int (t - a - b^{\intercal}\mathbf{x})^2 d\hat{F}(t, \mathbf{x})$$

If $(T, \mathbf{X})$ is independent from C, then in particular T is independent from C and $W_{i,n}$ designating the weights of Kaplan-Meier estimator of the univariate distribution function of T, we can use the following estimator

$$\hat{F}(t, \mathbf{x}) = \sum_{i=1}^{n} W_{i,n} \mathbf{1} y_{i \leq t, \mathbf{X}_i \leq \mathbf{x}}$$

5.3.8 Proportional hazards model: Cox regression model

Proportional hazards models are a class of survival models in statistics. As we just explained, survival regression models relate the time T before some event occurs, to one or more covariates $\mathbf{X}$ that may be associated with that quantity of time. In a proportional hazards model, the unique effect of a unit increase (or a level's value change) in a covariate is multiplicative with respect to the hazard rate. For example, as for accounting balances' perspective, being classified as type "cash call" instead of "Definitive / real", may double the hazard rate for its settlement to occur. Inversely, being registered in Mexican pesos instead of Swiss Franc may halve the hazard rate of transformation from EBNYS balance to cash transaction.

Survival models can be viewed as consisting of two parts: the underlying baseline hazard function, often denoted $\lambda_0(t)$, describing how the risk of event per time unit changes over time at baseline levels of covariates; and the effect parameters, describing how the hazard varies in response to explanatory covariates.

The Cox regression (or Cox proportional-hazards) model is a commonly used proportional hazards models, which works for both quantitative and categorical explanatory variables, and is able to assess simultaneously the effect of several risk factors on survival time.
Let $\lambda(t \mid \mathbf{x})$ denote the hazard rate of $T \mid \mathbf{X} = \mathbf{x} = (x_1, x_2, \ldots, x_p)$. The conditional hazard function for the Cox proportional hazards model has the form

$$\lambda(t \mid \mathbf{x}) = \lambda_0(t) \exp\left(\beta^{\intercal}\mathbf{x}\right)$$

where $\beta \in \mathbb{R}^p$ is an unknown parameter vector, and λ_0 is an unknown hazard baseline function (nonparametric part). The coefficients $(\beta_1, \beta_2, \ldots, \beta_p)$ measure the impact (i.e., the effect size) of covariates $(X_1, X_2, \ldots, X_p)$.

The quantities $\left(e^{\beta_j}\right)_{j \in [\![1,p]\!]}$ are called hazard ratios (HR). A value of β_j greater than zero, or equivalently a hazard ratio greater than one, indicates that as the value of the j^{th} covariate increases, the event hazard rate increases and thus the length of probable survival time decreases. Put another way, a hazard ratio above 1 indicates a covariate that is positively correlated with the event probability, and thus negatively associated with the length of survival. To summarize,

- HR = 1: No effect
- HR < 1: Reduction of the hazard rate
- HR > 1: Increase of the hazard rate

If λ_0 is known or of known form, the Cox regression approach becomes fully parametric. This model has the advantage of simplicity regarding the dependence in terms of $\mathbf{X}$, and flexibility due to the presence of λ_0, but it can sometimes turn out to be too rough to describe the dependence in terms of $\mathbf{X}$.

Based on the observations $(Y_i, \delta_i, \mathbf{X}_i)_{1 \leq i \leq n}$, one can estimate the parameter β through pseudo-likelihood maximization, that is maximization of the following function:

$$L(\beta) = \prod_{i:\delta_i=1} L_i(\beta) = \prod_{i:\delta_i=1} \left(\frac{\lambda\left(Y_i \mid \mathbf{X}_i\right)}{\sum_{j:Y_j \geq Y_i} \lambda\left(Y_i \mid \mathbf{X}_j\right)} \right) = \prod_{i:\delta_i=1} \left(\frac{\exp\left(\beta^\intercal \mathbf{X}_i\right)}{\sum_{j:Y_j \geq Y_i} \exp\left(\beta^\intercal \mathbf{X}_j\right)} \right)$$

where $L_i(\beta) = \left(\frac{\lambda(Y_i|\mathbf{X}_i)}{\sum_{j:Y_j \geq Y_i} \lambda(Y_i|\mathbf{X}_j)} \right)$ is the likelihood of the event of interest to occur for subject i at time Y_i.

Λ_0 can then be estimated via the Breslow estimator,

$$\hat{\Lambda}_0(t) = \sum_{i:Y_i \leq t} \left(\frac{\delta_i}{\sum_{j:Y_j \geq Y_i} \exp\left(\hat{\beta}^\intercal \mathbf{X}_j\right)} \right)$$

with $\hat{\beta} = \arg\min_\beta L(\beta)$

What of the case in which $\mathbf{X}$ corresponds to discrete (or categorical) covariates, whose values (levels or classes) represent population categories? In regression modeling, including proportional hazards regression, a useful way of modeling such categorical explanatory variables and their effect on the outcome hazard rate is to resort to dummy variables. For that we replace p categorical covariates $\mathbf{X} = (X_1, X_2, \ldots, X_p)$ by p binary vectors (with at most one component value equal to one) as such:

$$\left(X_j = (\mathbb{1}_{\text{individual is in population category } k})_{1 \leq k \leq n_j - 1} \right)_{j \in [\![1,p]\!]}$$

where n_j is the number of levels within the categorical variable X_j.
To avoid overparametrization, we consider only $n_j - 1$ dummy variables and a reference category (for which dummy variables are all equal to zero) for a given categorical predictor X_j.
We thus obtain the Cox conditional hazard function for categorical explanatory variables

$$\lambda(t \mid \mathbf{X}) = \lambda_0(t) \exp\left(\sum_{j=1}^{p} \beta_j^\intercal X_j \right) \quad \text{where } \beta_j \in \mathbb{R}^{n_j - 1}$$

For $j \in [\![1, p]\!]$, and $k \in [\![1, n_j - 1]\!]$, β_{jk} can be interpreted as the log hazard ratio between an individual in category k and an individual in the reference category with regard to categorical variable X_j, all other covariates being indentical. We can see that from

$$\frac{\lambda(t \mid X_j\text{'s level} = \text{k, other factors being equal})}{\lambda(t \mid X_j\text{'s ref level, other factors being equal})} = \frac{\lambda_0(t) \exp\left(\beta_{jk} + \sum_{l=1, l \neq j}^{p} \beta_l^\intercal X_l \right)}{\lambda_0(t) \exp\left(0 + \sum_{l=1, l \neq j}^{p} \beta_l^\intercal X_l \right)} = \exp\left(\beta_{jk}\right)$$

The hazard ratio between two arbitrary classes of X_j, k and k' in $[\![1, n_j - 1]\!]$ is given by

$$\frac{\lambda(t \mid X_j\text{'s level} = \text{k, other factors being equal})}{\lambda(t \mid X_j\text{'s level} = \text{k', other factors being equal})} = \frac{\lambda_0(t) \exp\left(\beta_{jk} + \sum_{l=1, l \neq j}^{p} \beta_l^\intercal X_l \right)}{\lambda_0(t) \exp\left(\beta_{jk'} + \sum_{l=1, l \neq j}^{p} \beta_l^\intercal X_l \right)} = \exp\left(\beta_{jk} - \beta_{jk'}\right)$$

In the semi-parametric Cox regression, the reference survival curve (hazard baseline function) is based on reference levels of categorical variables and values of 0 for continuous variables.

Other types of survival models that are not proportional exist, such as Accelerated Failure Time models (AFT models). Whereas a proportional hazards model assumes that the effect of an explanatory variable is to multiply the hazard by some constant, an AFT model assumes that the effect of a covariate is to accelerate or decelerate the life course by some constant. Unlike proportional hazards models, in which Cox's semi-parametric proportional hazards model is more widely used than parametric models, AFT models are predominantly fully parametric i.e. a probability distribution is specified.

5.4 Duration models implementation, performance criteria and backtest

5.4.1 Direct application to historical financial balances' data: model calibration

First of all, let us call to mind the fact that only particular balances fall inside the scope of this study: we will work exclusively on balances that will spawn what we called "technical cash flows", i.e. balances netting (or not) claims payments, adjustment premiums and commissions, at the exception of Pools Accounts and specific Quota-Shares (QS Germany and QS Mexico) related ones.

The raw balances' historical data set is accordingly trimmed to concur exactly with this scope.

We will place ourselves at the arbitrary date of reference d = 2020-01-01 to display afterwards numerical examples and figures illustrating the calibration of a duration model to the data at hand. We must nevertheless keep in mind for the moment that any reference date can be chosen (within "reasonable" bounds of the data historical depth). To achieve this time perspective shift, we accomplish the relatively straightforward "as-of-date" transformation described in the "sketch of a procedure" of subsection 5.2.1.

We are now on the 1st of January 2020 and wish to get an idea of the cash needs (demand for cash or fund) in the diverse currencies composing our portfolio, that are liable to pop up over the year. Existing But Not Yet Settled (EBNYS) accounting balances today (expressed in currencies) represent a valuable first accessible source of information to quantify and time near future currencies cash requirements. Hence, to forecast monthly technical cash flows (expressed in currencies) till December 2020, corresponding to the settlements of EBNYS balances on (as of) 2020-01-01.

As of date 2020-01-01 and within previously outlined perimeter, we count 149 576 historical lines, 16 956 of which are EBNYS balances.

For each line i of this as-of-date data set we define Y_i and δ_i as previously introduced in survival analysis mathematical framework.

- case subject i is a settled balance as of 2020-01-01: $Y_i = T_i$ = time lapse (in months) between writing and settlement, and the event indicator is set to 1, i.e. $\delta_i = 1$;
- case subject i is a EBNYS balance (right-censored observation) as of 2020-01-01: $Y_i = C_i$ = time lapse (in months) between writing and as-of-date d = 2020-01-01, and the event indicator is set to 0, i.e. $\delta_i = 0$.

One or two lines of code are then sufficient to fit a duration model on those historical right-censored observations, as of date 2020-01-01. The library "survival" on R has been used. To begin, let us review and interpret the corresponding Kaplan-Meier estimator.

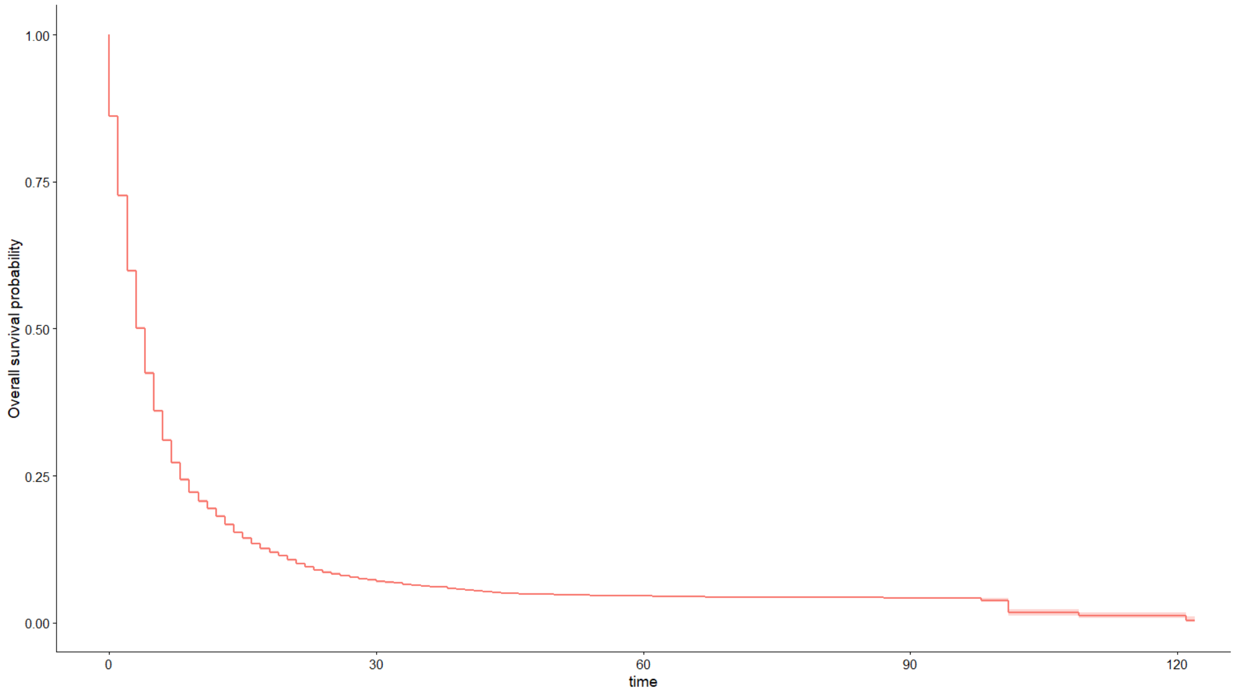


Figure 47: Kaplan-Meier survival function estimator

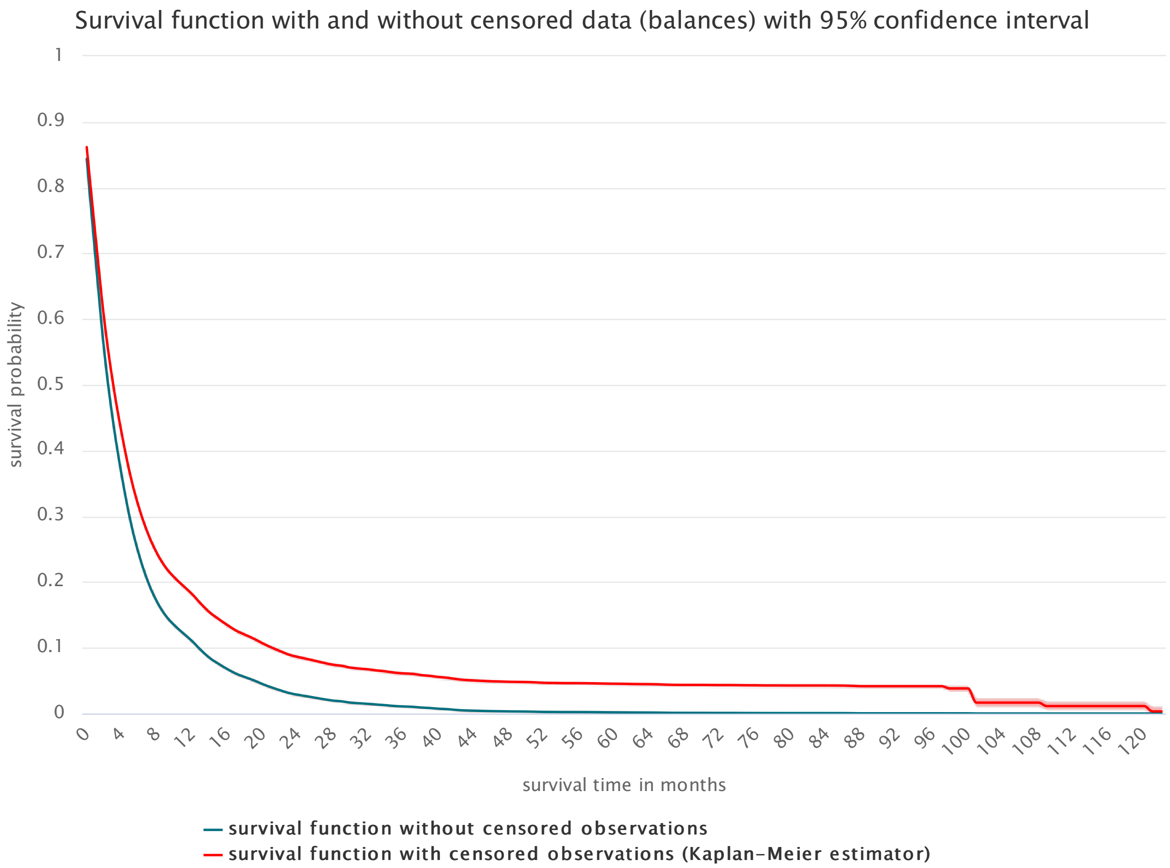


Figure 48: Survival function with and without censored data

The first of the two above figures (Figure 47) show the piecewise constant Kaplan-Meier estimator of the balances survival function as of 2020-01-01, that is $1 - \hat{F}(t)$. The second figure (Figure 48) displays the same estimator, but smoothed, as well as the smoothed Kaplan-Meier estimator of the survival function in the case censored observations (i.e. EBNYS) are removed from the calibration set (only data with $\delta = 1$ are kept). Ignoring censoring clearly leads to an underestimation of the overall survival probability. The two survival curves are constructed on the basis of exactly the same balances subjects (the ones for which the observation is complete, i.e. for which $\delta = 1$), but the weights associated to those complete lifetime observations are not the same (remember that Kaplan-Meier estimator attributes more mass to large uncensored observations).

time	**n.risk**	**n.event**	**survival**	**std.err**	**lower 95% CI**	**upper 95% CI**
0	149576	20713	0.862	0.000893	0.86	0.863
1	127904	20094	0.726	0.001155	0.724	0.728
2	106847	18853	0.598	0.001274	0.596	0.601
3	87223	14128	0.501	0.001302	0.499	0.504
4	72081	11071	0.424	0.001292	0.422	0.427
5	60375	9169	0.36	0.001259	0.357	0.362
6	50551	6934	0.31	0.001217	0.308	0.313
7	42919	5267	0.272	0.001176	0.27	0.275
8	36629	3904	0.243	0.001139	0.241	0.246
9	31694	2811	0.222	0.001108	0.22	0.224
10	28321	1945	0.206	0.001084	0.204	0.209
11	26040	1605	0.194	0.001063	0.192	0.196
12	24311	1570	0.181	0.00104	0.179	0.183

Table 13: Summary of the Kaplan-Meier model fit: First twelve months survival probabilities from Kaplan-Meier estimator calibrated on right-censored balances observations as 2020-01-01

From this table we can read that for time 12, their are 24 311 balances ("n.risk") with $Y >= 12$ months, of which 1570 ("n.event") were settled at exactly 12 months old. The probability of survival after 12 months ("survival" at time 12) is, according to Kaplan-Meier estimator, of 0.181 (i.e. $P(T > 12) = \hat{S}(12) = 1 - \hat{F}(12) = 0.181$) whereas it is of 0.110 if one only takes into account uncensored balances (i.e. settled ones). The Kaplan-Meier 1-year probability of survival is not equal to $1 - \frac{\sum_{time=0}^{12} n.event}{n.risk_0}$ as it is without censored observations.

The average survival time for settled balances (without censored data) is 5.3 months, whereas it is 7.3 when including EBNYS balances (censored observations).

Let us now fit a Cox proportional-hazards model on AGRe balances' historical observations, using the following categorical covariates (whose significations have already been described previously):

Factor variable	**Levels**
accept_retro	Accept / Retro
proportional	Propor. / Non Propor.
lr_scope	LiabilityUK+ACS / PoolPropertyParRisk / CatNatNonAuto / Engineering / Transport / Property / LiabilityLong / PoolPropertyParEvt / QSJapanMotor / QSKoreaMotor / other
perimetre	Local / CouvGpe
sec_cover_form	XP / QP / XS / SL
sign	Sign of the balance amount: -1 / 1
montant_initial_euro_class	(-1,1e+03] / (1e+03,1e+04] / (1e+04,1e+05] / (1e+05,1e+06] / (1e+06,6.46e+07]
type_compte	Definitif reel / Appel au comptant / other
currency	EUR / USD / SGD / GBP / HKD / THB / JPY / KRW / other
country	FR / SG / BE / ES / CH / DE / UK / IE / HK / JP / MX / KR / other
notation	good_notation / bad_notation / other

Table 14: Revamped factor variables and their retained classes inserted in Cox regression fitting

Some of the categorical variables exhibited in the above table have seen their initial number of levels been reduced, as discussed in section 7.2.1 ("Possible explanatory variables" part). The total absolute amount 85% quantile have been taken to narrow down the number of categories within currency, country, and LR scope factors.

	coef	exp(coef)	se(coef)	z	Pr(>|z|)	
accept_retroRetro	0.471458	1.602328	0.049066	9.609	<2.00E-16	***
proportionalPropor.	1.309942	3.705961	0.008071	162.311	<2.00E-16	***
lr_scopeEngineering	-0.367023	0.692794	0.020947	-17.521	<2.00E-16	***
lr_scopeLiabilityLong	-0.476952	0.620672	0.024436	-19.519	<2.00E-16	***
lr_scopeLiabilityUK+ACS	-0.549944	0.576982	0.027119	-20.279	<2.00E-16	***
lr_scopeother	-0.511717	0.599465	0.019505	-26.236	<2.00E-16	***
lr_scopePoolPropertyParEvt	-0.378876	0.684631	0.025919	-14.618	<2.00E-16	***
lr_scopePoolPropertyParRisk	-0.232035	0.792919	0.022019	-10.538	<2.00E-16	***
lr_scopeProperty	-0.524083	0.592098	0.020489	-25.579	<2.00E-16	***
lr_scopeQSJapanMotor	-0.590841	0.553861	0.076231	-7.751	9.14E-15	***
lr_scopeQSKoreaMotor	-0.931111	0.394116	0.212714	-4.377	0.000012	***
lr_scopeTransport	-0.833865	0.434367	0.020264	-41.151	<2.00E-16	***
perimetreLocal	-0.177076	0.837716	0.01706	-10.38	<2.00E-16	***
sec_cover_formSL	1.650291	5.208497	0.031973	51.616	<2.00E-16	***
sec_cover_formXP	-0.029279	0.971146	0.009408	-3.112	0.001858	**
sign1	0.078102	1.081233	0.00564	13.849	<2.00E-16	***
montant_initial_euro_class(1e+03,1e+04]	0.07645	1.079449	0.006951	10.998	<2.00E-16	***
montant_initial_euro_class(1e+04,1e+05]	0.171134	1.18665	0.007495	22.832	<2.00E-16	***
montant_initial_euro_class(1e+05,1e+06]	0.226299	1.25395	0.010624	21.301	<2.00E-16	***
montant_initial_euro_class(1e+06,6.46e+07]	0.30174	1.352209	0.023108	13.058	<2.00E-16	***
type_compteDefinitif_reel	-0.501287	0.60575	0.016196	-30.951	<2.00E-16	***
type_compteother	-0.362796	0.695728	0.018907	-19.189	<2.00E-16	***
notationgood_notation	0.800214	2.226018	0.023574	33.945	<2.00E-16	***
notationother	1.316931	3.73195	0.054184	24.305	<2.00E-16	***
currencyGBP	0.26831	1.307753	0.019325	13.884	<2.00E-16	***
currencyHKD	-0.104532	0.900746	0.015292	-6.836	8.16E-12	***
currencyJPY	0.975213	2.651733	0.069076	14.118	<2.00E-16	***
currencyKRW	0.960842	2.613897	0.13559	7.086	1.38E-12	***
currencyother	-0.193577	0.824006	0.008644	-22.394	<2.00E-16	***
currencySGD	-0.176144	0.838497	0.016971	-10.379	<2.00E-16	***
currencyTHB	-0.240142	0.786516	0.017369	-13.826	<2.00E-16	***
currencyUSD	-0.083567	0.919829	0.008563	-9.759	<2.00E-16	***
countryCH	0.157742	1.170864	0.024579	6.418	1.38E-10	***
countryDE	0.127137	1.135572	0.023846	5.332	9.74E-08	***
countryES	-0.057194	0.944411	0.025974	-2.202	0.027667	*
countryFR	0.140226	1.150534	0.023694	5.918	3.26E-09	***
countryHK	0.028139	1.028539	0.028683	0.981	0.326581	
countryIE	0.130119	1.138964	0.025276	5.148	2.63E-07	***
countryJP	0.058596	1.060347	0.047099	1.244	0.213459	
countryKR	0.135231	1.144801	0.031086	4.35	0.0000136	***
countryMX	0.190129	1.209406	0.043604	4.36	0.000013	***
countryother	0.088675	1.092726	0.023311	3.804	0.000142	***
countrySG	0.043535	1.044497	0.029194	1.491	0.135902	
countryUK	0.009536	1.009581	0.024633	0.387	0.698672	
countryUK	0.009536	1.009581	0.024633	0.387	0.698672	
Signif. codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’					
Concordance = 0.608 (se = 0.001)						
Likelihood ratio test = 14180 on 44 df, p=<2e-16						
Wald test = 63456 on 44 df, p=<2e-16						
Score (logrank) test = 15268 on 44 df, p=<2e-16						

Table 15: Output of Cox regression fitting: regression coefficients and statistical significance

The Cox regression results of Table 15 can be interpreted as follows:

- The first column corresponds of course to the different modalities (or classes) of the factors included in the regression as potential categorical explanatory variables.
- The regression coefficients ("coef") column corresponds to the β coefficients previously introduced in the mathematical representation of Cox proportional-hazards model. The sign of those coefficients are to be noticed: a positive sign means that the hazard (risk of death or propensity to settlement in our case) is higher for that class in comparison with the reference class of the corresponding factor variable. In fact, if $\beta > 0$ then $e^{\beta} > 1$ and the considered class increase the baseline hazard function with respect to the corresponding reference level. The converse is true.
- The exponentiated coefficients ("exp(coef)"), also known as hazard ratios, give the effect size of covariates.

- The column “z” gives the Wald statistic value, which is the ratio of a regression coefficient to its standard error (z = coef/se(coef)). The next column "Pr(>|z|)" is the corresponding p-value. The Wald statistic and the related p-value evaluate whether the coefficient β of a given variable (or level of a categorical variable) is statistically significantly different from zero. From the output above, we can conclude that almost all selected variables' classes are significant coefficients (significantly different from 0).

- The final output gives the p-values of three alternative tests (likelihood-ratio test, Wald test, and score logrank statistics) to measure the overall significance of the model. These three methods are asymptotically equivalent, i.e. for a large enough number of observations, they will give similar results.

In summary, for all rows j displayed, that is for all the model degrees of freedom (in this example $j \in [\![1, 44]\!]$), we have:

$\text{coef}_j = \hat{\beta}_j$; $\exp(\text{coef})_j = e^{\hat{\beta}_j}$; $\text{se(coef)}_\text{j} = \sqrt{\widehat{Var\left(\hat{\beta}_j\right)}}$; Test of $\text{H}_0 : \beta_j = 0$: $z_j = \frac{\sqrt{n}\hat{\beta}_j}{\sqrt{\widehat{Var\left(\hat{\beta}_j\right)}}} =$ value of Wald statistic; $p - value_j = P\left(|U| > z_j\right)$ with $U \sim \mathcal{N}(0,1)$

Finally, R gives the values of the three test statistics for the hypothesis $\text{H}_0 : \beta_1 = \ldots = \beta_{44} = 0$, together with the corresponding number of degrees of freedom for the statistics' chi2 limit distribution under H_0 and the associated p-values: here, $P\left(\chi^2(44) > \text{statistic value}\right)$.

Let us examine and interpret two interesting numerical examples from the output of Cox regression fitting.

We can see that the coefficients of both levels "Definitif_reel" and "other" from the factor variable "type_compte" are negatives ($\beta = -0.501287$ and $\beta = -0.362796$ respectively). Those values are computed with reference to the class "Appel au comptant" (cash call). This translate to

$$\frac{\lambda(t \mid \text{type\_compte} = \text{Definitif\_reel, other factors being equal})}{\lambda(t \mid \text{type\_compte} = \text{Appel au comptant, other factors being equal})} = e^{-0.501287} = 0.6057506$$

The hazard ratio (HR) for the balance group "Definitif_reel" relative to the balance group "Appel au comptant", is around 0.6, which means that "Definitif_reel" balances have a lower "risk of death" (i.e. tendency to transform into cash flow) than "Appel au comptant" balances. The hazard rate (the whole hazard base function) is indeed reduced by a factor of 0.6, which is consistent with what we have observed in the descriptive analysis of data: cash calls settle in average much faster than other balances types.

Regarding the amount categorical variable (montant_initial_euro_class), for which the model has considered the lowest amounts class (-1,1e+03] as the reference class, we can note that classes with higher amounts have higher regression coefficients, which translate to higher hazard ratios (comparatively to (-1,1e+03] class) and higher hazard function (i.e higher susceptibility for associated balances to be settled rapidly).

5.4.2 Direct application to historical financial balances' data: model prediction

We will begin this part with graphical results coming from Cox regressions fitted to the same data but taking into account one only explanatory variable at a time, in order to display how estimated survival depends upon the value of a covariate of interest.

Below (Table 16) are the Cox regression coefficients and their statistical significance, in the case of the balance amount class variables as sole explanatory factor. Immediately after, follows the corresponding graph (Figure 49) showing the predicted survival proportion (estimated survival function) as a function of time (months) and as a function of the balance amount category.

	coef	**exp(coef)**	**se(coef)**	**z**	**Pr(>|z|)**	
montant_initial_euro_class(1e+03,1e+04]	0.101391	1.106709	0.006867	14.77	<2e-16	***
montant_initial_euro_class(1e+04,1e+05]	0.253698	1.288782	0.007231	35.08	<2e-16	***
montant_initial_euro_class(1e+05,1e+06]	0.391781	1.479614	0.01001	39.14	<2e-16	***
montant_initial_euro_class(1e+06,6.46e+07]	0.621785	1.862249	0.022084	28.16	<2e-16	***

Table 16: Output of Cox regression fitting with the balance amount factor as sole explanatory variable: regression coefficients and statistical significance

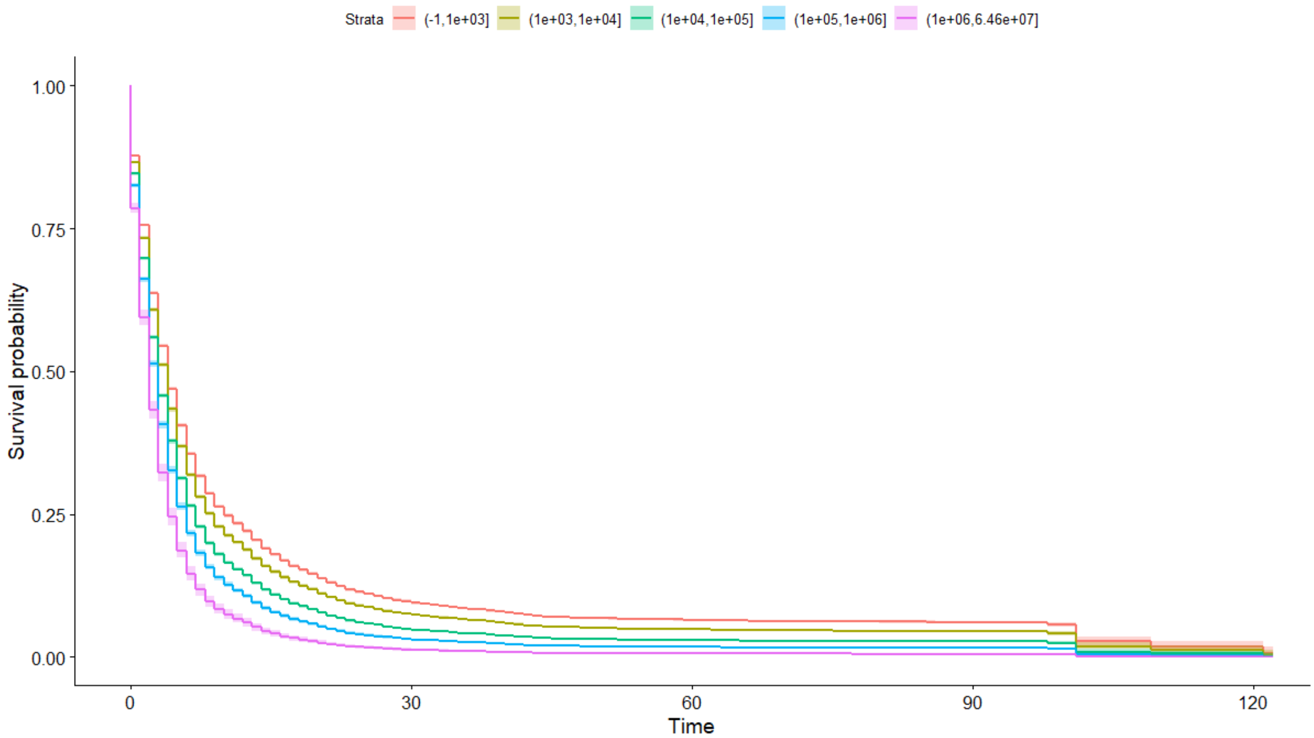


Figure 49: Predicted survival proportion as a function of time (months) for particular amount balance groups

Those regression coefficients' values and the associated very distinctive calibrated survival curves are in line with what had already come to our attention, namely that higher the balance amount, quicker its settlement is prone to occur. Each subsequent figure (Figures 50 to 52) represents other examples of estimated survival functions through Cox regression with a single categorical explanatory variable. We can discern clearly that the form of the calibrated survival curve depends strongly on the modalities making up a given factor variable. We can thus visually and easily assess the impact of those factors' values on the estimated survival probability function.

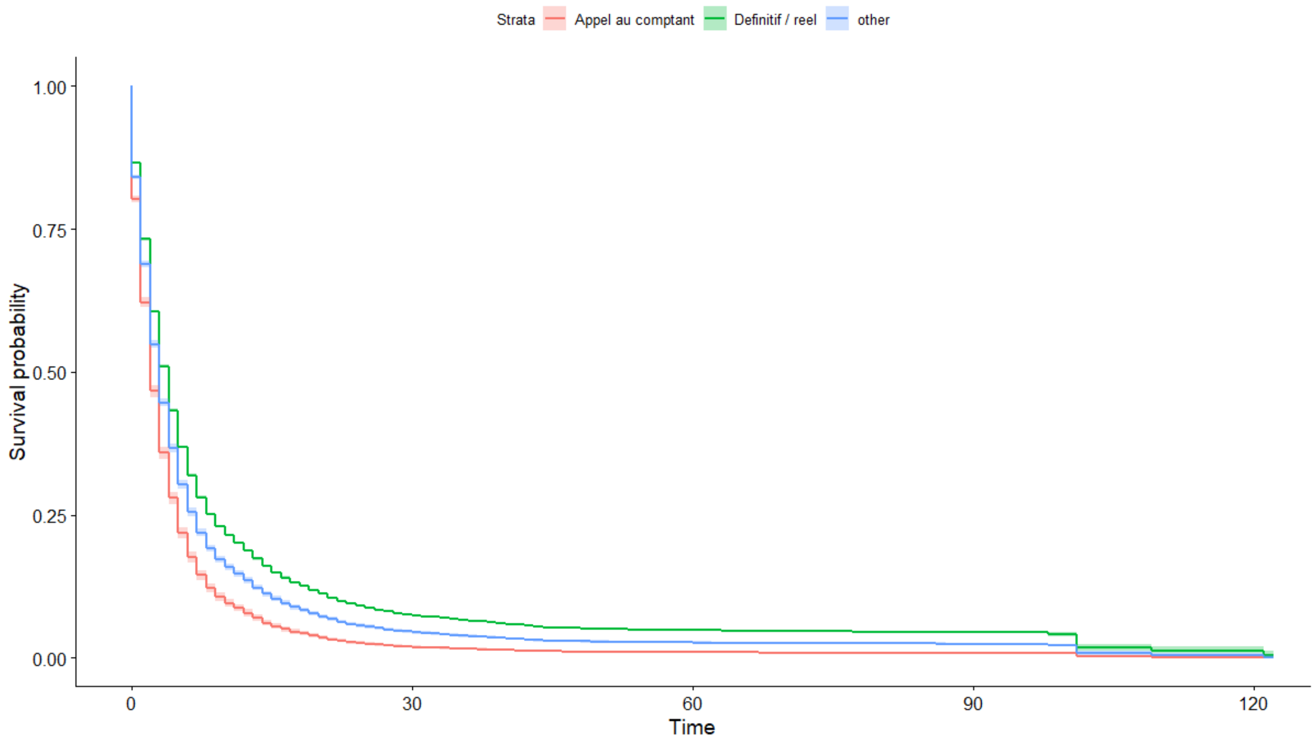


Figure 50: Predicted survival proportion as a function of time (months) for particular type of account balance groups

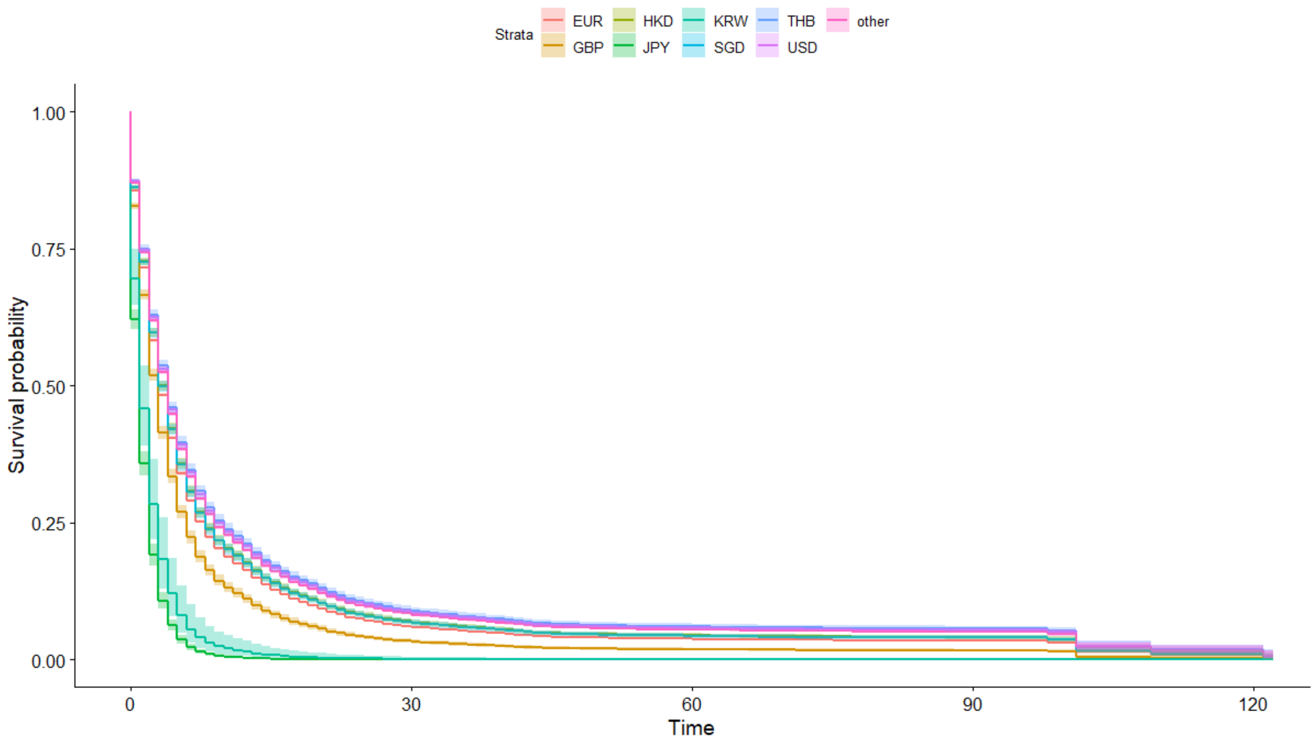


Figure 51: Predicted survival proportion as a function of time (months) for particular currency balance groups

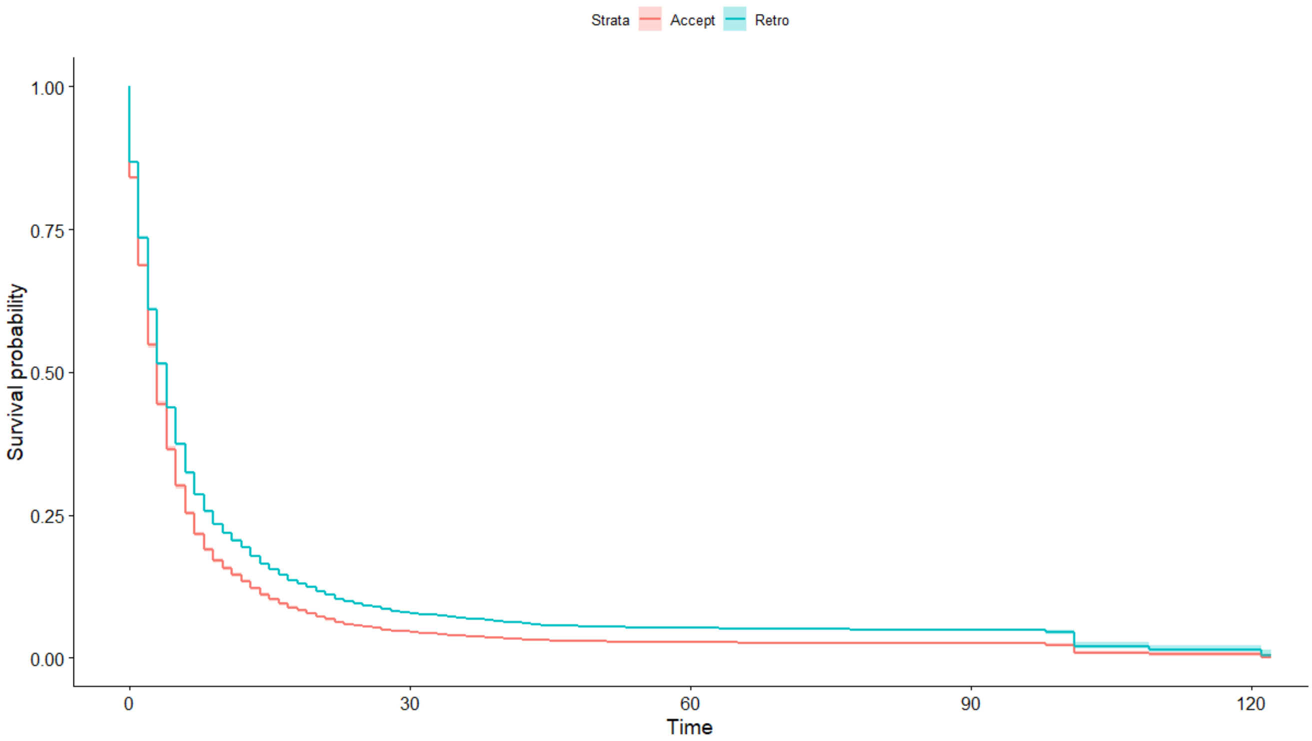


Figure 52: Predicted survival proportion as a function of time (months) for Accept or Retro balance groups

Once the Cox proportional-hazard model has been fitted on the as-of-date balances observations, we have access to a specific survival function estimator $\hat{S}_i$ for each EBNYS balances i. Those survival functions are not conditional to the time already lived by EBNYS balances. In other words, for a given EBNYS balances i, the corresponding survival function estimator $\hat{S}_i$ does not take into consideration the elapsed time from this balance entry date to the date of modeling and projection (here as of date 2020-01-01). For each EBNYS balances (as of 2020-01-01) we then have to adjust their estimated survival probability function incorporating the information of their current age (as of 2020-01-01). That is, we have to compute their estimated conditional survival probability function given that they have already survived a certain number of months.

Based on the conditional survival probability formula

$$\forall t_2 \geq t_1, \quad S(t_2 \mid t_1) = P(T > t_2 \mid T > t_1) = \frac{P(T > t_2)}{P(T > t_1)} = \frac{S(t_2)}{S(t_1)}$$

we compute for each as-of-date EBNYS balances i, and for all t_2 (in months) $\geq$ already elapsed months (i.e current age of balance i):

$$\hat{S}_i(t_2 \mid \text{already elapsed months}) = \frac{\hat{S}_i(t_2)}{\hat{S}_i(\text{already elapsed months})}$$

from which we can retrieve the estimated conditional probability distribution for all t_2 (in months) $>$ current age of balance i:

$$\begin{aligned}
\hat{P}(T_i = t_2 \mid \text{age of balance } i) &= \hat{P}((t_2 - 1 < T_i \leq t_2) \mid \text{age of balance } i) \\
&= \hat{S}_i(t_2 - 1 \mid \text{age of balance } i) - \hat{S}_i(t_2 \mid \text{age of balance } i) \\
&= \frac{\hat{S}_i(t_2 - 1) - \hat{S}_i(t_2)}{\hat{S}_i(\text{age of balance } i)}
\end{aligned}$$

Denoting $T_i' = T_i -$current age of balances i, the survival time of EBNYS balance i (from as-of-date d), the above relationship is equivalent to say that for all $t \in \mathbb{N}^*$ (in our case discrete t since survival is measured in number of months):

$$\hat{P}(T_i' = t) = \frac{\hat{S}_i(t + \text{age of balance } i - 1) - \hat{S}_i(t + \text{age of balance } i)}{\hat{S}_i(\text{age of balance } i)}$$

We finally get a discrete survival probability distribution for each EBNYS lines (whose amounts we aim to project into cash flows), taking into account their individual characteristics (combination of values taken by selected explanatory variables) and their current age at d.

A Monte Carlo simulation is then performed to get as many slices of probable survival times over all EBNYS lines as the number of simulation we choose. Taking as an example the number of EBNYS lines as of 2020-01-01 (16 956) and carrying out N simulations, we obtain the following matrix proceeding from the Monte Carlo simulation:

Balance	**Currency and other characteristics**	**Simulation 1**	**Simulation 2**	...	...	**Simulation N**
$EBNYS_1$	...	$T'_{1,1}$	$T'_{1,2}$	...	...	$T'_{1,N}$
$EBNYS_2$	...	$T'_{2,1}$	$T'_{2,2}$	...	...	$T'_{2,N}$
...	...	...	...	...	...	...
...	...	...	...	...	...	...
$EBNYS_{16956}$	...	$T'_{16956,1}$	$T'_{16956,2}$	...	...	$T'_{16956,N}$

Table 17: Monte Carlo procedure applied to EBNYS balances' survival times sampling

From this Monte Carlo process output, we achieve, via a few data aggregations and transformations, the following table giving for each currency the aggregated cash flows over the next 12 months (a column), this for every simulations run.

Currency	**Time to settlement (in months)**	**Simulation 1**	**Simulation 2**	...	...	**Simulation N**
EUR	1	CF_$EUR_{1,1}$	CF_$EUR_{1,2}$	...	...	CF_$EUR_{1,N}$
EUR	2	CF_$EUR_{2,1}$	CF_$EUR_{2,2}$	...	...	CF_$EUR_{2,N}$
...	...	...	...	...	...	...
EUR	12	CF_$EUR_{12,1}$	CF_$EUR_{12,2}$	...	...	CF_$EUR_{12,N}$
USD	1	CF_$USD_{1,1}$	CF_$USD_{1,2}$	...	...	CF_$USD_{1,N}$
USD	2	CF_$USD_{2,1}$	CF_$USD_{2,2}$	...	...	CF_$USD_{2,N}$
...	...	...	...	...	...	...
USD	12	CF_$USD_{12,1}$	CF_$USD_{12,2}$	...	...	CF_$USD_{12,N}$
GBP	1	CF_$GBP_{1,1}$	CF_$GBP_{1,2}$	...	...	CF_$GBP_{1,N}$
GBP	2	CF_$GBP_{2,1}$	CF_$GBP_{2,2}$	...	...	CF_$GBP_{2,N}$
...	...	...	...	...	...	...
GBP	12	CF_$GBP_{12,1}$	CF_$GBP_{12,2}$	...	...	CF_$GBP_{12,N}$
...	...	...	...	...	...	...
...	...	...	...	...	...	...

Table 18: Aggregate cash flows in currencies overs next 12 months stemming from survival times Monte Carlo simulations

Those N simulation of next 12 months aggregate cash flows by currency can visually be rendered into the following distribution densities examples (one density curve for each future month cash flow random value):

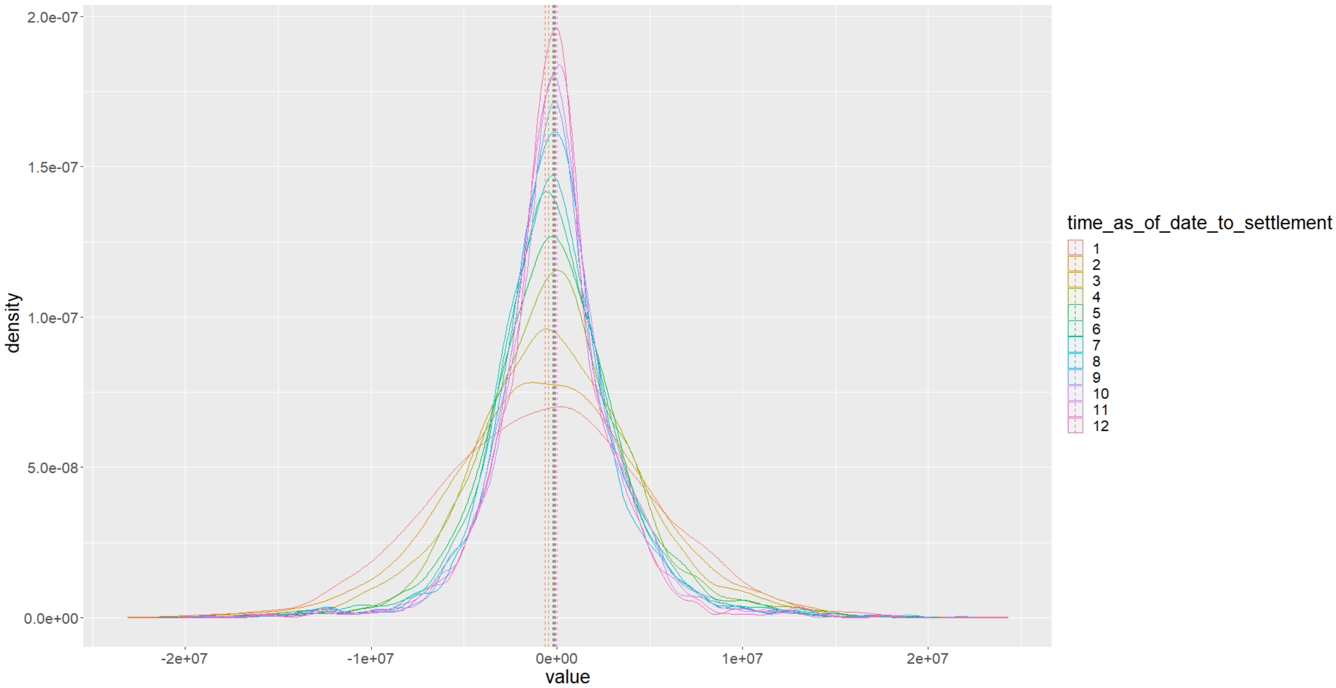


Figure 53: Simulated distribution densities and averages of future 12 monthly cash flows random values – all balances' amounts – currency EUR

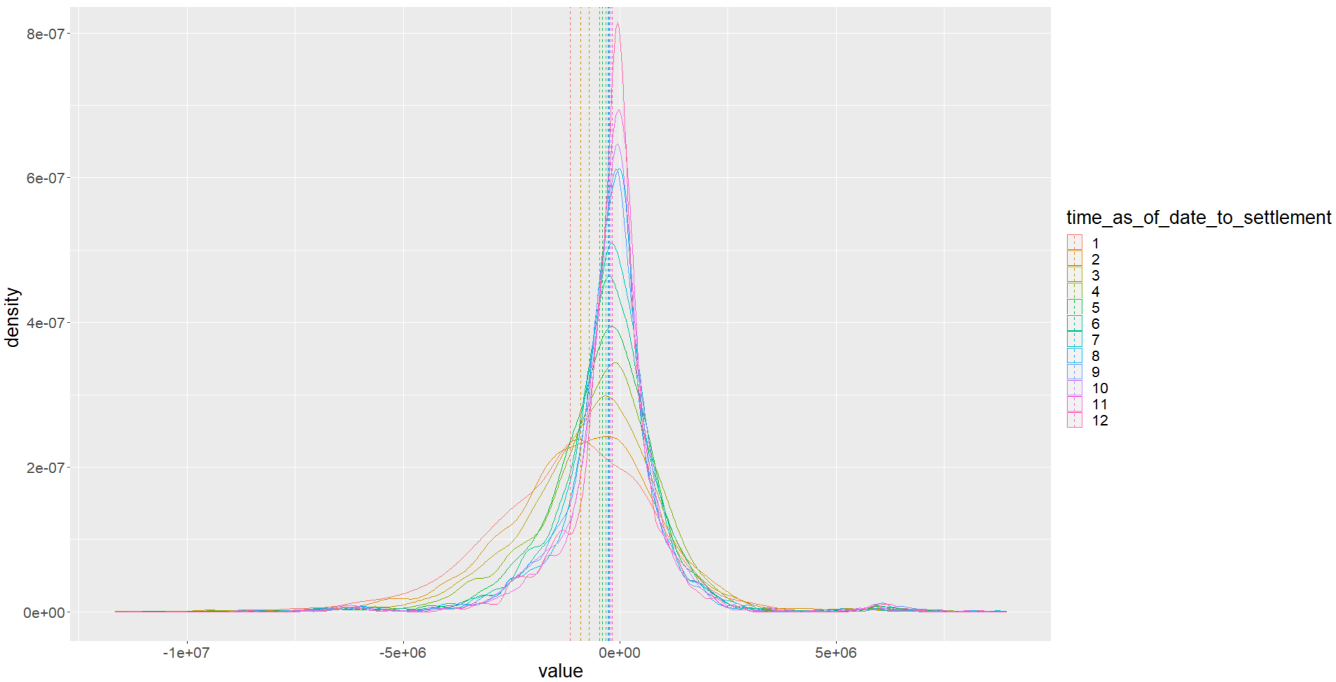


Figure 54: Simulated distribution densities and averages of future 12 monthly cash flows random values – all balances' amounts – currency USD

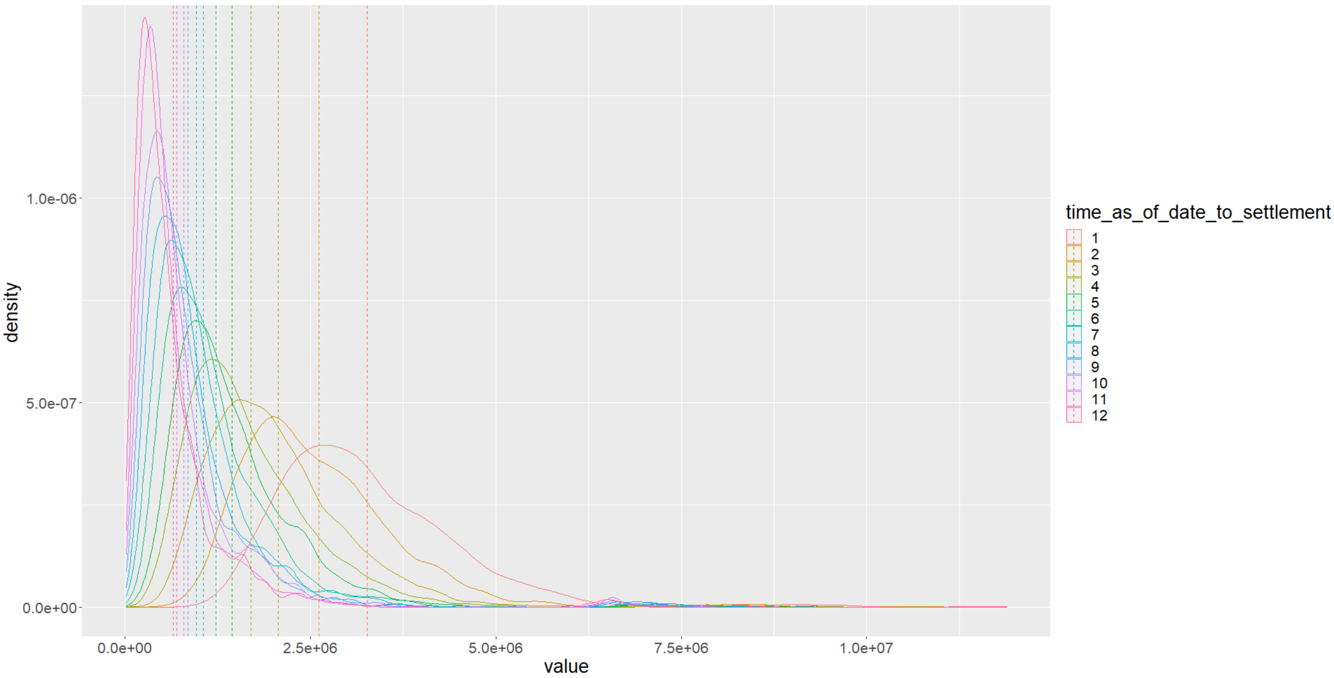


Figure 55: Simulated distribution densities and averages of future 12 monthly cash flows random values – positive balances' amounts – currency USD

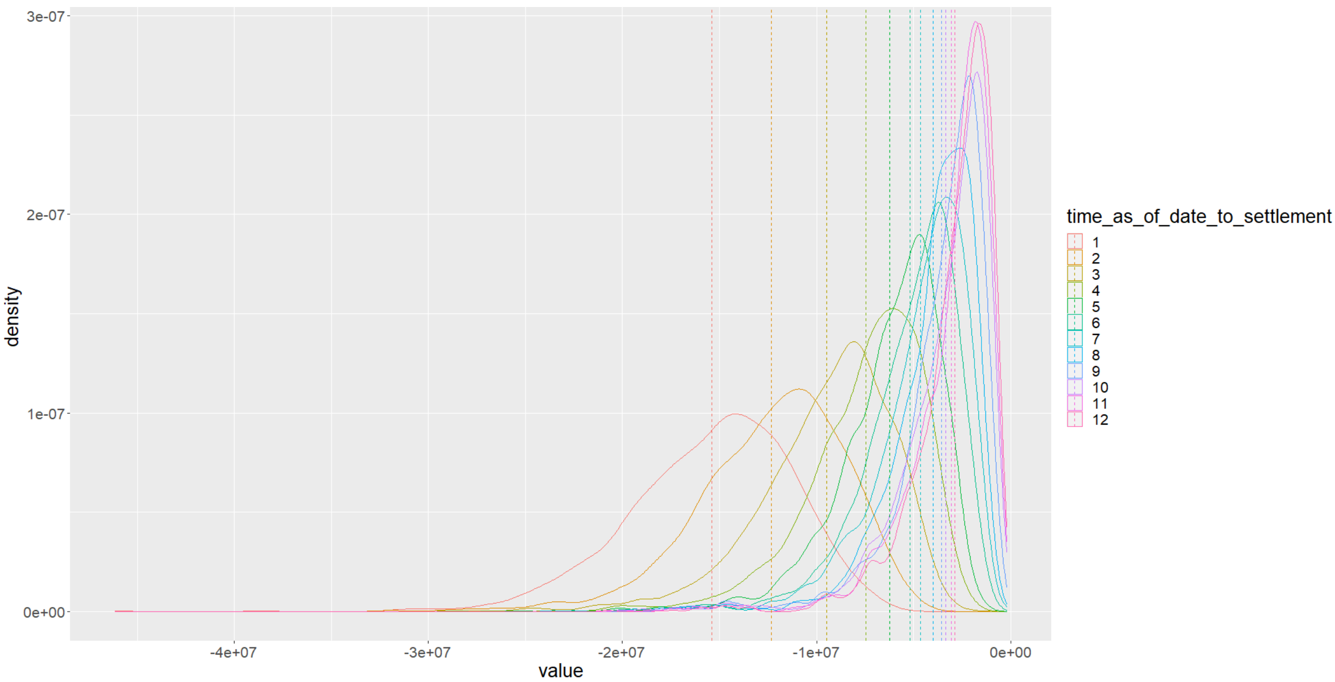


Figure 56: Simulated distribution densities and averages of future 12 monthly cash flows random values – negative balances' amounts – currency EUR

If the currency considered is different from the euro, the value on the x-axis is the value of cash flows in euro equivalent. Figures 53 and 54 show the predicted densities, after N = 2000 simulations, of the 12 future monthly cash flows values (in euro equivalent), taking into account both positive and negative balances amounts (for EUR and USD labelled accounting balances respectively). Those estimated densities are almost centered to zero (on average, settlement of balances with positive amounts offsetting settlement of balances with negative amounts) and are rather dispersed (not highly informative).

Figure 55 presents those predicted densities only for positive amount balances labelled in USD. As for Figures 56, we can see simulated densities for next-12-month cash flows values corresponding exclusively to negative amount balances expressed in EUR.

For all those displayed configurations, one can notice that the predicted density of remote (in terms of number of months from the date of projection) cash flows are closer to 0 and less disperse. To put it differently, the random values of far off cash flows (in relation to as-of-date d) are less volatile and are more likely to be small in a run-off set up. This is indeed to be expected, since we saw that most not problematic EBNYS financial balances are settled during the first months after being written, which tend to generate more settlements and thus higher absolute amounts and more dispersion of projected cash flows values during the first months after the date d of projection. As for problematic and aged EBNYS financial balances, the duration model will generally (over the N simulations) throw their amounts at later months (after the 12 months following d). In a run-off configuration, that is basing ourselves solely on existing balances at d to forecast future (as of d) cash flows, higher is the time lapse from d, narrower and closer to zero is the distribution of the corresponding cash flows' aggregate amounts resulting from our model (as it is also observed historically).

The averages over all simulations of every simulated cash flows values are then computed and constitute the prediction within 12 months of run-off monthly cash flows amounts by currency.

Currency	**Time to settlement (in months)**	**Average over all simulations**
EUR	1	$\frac{1}{N}\sum_{i=1}^{N}$ CF_$EUR_{1,i}$
EUR	2	$\frac{1}{N}\sum_{i=1}^{N}$ CF_$EUR_{2,i}$
...	...	...
EUR	12	$\frac{1}{N}\sum_{i=1}^{N}$ CF_$EUR_{12,i}$
USD	1	$\frac{1}{N}\sum_{i=1}^{N}$ CF_$USD_{1,i}$
USD	2	$\frac{1}{N}\sum_{i=1}^{N}$ CF_$USD_{2,i}$
...	...	...
USD	12	$\frac{1}{N}\sum_{i=1}^{N}$ CF_$USD_{12,i}$
GBP	1	$\frac{1}{N}\sum_{i=1}^{N}$ CF_$GBP_{1,i}$
GBP	2	$\frac{1}{N}\sum_{i=1}^{N}$ CF_$GBP_{2,i}$
...	...	...
GBP	12	$\frac{1}{N}\sum_{i=1}^{N}$ CF_$GBP_{12,i}$
...	...	...
...	...	...

Table 19: Average cash flows in currencies over next 12 months stemming from survival times Monte Carlo simulations

To close this part, let us summarize the different steps of the duration model (Cox regression or Kaplan-Meier) algorithm, leading to the prediction of near future cash flows by currency (as of date d):

1. Select predictor variables (for Cox model);
2. Fit the model to all as-of-date balances observations;
3. Extract estimated survival functions for EBNYS balances: Cox regression provides each of the observations with a proper survival function depending on the combination of classes

it belongs to, while Kaplan-Meier method produces only one common survival function estimator for all balances;

4. Rearrange the survival functions estimators so as to consider the balances' seniorities as of date d.

5. Deduce the survival time discrete distribution proper to each EBNYS financial balances (conditional on the age of the balance at d, plus its particularities in the case of Cox model);

6. Simulate survival times samples from those conditional probabilities through Monte Carlo method;

7. Aggregate and format previous data in order to obtain samples of monthly (up to 12 months after d) technical cash transactions aggregated amounts by currency.

8. Finally, take the mean over all simulations of the previously derived cash flows samples, which brings about what we have been looking for: the prediction of monthly cash flows (in currencies) within a year originating from EBNYS accounting balances at d.

5.4.3 Performance criteria and backtest

In order to compare our duration model predictions with real observed run-off cash flows and measure its performance with respect to the currently applied "dummy model", we need to define first the yardsticks we will use to that end. Let us introduce frequently used measures of the differences between the values predicted by a model (we will denote $(\hat{y}_i)_{i\in[\![1,n]\!]}$) and the values observed ($(y_i)_{i\in[\![1,n]\!]}$), thus assessing the quality of a predictor (i.e. a function mapping arbitrary inputs to a sample of values of some random variable):

- Normalized Mean Square Error (NMSE)

$$\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i)^2}$$

- Normalized Root Mean Square Error (or Deviation) (NRMSE or NRMSD)

$$\frac{\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}}{\max_i(y_i)-\min_i(y_i)}$$

- Normalized Mean Absolute Error (NMAE)

$$\frac{\sum_{i=1}^{n}|y_i-\hat{y}_i|}{\sum_{i=1}^{n}|y_i|}$$

- Mean Absolute Percentage Error (MAPE)

$$\frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right|$$

Those different error measurements between estimated/predicted and actual values are measures of accuracy allowing to compare forecasting errors of different models for a particular data set. They all serve to aggregate the magnitudes of the errors in predictions for various data points into a single measure of predictive power. Nevertheless, they have their own characteristics and turn out to be complementary. Being all normalized, those indicators allow the error to be compared across data with different scales.

The MSE represents the quadratic mean of the differences (or deviations) between predicted and observed values. These deviations are called residuals when the calculations are performed over the data sample that was used for estimation and are called prediction errors when computed out-of-sample. The RMSE (or RMSD) take the root of the MSE, thus bringing the unit back to observations scale, which may help to interpret the model accuracy. Normalizing the MSE into NMSE and the RMSE into NRMSE facilitates the comparison between data sets or models with different scales. A common choices of normalization for the NRMSE is dividing by the range (maximum minus minimum value) of the observed data. A lower NMSE or NRMSE indicates less residual or prediction error variance.

MAPE is the absolute error normalized over the absolute observed value, computed for every data point and then averaged. NMAE is different from MAPE in that the mean error is normalized over the average of the actual values. The small difference in the way the error is computed between the NMAE and MAPE error measurements, can produce very different results. MAPE being computed over every data point and averaged, can therefore captures more errors and outliers. NMAE on the other hand can lose some of the detail because of the aggregation of errors performed before the averaging. Moreover, MAPE is asymmetric and puts a heavier penalty on negative errors (when forecast values are higher than actual ones) than on positive errors.

NMSE and NRMSE will favor predictions that are correct on average (i.e unbiased) and will penalize greatly the highest errors (sensitivity to outliers), whereas NMAE and MAPE will not give as much importance to the most significant errors.

Nothing better than a simple numerical example to see how those 4 metrics behave under "extreme" values. If we take $y = (1, 1, 100)$ and $\hat{y} = (100, 100, 1)$ we obtain $NMSE(y, \hat{y}) \approx 2.9, NRMSE(y, \hat{y}) \approx 1, NMAE(y, \hat{y}) \approx 2.9, MAPE(y, \hat{y}) \approx 66.3$ (asymmetry of MAPE stressing negative errors).

Specifically, if we denote for all $i \in [\![1, 12]\!]$ $CF_i\_XXX\_predicted\_duration$, $CF_i\_XXX\_predicted\_dummy$, and $CF_i\_XXX\_observed$, the cash flows amounts relative to currency XXX balances taking place at month i after the date of reference d, respectively either predicted by a duration model, or projected by the dummy model, or actually observed, we obtain in agreement with the general definitions of the NMSE:

$$\text{NMSE(dummy model)}_{XXX} = \frac{\sum_{i=1}^{12} (CF_i\_XXX\_observed - CF_i\_XXX\_predicted\_dummy)^2}{\sum_{i=1}^{12} (CF_i\_XXX\_observed)^2}$$

$$\text{NMSE(duration model)}_{XXX} = \frac{\sum_{i=1}^{12} (CF_i\_XXX\_observed - CF_i\_XXX\_predicted\_duration)^2}{\sum_{i=1}^{12} (CF_i\_XXX\_observed)^2}$$

In the current framework, NRMSE, NMAE and MAPE are likewise computed, according to their general expressions for predictors and estimators.
We can similarly assess our prediction models performance on the basis of aggregated cash flows gathering all currencies in euro equivalent. Employing the notations $CF_i\_predicted\_duration$, $CF_i\_predicted\_dummy$, and $CF_i\_observed$, we have taken this time the NMAE as an exam-

ple:

$$\text{NMAE(dummy model)} = \frac{\sum_{i=1}^{12} |CF_i\_observed - CF_i\_predicted\_dummy|}{\sum_{i=1}^{12} |CF_i\_observed|}$$

$$\text{NMAE(duration model)} = \frac{\sum_{i=1}^{12} |CF_i\_observed - CF_i\_predicted\_duration|}{\sum_{i=1}^{12} |CF_i\_observed|}$$

In addition to comparing the closeness of predicted (whether by the dummy model or by a duration model) cash flows with observed ones, it is equally relevant for us to compare the cumulative values of those cash flows. That is, for all $k \in [\![1, 12]\!]$, to juxtapose and gauge the distances (through NMSE, NRMSE, NMAE and MAPE) between $\sum_{i=1}^{k}(CF_i\_observed)$ vs $\sum_{i=1}^{k}(CF_i\_predicted)$ (dummy or duration model). Cumulative amounts of cash flows up to a certain horizon of time are indeed an interesting perspective to inspect since it answers the question: over the next k months (from as-of-date d to $d + k$ months) how much cash will we need (or receive) in XXX (EUR, USD, GBP, ...)?

This is how we offer to appraise the performance of implemented duration models (Kaplan-Meier estimator and Cox proportional-hazards model): through the above presented criteria applied to both cash flows values and their cumulative counterparts, over the 12 months following the date d of model calibration and prediction. Nevertheless, in order to have a more robust performance assessment that does not depend on the date d, we will estimate the average of those distance measurements over many as-of-dates d. To reach this goal, we built a loop making d covers the first of every months between 2015 and 2020 ($d =$ 2015-01-01, $d =$ 2015-02-01, $d =$ 2015-03-01, ..., $d =$ 2020-11-01). For each of those dates d, we train the duration model on as-of-date data and project the amounts of EBNYS balances at d according to this model as well as according to the dummy model. NMSE(d), NRMSE(d), NMAE(d) and MAPE(d) are then computed on the basis of those predicted cash flows (issued from EBNYS balances passed through the as-of-date model an the dummy model) and the actual cash flows observed after d (from d to d + 12 months).

Let us display the average of those indicators over all chosen as-of-dates (first of every month between January 2015 and September 2020) based on predictions issued from 2000 Monte Carlo simulations (for duration models). Each of the following tables (Tables 20 to 25) shows the average NMSE, NRMSE, NMAE and MAPE for each of the 5 top currencies (as well as cash flows labelled in "other" currencies), and for each of the 3 predictive models considered, namely the dummy model, Kaplan-Meier one, and Cox regression based model (with all the categorical predictors resulting in Table 15: we have tested other combinations of categorical variables but the results are similar). There are 6 tables, 3 of them are related to individual monthly cash flows and the 3 others to cumulative monthly cash flows. The first two tables refer to models calibrated and used for prediction on the basis of all balances. Tables 22 and 23 are associated to predictions based on models trained and tested on balances with positive amounts. Finally, the two last tables (Tables 24 and 25) are linked to predictions based on negative amounts balances.

Individual cash flows	Average over as-of-dates dummy_model				Average over as-of-dates Kaplan_Meier				Average over as-of-dates Cox_regression			
Currency	NMSE	NRMSE	NMAE	MAPE	NMSE	NRMSE	NMAE	MAPE	NMSE	NRMSE	NMAE	MAPE
CHF	0,95	0,26	1,43	21753,52	0,95	0,26	1,29	13091,57	0,99	0,26	1,30	13556,93
EUR	0,90	0,28	1,11	1677,75	0,85	0,27	1,00	9214,83	0,82	0,26	0,98	11697,76
GBP	1,76	0,33	1,84	34725,20	7,77	0,56	2,97	50334,62	6,42	0,54	2,85	55836,21
HKD	0,99	0,26	1,26	54263,92	0,96	0,26	1,12	29691,29	0,95	0,26	1,13	32439,69
USD	0,99	0,28	1,11	1547,31	0,98	0,28	1,06	2345,83	0,97	0,28	1,06	4412,40
other	0,95	0,25	1,09	6761,99	0,97	0,25	1,05	2903,45	0,96	0,25	1,06	1776,95

Table 20: Average (over as-of-dates) NMSE, NRMSE, NMAE, and MAPE of predicted (by dummy, Kaplan-Meier, and Cox models) vs observed cash flows - balances of all amounts

Cumulative cash flows	Average over as-of-dates dummy_model				Average over as-of-dates Kaplan_Meier				Average over as-of-dates Cox_regression			
Currency	NMSE	NRMSE	NMAE	MAPE	NMSE	NRMSE	NMAE	MAPE	NMSE	NRMSE	NMAE	MAPE
CHF	0,48	0,88	0,56	8406,14	0,46	0,75	0,57	16356,85	0,41	0,68	0,54	17355,12
EUR	0,50	0,46	0,63	1670,30	0,42	0,37	0,56	9212,33	0,42	0,35	0,53	11695,90
GBP	11,48	1,68	2,34	1096,83	20,16	1,74	2,56	3171,99	18,51	1,65	2,52	5201,17
HKD	1,73	0,89	1,11	5462,13	0,99	0,71	0,89	8983,41	1,02	0,71	0,91	9659,77
USD	2,26	0,63	1,11	1545,21	1,85	0,61	1,11	2346,23	2,09	0,62	1,15	4413,27
other	0,90	0,48	0,83	5,61	0,87	0,46	0,80	496,90	0,85	0,45	0,80	107,28

Table 21: Average (over as-of-dates) NMSE, NRMSE, NMAE, and MAPE of predicted (by dummy, Kaplan-Meier, and Cox models) vs observed cumulative cash flows - balances of all amounts

Taking into consideration both positive and negative amounts balances we can remark from above tables that the duration models do not clearly outperform the dummy model regarding the prediction of near future individual monthly cash flows. They seem to do a bit better concerning the projection of cumulative cash flows, at least according to 3 (NMSE, NRMSE and NMAE) out of 4 performance measures.

Individual cash flows	Average over as-of-dates dummy_model				Average over as-of-dates Kaplan_Meier				Average over as-of-dates Cox_regression			
Currency	NMSE	NRMSE	NMAE	MAPE	NMSE	NRMSE	NMAE	MAPE	NMSE	NRMSE	NMAE	MAPE
CHF	321,57	1,09	10,52	86225,47	25,40	0,45	3,40	41109,54	30,98	0,47	3,60	40770,21
EUR	3,12	0,63	2,23	33789,83	0,63	0,29	0,88	50782,86	0,68	0,30	0,90	61038,36
GBP	3237,31	13,78	80,28	2537835,79	109,80	2,70	13,96	436951,14	108,47	2,72	14,18	441109,18
HKD	25,59	0,62	4,68	350059,82	3,85	0,38	2,13	174690,81	3,52	0,38	2,14	184507,89
USD	1,71	0,51	1,77	16970,69	0,57	0,29	0,87	15517,52	0,63	0,31	0,90	16985,09
other	2,14	0,45	1,69	61969,35	0,68	0,31	0,95	32397,62	0,72	0,31	0,97	34996,64

Table 22: Average (over as-of-dates) NMSE, NRMSE, NMAE, and MAPE of predicted (by dummy, Kaplan-Meier, and Cox models) vs observed cash flows - balances of positive amounts

Cumulative cash flows	Average over as-of-dates dummy_model				Average over as-of-dates Kaplan_Meier				Average over as-of-dates Cox_regression			
Currency	NMSE	NRMSE	NMAE	MAPE	NMSE	NRMSE	NMAE	MAPE	NMSE	NRMSE	NMAE	MAPE
CHF	4533,98	16,85	15,41	17719,58	336,20	8,75	4,25	5904,92	386,74	8,04	4,43	6452,43
EUR	3,14	1,22	1,20	33750,38	0,34	0,36	0,37	50774,63	0,40	0,38	0,41	61030,60
GBP	14584,66	60,26	66,03	176560,10	317,32	10,67	11,17	36687,70	337,90	11,04	11,59	42731,08
HKD	191,24	2,54	5,05	44331,79	24,72	1,07	2,01	79222,67	23,59	1,03	2,00	99272,35
USD	1,51	0,92	0,99	10726,21	0,19	0,29	0,35	13757,42	0,25	0,34	0,42	15299,37
other	2,40	0,83	0,82	9479,48	0,33	0,37	0,39	18467,18	0,37	0,38	0,42	21586,44

Table 23: Average (over as-of-dates) NMSE, NRMSE, NMAE, and MAPE of predicted (by dummy, Kaplan-Meier, and Cox models) vs observed cumulative cash flows - balances of positive amounts

This is where we can see a real difference made by the duration models in comparison with the dummy one. Above results (Tables 22 and 23) show a steep increase in predictions accuracy passing from the dummy model to duration ones, especially for cumulative projections. All four error metrics are reduced by half and more, for all currencies. Nevertheless, we do not notice better results for Cox regression model with respect to the simpler model Kaplan-Meier.

Individual cash flows	Average over as-of-dates dummy_model				Average over as-of-dates Kaplan_Meier				Average over as-of-dates Cox_regression			
Currency	NMSE	NRMSE	NMAE	MAPE	NMSE	NRMSE	NMAE	MAPE	NMSE	NRMSE	NMAE	MAPE
CHF	22,10	0,48	4,20	82846,82	2,06	0,31	1,85	42127,10	1,91	0,30	1,82	42204,49
EUR	1,56	0,48	1,75	33256,03	0,47	0,26	0,78	35825,23	0,49	0,26	0,79	44543,90
GBP	4692,68	13,04	70,02	1516385,81	305,82	2,98	15,56	287223,82	312,58	2,99	15,72	286297,71
HKD	1,06	0,35	1,70	453297,02	0,83	0,31	1,27	236455,84	0,89	0,32	1,30	256119,53
USD	1,44	0,50	1,71	49004,68	0,64	0,33	0,96	30836,85	0,70	0,34	1,00	35439,37
other	1,34	0,41	1,47	104679,05	0,60	0,30	0,89	54679,22	0,65	0,31	0,92	61872,12

Table 24: Average (over as-of-dates) NMSE, NRMSE, NMAE, and MAPE of predicted (by dummy, Kaplan-Meier, and Cox models) vs observed cash flows - balances of negative amounts

Cumulative cash flows	Average over as-of-dates dummy_model				Average over as-of-dates Kaplan_Meier				Average over as-of-dates Cox_regression			
Currency	NMSE	NRMSE	NMAE	MAPE	NMSE	NRMSE	NMAE	MAPE	NMSE	NRMSE	NMAE	MAPE
CHF	144,97	3,96	2,82	6813,71	8,79	1,95	0,92	8121,67	7,35	1,82	0,87	9550,11
EUR	1,28	0,87	0,85	33060,19	0,14	0,28	0,28	35782,84	0,16	0,28	0,29	44506,46
GBP	9523,90	56,27	57,18	103623,99	607,05	12,78	12,75	20782,54	643,91	12,89	12,94	20922,86
HKD	2,82	0,88	1,02	61250,65	0,82	0,64	0,73	96022,03	0,92	0,64	0,77	123062,68
USD	1,28	0,83	0,94	19279,82	0,26	0,33	0,40	24257,61	0,32	0,36	0,45	29103,09
other	1,15	0,61	0,68	21297,62	0,20	0,30	0,35	34007,27	0,24	0,31	0,38	42613,07

Table 25: Average (over as-of-dates) NMSE, NRMSE, NMAE, and MAPE of predicted (by dummy, Kaplan-Meier, and Cox models) vs observed cumulative cash flows - balances of negative amounts

From Tables 24 and 25 we can make the same observations for negative balances related results as for positive balances ones, that is, a higher performance of duration models for predicting future individual and cumulative cash flows (in currencies) in relation to the dummy model, but no sensible improvement between Kaplan-Meier method and Cox regression one.

Let us now have a look on specific 12-month projections of cash flows at specific dates, as individual underlying elements comprising the average metrics we have just exhibited. The graphs we present below will allow us to illustrate some instances of predictions made by the models, confront them directly and visually to corresponding observed monthly cash flows, and better interpret previously introduced overall and aggregated results. The x-axis of those graphs represents the time (in months) from the date of reference considered to the future monthly cash flows, and the y-axis unit is the euro equivalent amount of those monthly transactions. Other graphs displaying other samples (different mixtures of as-of-date x currency x sign of balances amounts), of predicted near future monthly cash flows versus their actually observed counterparts, are to be found in Appendix D.

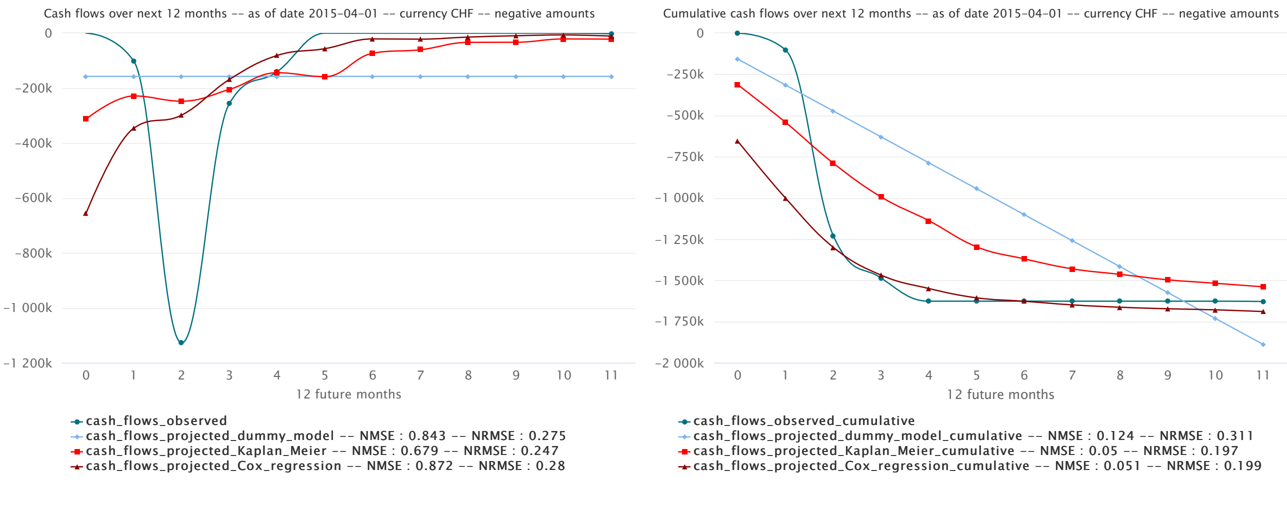


Figure 57: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2015-04-01 – currency CHF – negative amounts

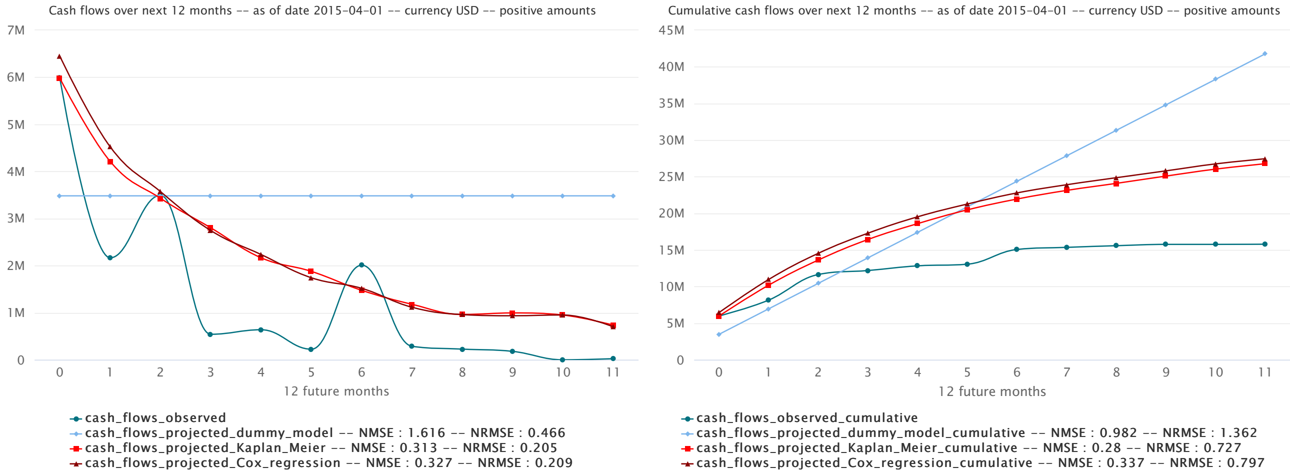


Figure 58: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2015-04-01 – currency USD – positive amounts

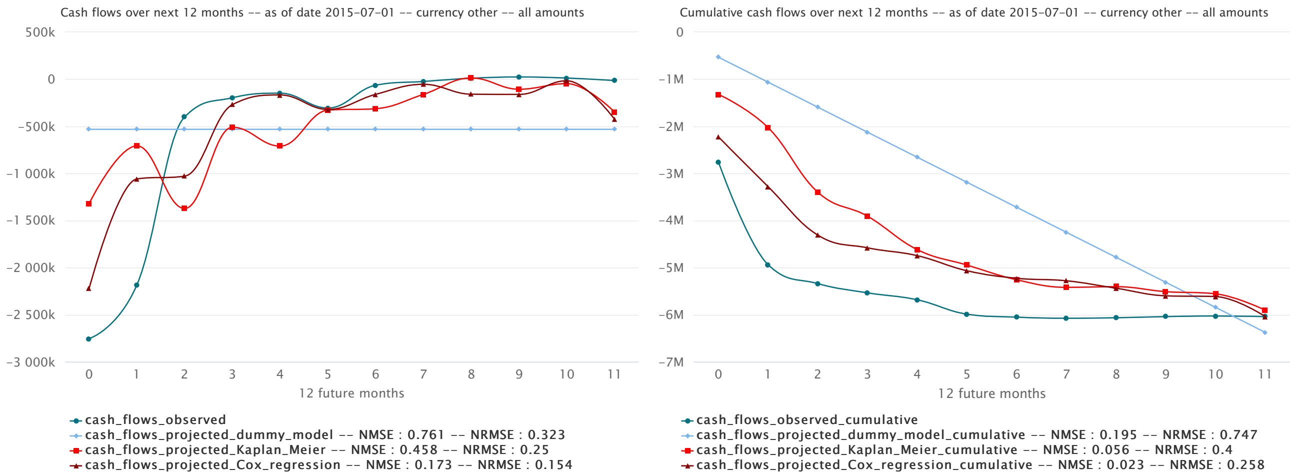


Figure 59: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2015-07-01 – currency other – all amounts

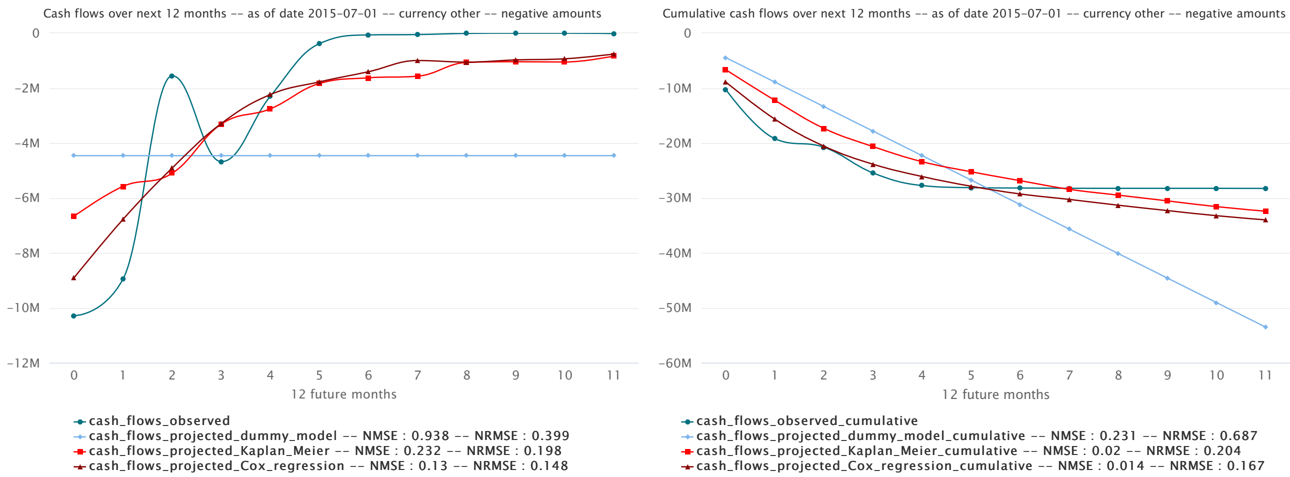


Figure 60: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2015-07-01 – currency other – negative amounts

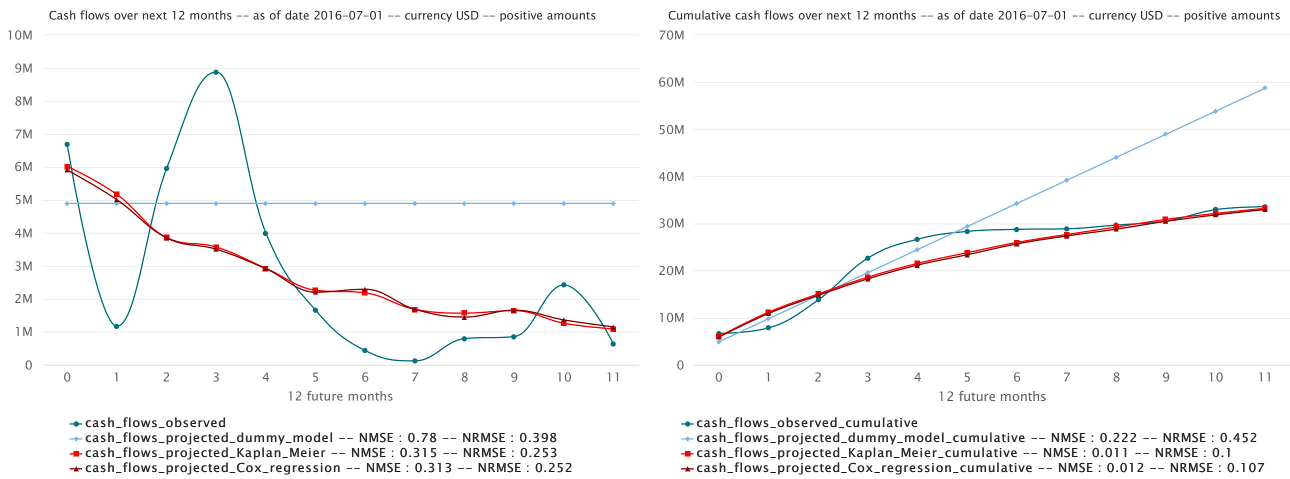


Figure 61: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2016-07-01 – currency USD – positive amounts

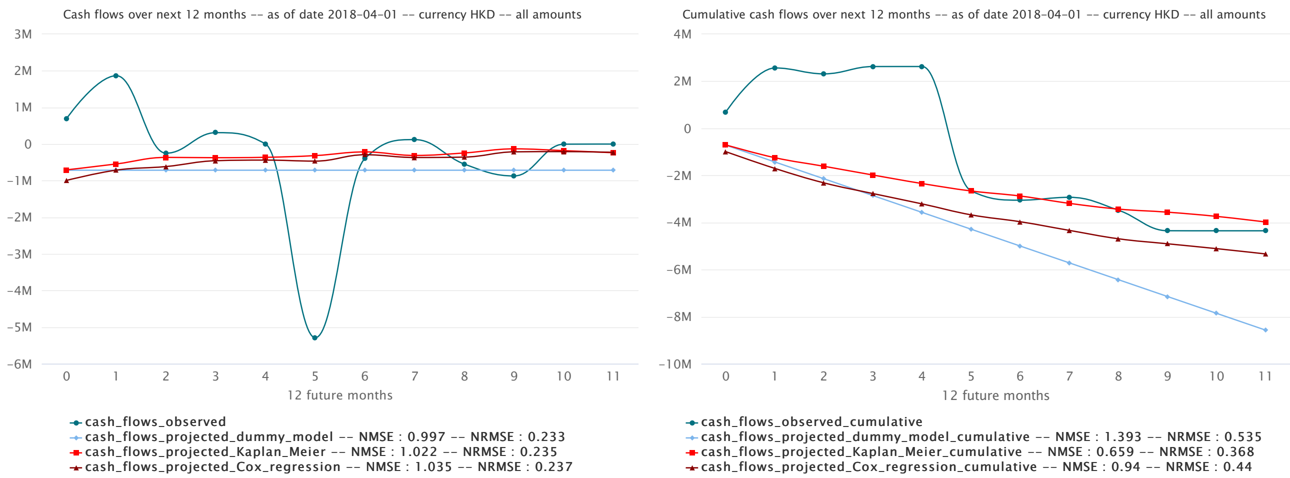


Figure 62: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2018-04-01 – currency HKD – all amounts

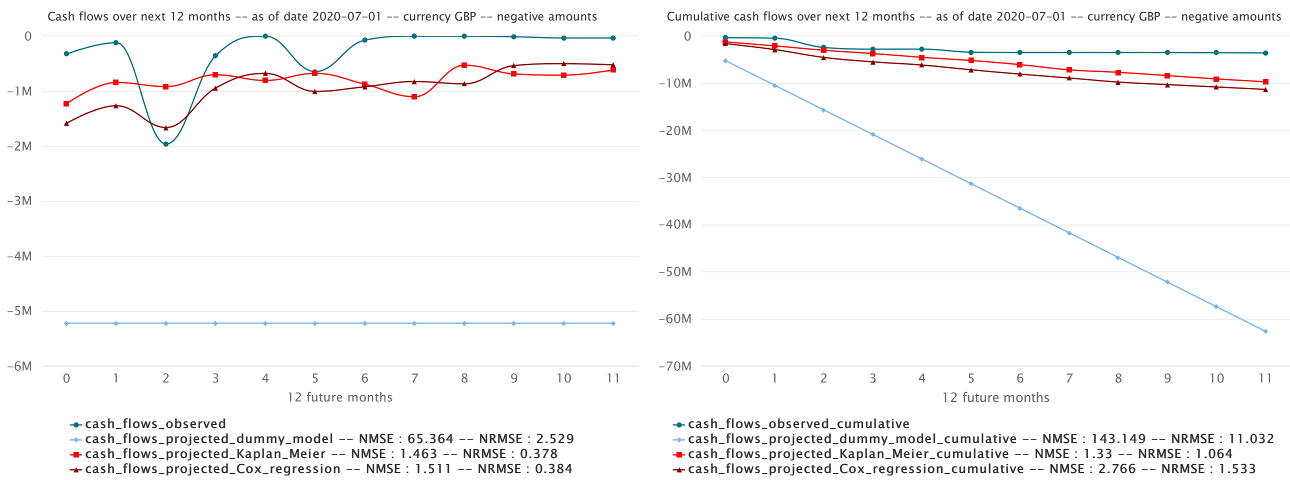


Figure 63: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2020-07-01 – currency GBP – negative amounts

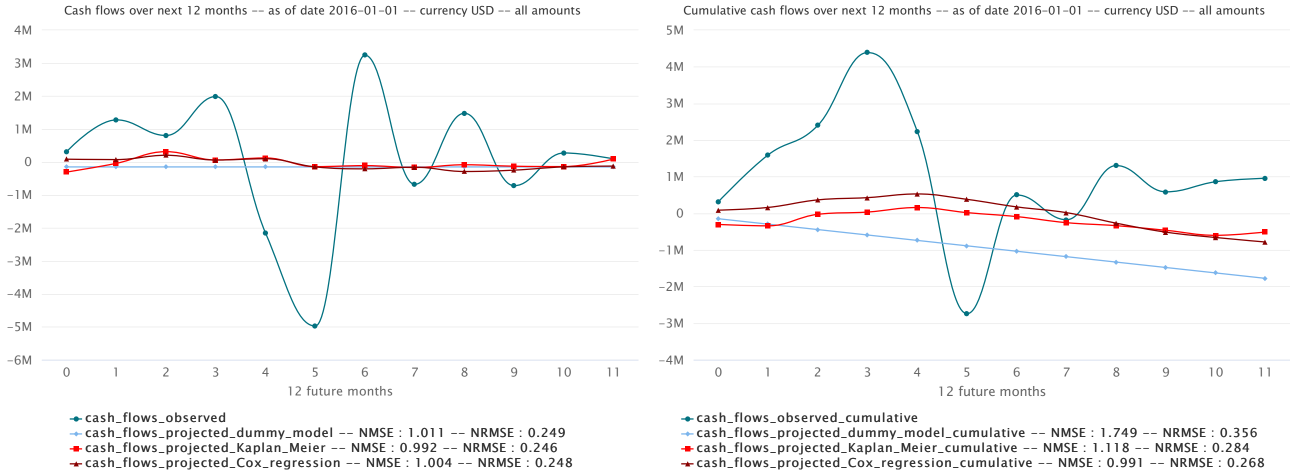


Figure 64: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2016-01-01 – currency USD – all amounts

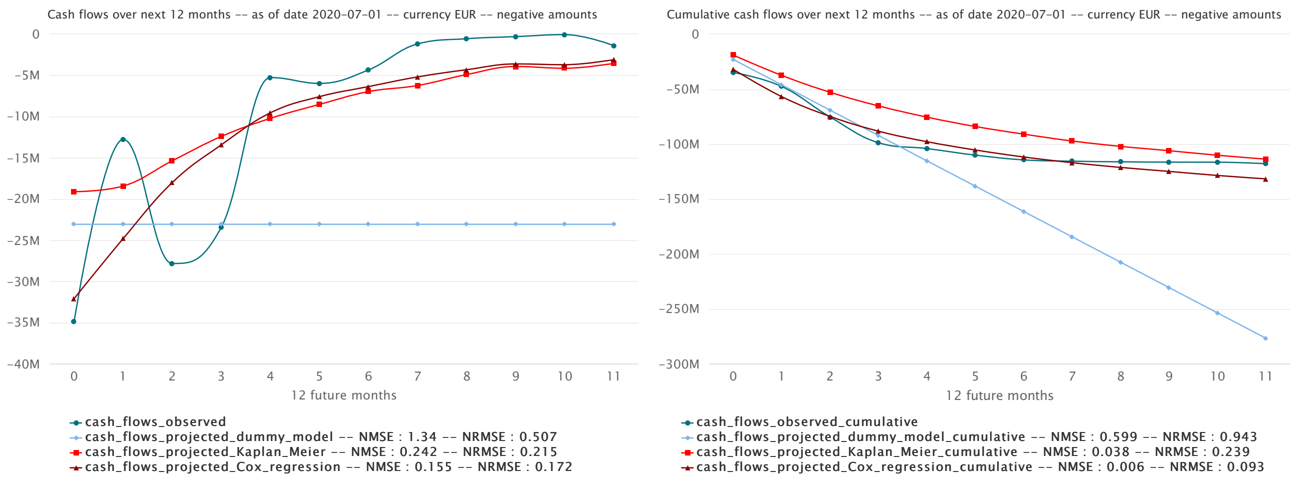


Figure 65: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2020-07-01 – currency EUR – negative amounts

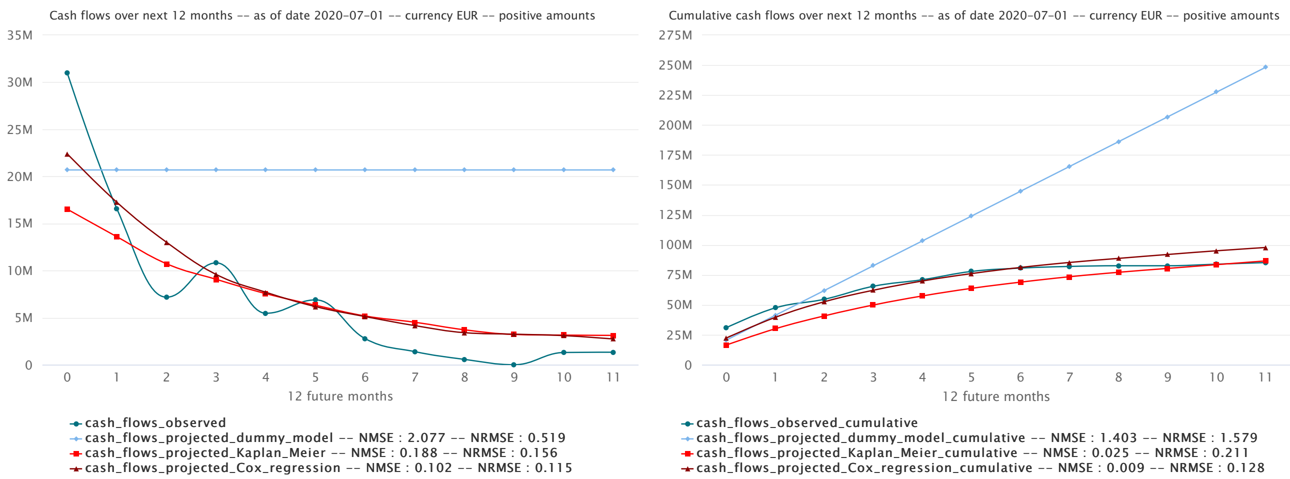


Figure 66: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2020-07-01 – currency EUR – positive amounts

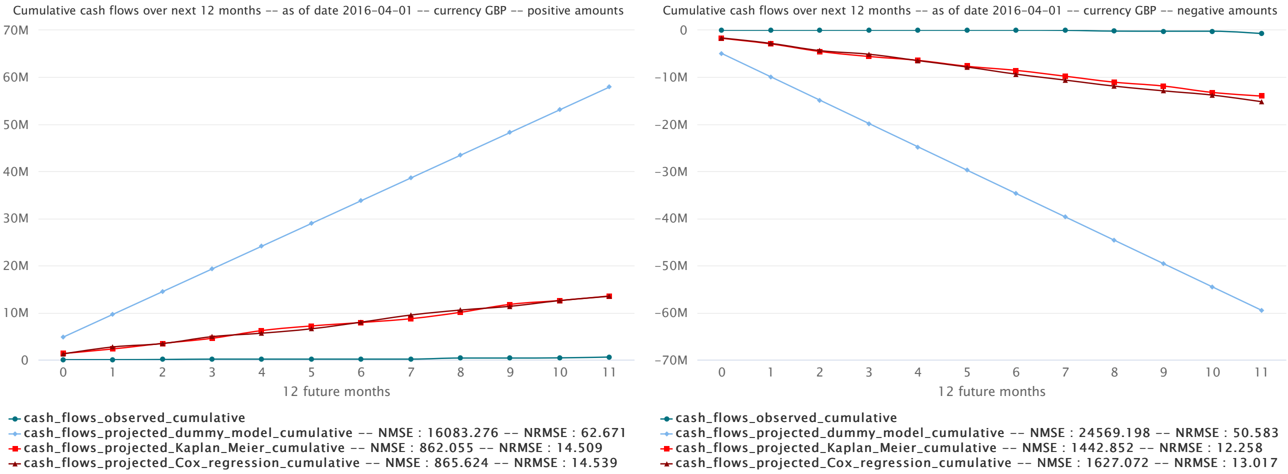


Figure 67: Cumulative cash flows over next 12 months – as of date 2016-04-01 – currency GBP – positive (left) and negative (right) amounts

We will not comment every figures and cash flows curves individually. Nonetheless, let us make the following general observations, that will serve as a conclusion to this chapter:

- We note erratic fluctuations in the patterns of observed individual cash flows, mostly in non euro currencies. We could assume (and maybe test) that the least represented is the currency among EBNYS balances, the more turbulent is the development of corresponding individual transactions with possible occurrences of extreme fluctuations corresponding to one consequent amount (in absolute value) being settled and not being compensated by many other settlements. In other words, the more EBNYS balances there is in a currency, the higher the chance for big settlements to not become preponderant in a given aggregate monthly cash flow, and the smoother the curve describing the twelve future monthly cash flows in that currency will be;

- We remark generally steadier patterns for positive-amounts-only and negative-amounts-only EBNYS balances transformation into observed cumulative amounts of monthly cash transactions;

- The curves representing both actually observed individual and cumulative monthly amounts of cash transactions related to all EBNYS balances, whatever the sign of their value, manifest more and steeper oscillations and shifts over the the twelve months following the date of reference considered;

- Concerning the closeness of predicted movements of cash in respect of real ones, we see that for positive-amounts-only and negative-amounts-only EBNYS balances, the patterns of cash flows issued from the predictions of the duration models are indeed closer in general to the observed ones than those of the dummy model;

- We furthermore notice that both Kaplan-Meier and Cox regression models (and of course the dummy model) do not capture brisk and large changes and fluctuations in the developments of monthly cash flows, either individual or cumulative. This explains partly why performance metrics reveal poor predicting performance of those models regarding the scopes of all balances amounts and under represented currencies, more incline to be characterized by erratic movements of monthly cash transactions;

- The difficulty to predict cash flows issued from all EBNYS balances (with positive and negative amounts), resides in the fact that both positive and negative balances and their subsequent settlements tend to neutralize each other which results in close to zero and highly volatile amounts of aggregated monthly cash flows (see for example Figure 64). Those oscillations of real monthly transactions of cash around small absolute values in comparison with the absolute values of original EBNYS balance (and the fact that the duration models monthly predictions are averages stemming from a Monte Carlo procedure) most of the time lead to equivalent averaging curves between the dummy model and the duration ones. This is consistent with the error measurements analogous numerical results between the models for balances of all amounts;

- From a general point of view, Kaplan-Meier model seems to perform as well as Cox proportional hazards model on our data. This could be explained by a bias-variance trade-off: Kaplan-Meier model a priori has high bias and low variance whereas Cox regression model has lower bias but higher variance (low/high bias suggests less/more assumptions about the form of the target function and low/high variance suggests small/high changes to the estimate of the target function with changes to the training dataset);

- One last observation corroborating the huge average values of MAPE (and NMSE) encountered in previous results tables (Tables 20 to 25): in many examples of predicted versus observed cash flows, we find that forecast monthly amounts are larger (in absolute value) than observed ones (Figure 67 serves as an extreme instance for illustrative purposes). In addition, we saw that the mean absolute percentage error, as a measure of forecast accuracy, treats negative and positive forecast errors differently (higher weigh on negative normalized errors, i.e. on situations where forecast quantities have larger absolute values than observed ones). This asymmetry of MAPE together with particular single cases of higher forecast amounts bring about extremely high values of individual MAPE, disrupting its average value over several dates of reference.

6 Technical cash flows prediction from not yet constituted balances

In subsection 5.1.3, we saw that numerous balances were settled rapidly after their writing date. In other words, a significant part of next-12-month cash flows does not come from EBNYS balances, as it would be expected.

A more direct and convincing way to see that EBNYS balances do not explain all future cash flows within a year, is to look at the as-of-date d ratios

$$\left(\frac{\text{Settlements aggregate amount (in euro) within 12 months coming from not yet existing balances at } d}{\text{Settlements aggregate amount (in euro) within 12 months}}\right)_d$$

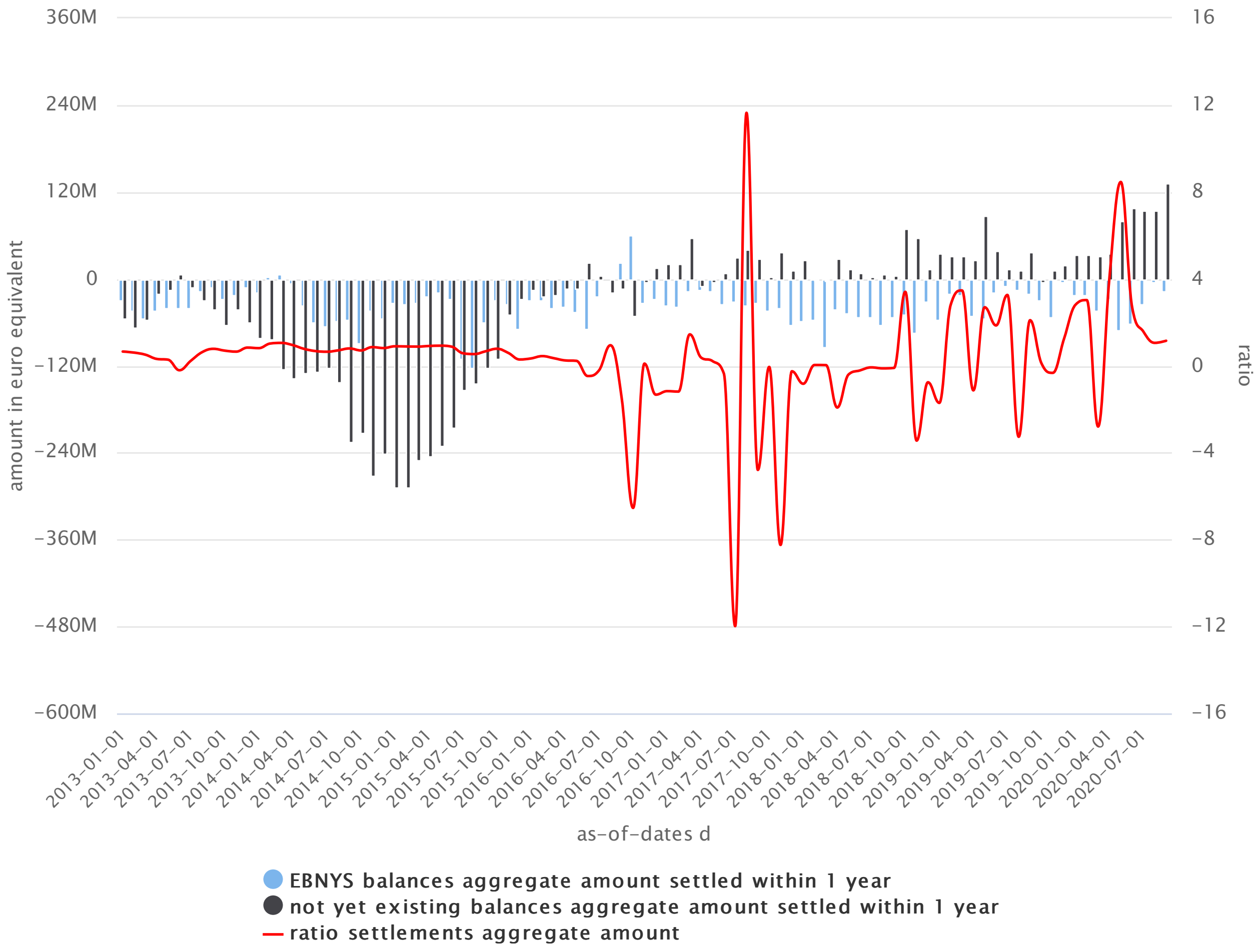


Figure 68: Ratio settlements within 12 months aggregate amount (in euro) - EBNYS vs non existing balances as of date

We can observe that this ratio is non-negligible most of the time, and can take extreme values corresponding to situations where, within 12 months, the settlements aggregate amount issued from not yet existing balances is of the same order as the settlements aggregate amount issued from EBNYS balances, but of opposite sign.

As a consequence, we chose to break down the issue of anticipating near future treasury needs in currencies, into one hypothesis and three distinct and complementary axes:

Hypothesis: Future cash flows within the next twelve months stem from either existing but not yet settled (EBNYS) balances, or existing technical reserves, or last first notice and evaluations of claims and cash calls (not yet materialized as reserves nor considered in AGRe's accounting system).

Sub-goals:

1. Prediction of settlements issued from EBNYS technical balances;
2. Prediction of not yet existing technical balances settlements issued from existing technical reserves;
3. Inclusion of first notices of losses (recent claims and cash calls evaluations) liable to be settled within next few months but not yet contemplated in AGRe's accounting and financial systems.

The hypothesis we endorse is a strong assumption and is clearly a simplification of reality that focuses only on the most important features of possible future cash flows sources. It serves mostly to restrain and define the outlines of our attempt at forecasting near future cash requirements in currencies, and it won't be tested.

The first goal was achieved partially in previous chapter through duration models calibrated on the basis of AGRe historical balances. We will try to reach the second one in this chapter, or at least offer original proposals and delineate an explicit procedures to that end: for lack of reliable historical data, numerical results won't be displayed. The third goal is more qualitative and efforts towards its accomplishment will not be laid out in this thesis. For now, and for lack of data as well as for lack of a better method (or model), first notice of claims and cash calls will be integrated "manually", in conformity with experts judgments, in the predictions made by the two previous courses of action.

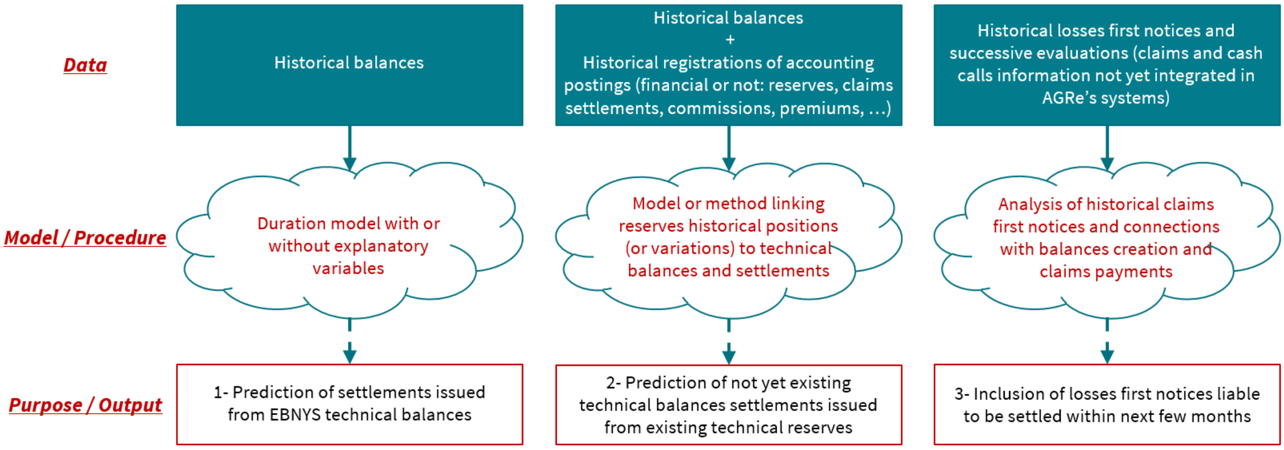


Figure 69: Short-term cash flows prediction goal breakdown - models structure and data sources

Let us then have a try at the second approach, an inevitable and complementary step on the road to the forecast of future monthly technical transactions.

6.1 Source of information and data available

6.1.1 Data description

We have at our disposal a database prototype keeping track of all historical (in theory from 1990 to 2021) records of financial and non-financial statements for each reinsurance treaty's section x third party business ID. That is to say, each row of this database corresponds to a given historical account entry (whether a premium or a loss payment, a cash call, a change in the stock of a particular reserve type, a cash deposit, a commission, a refund, etc.) related to a specific third party of AGRe and a specific treaty's section.

It is worth noticing that the historical net balances reviewed at length in previous chapter are, for most of them, composed of several of this current database financial entries. A given historical balance amount could be the result of a netting process between various financial entries of different natures (adjustment premiums + losses for example). The financial part of this new database can be seen as an decomposition of historical balances, distinguishing their underlying financial elements according to their qualities.

The main fields of that historical-statements-record-keeping database are:

- variables linked to the reinsurance treaty x section (reference, name, section number, ...);
- variables linked to the cedent entity (business ID, name, ...);
- variables linked to the retrocession entity (business ID, name, ...);
- type of posting, i.e. nature of the registering (specific kind of premium or loss settlement, variation in the stock of a particular sort of provision, cash call, cash deposit, commission, ...);
- statement's reference currency;
- date of registration / line writing;
- amount valued in original currency as well as in euro equivalent according to FX rates at the time of the posting entry.

Basically, every time there is a variation in the stock of a particular type of reserves for a given reinsurance treaty's section x third party (cedent or retrocessionaire), an entry (a posting) is created in this database, informing of the new reserves stock, its nature, and the date of this movement. Similarly, every time a particular type of premium, a commission, a fee, or a claim is paid under a given treaty's section x third party, a posting is generated, informing of the amount, the nature and the date of the transaction.

Work is still under progress with AGRe's data department to complete, consolidate and ensure the reliability of this historical database. A great deal of efforts have been spent, and are even now ongoing, in helping to reconcile those historical postings data with historical accounting balances together with closures' contractual and accounting data. Links between those databases, for a given scope, have to be as clean, as transparent and as coherent as possible before applying a model or a methodology aiming at predicting future monthly cash flows deriving from technical reserves past positions or variations.

The consistency of those historical entries amounts with respect to other reference and mature AGRe's source of data is indeed of paramount importance, and is now a key issue to serenely employ and test the recipes and techniques we are now going to put forward.

Let us imagine from now on that we secured the soundness and coherence of the historical-statements-record-keeping registered postings and that they cover correctly and thoroughly the required scope. We first have to identify and select the relevant types of posting we needed in our quest to connect technical reserves to near future cash flows that do not relate to EBNYS balances. That being done, we have to devise a way of reorganizing and structuring those data to be able to extract fruitful information either through simple computations or through model training algorithms.

6.1.2 Data structuring

As a reminder, we seek to predict next-12-month technical cash flows that are not issued from EBNYS balances, but that are instead issued from existing technical provisions as of date. A seemingly obvious and proficient manner to discover relationships between technical reserves amounts at a given date and following cash flows, is to examine at different historical moments how prior variations (or stock levels) of reserves are connected to subsequent balances writings and settlements.

The data structuring process will be sensibly the same whatever information extracting strategy will be used:

1. Choose a time interval between as-of-dates pictures of historical observations, i.e. choose historical dates d_1, d_2, ..., d_n evenly spaced (each one separated by a month or a quarter for instance);

2. For each reinsurance treaty's individual section, each related third party and each date d_i, consider the evolution of past (with respect to d_i) reserves registering, along with future (with respect to d_i) settlements that are not associated with as-of-date d_i EBNYS balances;

3. For each combination of reinsurance treaty's section x third party x as-of-date d_i, build a row rearranging and summarizing previous step pieces of information into amounts encapsulated into specified periods surrounding d_i: for example a row displaying the last technical reserves stocks in period [d_i - 4 years, d_i - 3 years[, followed by last technical reserves stocks in periods [d_i - 3 years, d_i - 2 years[, [d_i - 2 years, d_i - 3 semesters[, [d_i - 3 semesters, d_i - 2 semesters[, [d_i - 2 semesters, d_i - 1 semester[and [d_i - 1 semester, d_i[, and then the amounts of settlements (not linked to EBNYS balances as of d_i) aggregated over the periods [d_i, d_i + 1 quarter[, [d_i + 1 quarter, d_i + 2 quarters[, [d_i + 2 quarters, d_i + 3 quarters[, [d_i + 3 quarters, d_i + 4 quarters[. Independently of the date d_i we place ourselves at, we will designate those periods more simply as Y_{-4}, Y_{-3}, S_{-4}, S_{-3}, S_{-2}, S_{-1}, and Q_1, Q_2, Q_3, Q_4 respectively.

Let us detail a bit the algorithm behind this structuring procedure of historical statements and balances data. For each as-of-date d_i:

1. Compute the delta in days (and then in months) on one hand between d_i and the entry dates of technical-reserves-types postings in the historical-statements-record-keeping database, and on the other hand between d_i and the settlements dates of balances that have been written after d_i;

2. Create a categorical variable classifying all of those delta into a set of predetermined periods preceding and succeeding d_i, for instance: Y_{-max}, Y_{-8},...,Y_{-4}, Y_{-3}, S_{-4}, S_{-3}, S_{-2}, S_{-1} for negative delta, and Q_1, Q_2, Q_3, Q_4, Y_+ for positive delta. One could slice the past in quarters instead of semesters and years, and not dive as far into the past (limit oneself to the 4 last years and not go beyond). Likewise, one could slice the future in monthly periods instead of quarterly;

3. Once every historical balances written after d_i and every historical technical reserves postings have been categorized into a given past or future period relatively to d_i, retain in the scope only the reserves postings classified in a past period and the balances categorized as Q_1, Q_2, Q_3, and Q_4 (since we are only interested here in a 1-year horizon cash flows forecast);

4. For each treaty x section x third party (cedent or reinsurer) and for each past period, keep the last entry of each type of technical reserves considered, and sum them;

5. For each treaty x section x third party (cedent or reinsurer), sum all the related settlements amounts (from not yet existing balances as of d_i) by future period class inside which they are assigned;

6. Concatenate previously obtained past periods' reserves amounts with future periods' settlements amounts, and this by treaty x section x third party (and implicitly x currency);

7. As a last optional step, one can decide to transform sequences of past reserves stocks into sequences of past reserves variations from one period to the next.

Repeat this step-by-step procedure for all the chosen historical as-of-dates d_i's, thus building a newly structured data set having the following aspect:

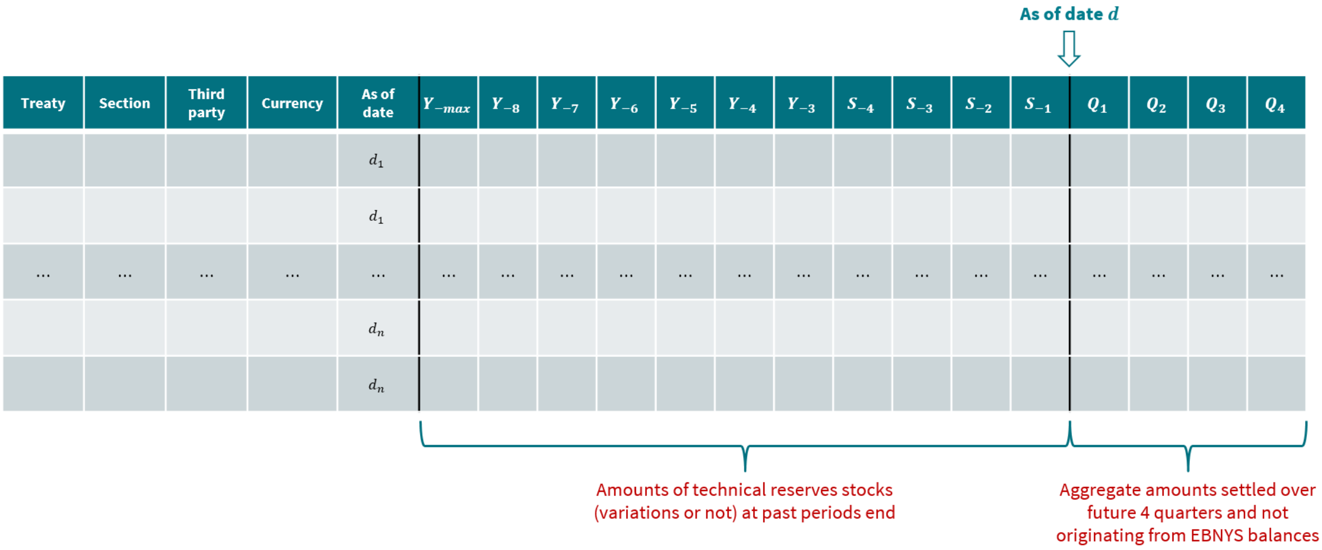


Figure 70: Data structuring to connect past variations or stocks of technical reserves to future settlements from not yet existing balances

Thanks to the treaty and section numbers together with the business ID of the third party, we can add to this data set as much technical and contractual variables as we wish (reserving segment i.e. LR scope, perimeter, proportional or non proportional, type of reinsurance treaty, underwriting year, etc).

6.2 Approach suggestions to forecast near future cash flows stemming from not yet existing balances

We offer in this subsection different possible approaches to extract useful information from the newly structured database whose construction process has just been drawn up. Q_1, Q_2, Q_3, and Q_4 can be seen as target variables possibly dependent, while other fields (past reserves evolution as well as attached technical and contractual information) can be used in this context as potential explanatory variables.

Once more, it is worth emphasizing that the time periods shaping the above derived database are flexible and constitute hyperparameters to be played with or to define wisely on the basis of expert intuition for example.

In the case we achieve to build a model or a method able to predict accurately aggregate amounts settled at Q_1, Q_2, Q_3, and Q_4, we would then split each of those 4 predicted amounts equally in 3 sub-components corresponding to monthly cash flows within the considered future quarter. We would finally aggregate those outputs monthly cash flows by currency and add them to the ones predicted from EBNYS balances.

A problem arising when structuring the historical postings this way, at the finest possible grain level that is treaty x section x third party, is the scarcity of meaningful observations. To express it differently, most of the rows of the newly built database only hold zero values (i.e. for a given treaty x section x third party and a given reference date, there is no past reserves and no near future settlements to be linked together). We can either remove those barren and fruitless observations or gather observations by summing them at a higher aggregation level, the LR SCOPE x CURRENCY level for instance:

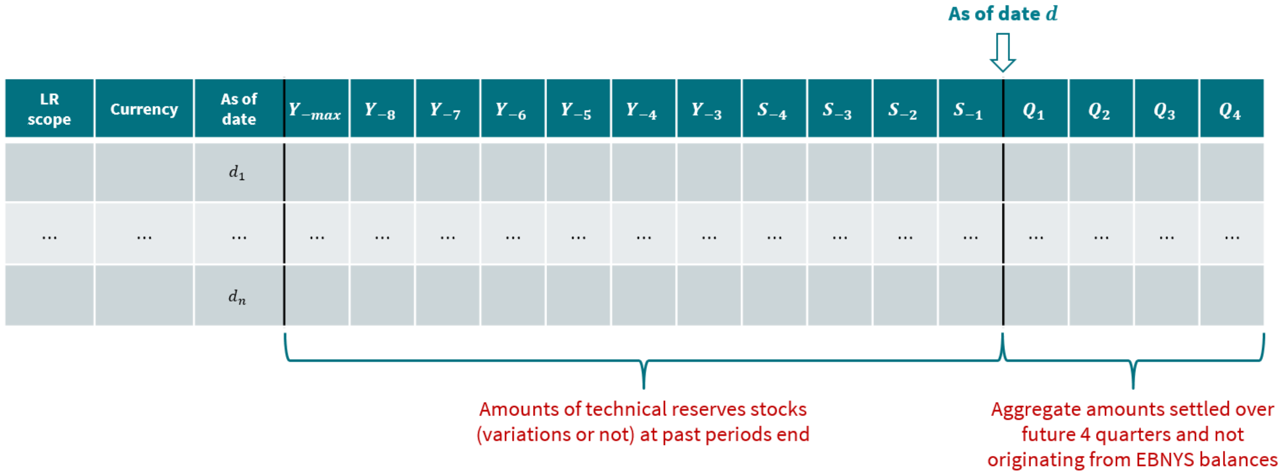


Figure 71: Data structuring to connect past variations or stocks of technical reserves to future settlements from not yet existing balances - data aggregation at LR SCOPE X CURRENCY level

Descriptive analysis on the structured data, such as the study of the Pearson as well as the Spearman Rank correlation matrices, can be performed.

R codes embryo implementing most of the subsequent methods and models on the data at hand, have been developed. Nevertheless, for lack of reliable data up to now, we will not present output results. We will deliberately remain general in the description of said models and methodologies, since no precise and optimal way towards the stated goal of this chapter has been determined.

6.2.1 General linear model, GLM and multiple response variables GLM

General linear model

The general linear model or general multivariate regression model is a generalization of multiple linear regression to the case of more than one dependent response variable, and can be written $\mathbf{Y} = \mathbf{XB} + \mathbf{U}$, where,

- Y is a matrix with series of multivariate observations (each column being a set of measurements on one of the dependent variables): in our case Y would be equal to the columns Q_1, Q_2, Q_3, and Q_4;
- X is a matrix of observations on independent variables: if reserves variations (at a aggregate level) from one period to the next were to be proven independent, then one could consider that X is equal to those past periods variations coupled with the categorical variables carrying technical and contractual information;
- B is a matrix containing the parameters to be estimated;
- U is a matrix containing errors assumed to be uncorrelated across measurements and to follow a multivariate normal distribution.

Given n observations of m response variables and p explanatory variables, the general multivariate linear regression yields n x m equations:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{i1} + \beta_{2j}X_{i2} + \ldots + \beta_{pj}X_{ip} + \epsilon_{ij}$$

with Y_{ij} the ith observation of the jth response (or dependent) variable, X_{ij} the ith observation of the jth possibly explanatory (or independent) variable, β_{ij} the parameters to be estimated, and ϵ_{ij} the independent identically distributed normal errors.

This amounts simply to a sequence of standard multiple linear regressions using the same predictors, and would translate to our current data as the following model composed of n systems of 4 equations:

$$\begin{cases} Q_{1,i} = & \beta_{0,1} + \beta_{1,1} \cdot Y_{-max,i} + \beta_{2,1} \cdot Y_{-8,i} + \beta_{3,1} \cdot Y_{-7,i} + \ldots + \beta_{10,1} \cdot S_{-2,i} + \beta_{11,1} \cdot S_{-1,i} + \beta_{\ldots,1} \cdot \text{technical and contractual variables}_i + \epsilon_{i,1} \\ Q_{2,i} = & \beta_{0,2} + \beta_{1,2} \cdot Y_{-max,i} + \beta_{2,2} \cdot Y_{-8,i} + \beta_{3,2} \cdot Y_{-7,i} + \ldots + \beta_{10,2} \cdot S_{-2,i} + \beta_{11,2} \cdot S_{-1,i} + \beta_{\ldots,2} \cdot \text{technical and contractual variables}_i + \epsilon_{i,2} \\ Q_{3,i} = & \beta_{0,3} + \beta_{1,3} \cdot Y_{-max,i} + \beta_{2,3} \cdot Y_{-8,i} + \beta_{3,3} \cdot Y_{-7,i} + \ldots + \beta_{10,3} \cdot S_{-2,i} + \beta_{11,3} \cdot S_{-1,i} + \beta_{\ldots,3} \cdot \text{technical and contractual variables}_i + \epsilon_{i,3} \\ Q_{4,i} = & \beta_{0,4} + \beta_{1,4} \cdot Y_{-max,i} + \beta_{2,4} \cdot Y_{-8,i} + \beta_{3,4} \cdot Y_{-7,i} + \ldots + \beta_{10,4} \cdot S_{-2,i} + \beta_{11,4} \cdot S_{-1,i} + \beta_{\ldots,4} \cdot \text{technical and contractual variables}_i + \epsilon_{i,4} \end{cases}$$

The general linear model is a commonly used statistical method to relate some number of continuous and/or categorical predictors to a single outcome variable and constitutes a special case of the GLM models, since it assumes that the distribution of the residuals follow a conditionally normal distribution, while the GLM loosens this assumption and allows for a variety of other distributions from the exponential family for the residuals.

Before applying this model, one would have to check that the assumptions are met, that is independent observations, normally distributed errors with mean of 0 and a constant variance (equal population variances), and predictor variables independent of one another (or at least uncorrelated). All these assumptions are not liable to be satisfied by our data.

Generalized linear model (GLM)

A rapid overview of classic generalized linear models has already been presented in subsection 5.2.2. As a reminder, outcomes of the response variables Q_1, Q_2, Q_3, and Q_4 are assumed to be generated from a particular distribution in the exponential family (the same for Q_1, Q_2, Q_3, and Q_4, or not). In this framework, one could consider Q1 as a potential explanatory variable to response Q2. Likewise, we could include Q1 and Q2 in the predictors of Q3, and Q1, Q2, Q3 in those of Q4. We would then have the model:

$$\begin{aligned} \mathrm{E}(\mathbf{Q1} \mid \mathbf{X_1}) &= g_1^{-1}(\boldsymbol{\beta}_1^{\intercal}\mathbf{X_1}) \\ \mathrm{E}(\mathbf{Q2} \mid \mathbf{X_2}) &= g_2^{-1}(\boldsymbol{\beta}_2^{\intercal}\mathbf{X_2}) \\ \mathrm{E}(\mathbf{Q3} \mid \mathbf{X_3}) &= g_3^{-1}(\boldsymbol{\beta}_3^{\intercal}\mathbf{X_3}) \\ \mathrm{E}(\mathbf{Q4} \mid \mathbf{X_4}) &= g_4^{-1}(\boldsymbol{\beta}_4^{\intercal}\mathbf{X_4}) \end{aligned}$$

with

$$\mathbf{X_1} = \begin{bmatrix} S_{-1} \\ S_{-2} \\ S_{-3} \\ S_{-4} \\ Y_{-3} \\ Y_{-4} \\ Y_{-5} \\ Y_{-6} \\ Y_{-7} \\ Y_{-8} \\ Y_{-max} \\ LRscope \\ Currency \\ Treatytype \\ \vdots \end{bmatrix}, \mathbf{X_2} = \begin{bmatrix} Q_1 \\ S_{-1} \\ S_{-2} \\ S_{-3} \\ S_{-4} \\ Y_{-3} \\ Y_{-4} \\ Y_{-5} \\ Y_{-6} \\ Y_{-7} \\ Y_{-8} \\ Y_{-max} \\ LRscope \\ Currency \\ Treatytype \\ \vdots \end{bmatrix}, \mathbf{X_3} = \begin{bmatrix} Q_2 \\ Q_1 \\ S_{-1} \\ S_{-2} \\ S_{-3} \\ S_{-4} \\ Y_{-3} \\ Y_{-4} \\ Y_{-5} \\ Y_{-6} \\ Y_{-7} \\ Y_{-8} \\ Y_{-max} \\ LRscope \\ Currency \\ Treatytype \\ \vdots \end{bmatrix}, \mathbf{X_4} = \begin{bmatrix} Q_3 \\ Q_2 \\ Q_1 \\ S_{-1} \\ S_{-2} \\ S_{-3} \\ S_{-4} \\ Y_{-3} \\ Y_{-4} \\ Y_{-5} \\ Y_{-6} \\ Y_{-7} \\ Y_{-8} \\ Y_{-max} \\ LRscope \\ Currency \\ Treatytype \\ \vdots \end{bmatrix}$$

The link functions g_i's could be equal or not.

As in the general linear model, it is crucial to check beforehand the GLM models assumptions concerning the data structure:

- Independence of each data points, i.e. of observations;
- Correct distribution and independence (at least non correlation) of residuals;
- Correct specification of the variance structure;
- Linear relationship between the response and the linear predictor.

We did not verify if those hypothesis were met on the current data set at our disposal, but intuitively the independence of observations and linear relationship between the response and the linear predictor may not be satisfied.

Multivariate Covariance Generalized Linear Models

Another road we did not explore is the multiple response variables GLM or Multivariate Covariance Generalized Linear Models developed in [10] and with a dedicated R package for implementation (see [11]). The multivariate covariance generalized linear models (McGLMs) is a general framework for non-normal multivariate data analysis, designed to handle multivariate response variables, along with a wide range of temporal and spatial correlation structures.

6.2.2 Ratios method, time series and curve fitting

A much simpler method than the others presented in this section is the use of historical ratios to extrapolate future aggregated settlements (Q_1, Q_2, Q_3, and Q_4) originating from not yet constituted balances, on the basis of the last technical reserves positions at date. From the newly obtained data structure of Figure 70 (or Figure 71 depending on the aggregation level we wish), we would only consider technical and contractual variables as well as the past period variable S_{-1}, which corresponds to the actual technical reserves put aside as of date.

The ratios $r_1 = \frac{Q_1}{S_{-1}}$, $r_2 = \frac{Q_2}{S_{-1}}$, $r_3 = \frac{Q_3}{S_{-1}}$, and $r_4 = \frac{Q_4}{S_{-1}}$ would then be computed for each row of the database in question, which is to say for each date of reference d_i and each chosen level of data aggregation (treaty x section x third party or LR scope x currency for instances). We would then analyse the movements of those ratios as a function of time, their average and variance.

To take a concrete an example, one could calculate those ratios taking the 1st day of each month between 2015 and 2020 as dates of reference to get as many historical slices of the past, and consider (in addition to the currency dimension) a short tail vs long tail development separation of reinsurance treaties for aggregation purpose of observations. That would give us two ratios (one for all short tail treaties observations and one for all long tail treaties observations) for each currency (or at least the major ones), for the 60 dates under consideration and for the 4 future quarters. We would normally see higher values for short tail ratios in comparison with long tail ones. Tendencies in the paths taken by those two ratios over time could also be observed. One could then consider the mean of the values taken by those ratios over time. We would thus have two average ratios for each currency, to be made use of in order to estimate future quarterly cash flows in currencies from not yet existing balances, and this on the basis of existing technical reserves aggregated by currencies along short tail or long tail characteristics of treaties. Another approach, a bit more complex, could be to handle the above time function ratios in a time series analysis framework, to predict relevant future ratios.

The ratios estimates (average over a given period or time series prediction) are then multiplied to current stocks of reserves leading to estimates of future aggregated settlements Q_1, Q_2, Q_3, and Q_4, as such:

$$\widehat{Q_1} = \widehat{r_1} \cdot S_{-1}$$
$$\widehat{Q_2} = \widehat{r_2} \cdot S_{-1}$$
$$\widehat{Q_3} = \widehat{r_3} \cdot S_{-1}$$
$$\widehat{Q_4} = \widehat{r_4} \cdot S_{-1}$$

Instead of splitting each $\widehat{Q_i}$ in 3 equal parts to obtain estimated monthly cash flows, we could contemplate directly the ratios $\left(\frac{M_i}{S_{-1}}\right)_{i \in [\![1,12]\!]}$, with M_i the future monthly cash flows (as of the chosen historical dates). The aggregate effect of the quarterly consideration would be lost, but one could resort to the process of curve fitting, trying to find a mathematical function (from a particular family or not) that has the best fit to the series of those average ratios data points.

This ratio methodology has a serious drawbacks: it does not take into account the maturities of past technical reserves, but only their stock amounts at date. To put it another way, this model does not incorporate the chronology of the different time layers of reserves that have been piled up or removed up to date. This loss of information on reserves constitution times and on past movements can potentially hinder the performance and accuracy of such an approach.

In spite of its apparent simplicity and limitations, this ratio method could turn out to be the one retained at the end, if the other strategies laid out in this chapter do not achieve satisfactory results.

6.2.3 Long short-term memory (LSTM) artificial recurrent neural network

The LSTM algorithm, being explicitly designed and particularly suited to address sequences of data with long-term dependencies, appears at first glance well adapted to associate past variations of technical reserves with future cash flows. This, of course, remains to be proved, by training and testing different configurations of LSTM structures on reliable AGRe's historical data. Other machine learning techniques, such as the ensemble learning method "random forest", could also be implemented and compared against one another. Meanwhile, let us focus only on the seemingly propitious LSTM algorithm for the matter at hand.

Brief presentation of LSTM network

The following brief presentation of the LSTM network is designed to be a qualitative overview. We therefore won't delve into the technical peculiarities of this tool and the mathematics underlying it. A step-by-step walk through of information processing inside a LSTM unit is summarized in appendix E.

Long short-term memory (LSTM) has an artificial recurrent neural network (RNN) architecture, which differs from standard feedforward neural networks by having feedback connections enabling it to not only process single data points, but also entire sequences of data.

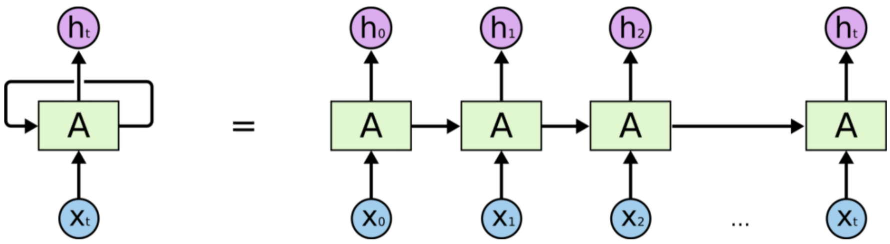


Figure 72: Unrolled recurrent neural network

A RNN network can be thought of as multiple copies of the same network, or loops (as displayed by Figure 72), each one transmitting a message to its successor, allowing information to be passed and persist from one step of the network to the next. Remembering information for long periods of time is what makes LSTM distinctive among RNNs. It is indeed able to connect distant prior information to present task/prediction, and is then adapted to applications such as:

- predicting the next word based on previous ones;
- predicting a given market asset future return as a function of past returns;
- predicting next period case reserves and paid claims from the sequence of past developments;
- speech recognition, rhythm and grammar learning, protein homology detection, . . . ;
- more generally: classifying, processing and making predictions based on time series data.

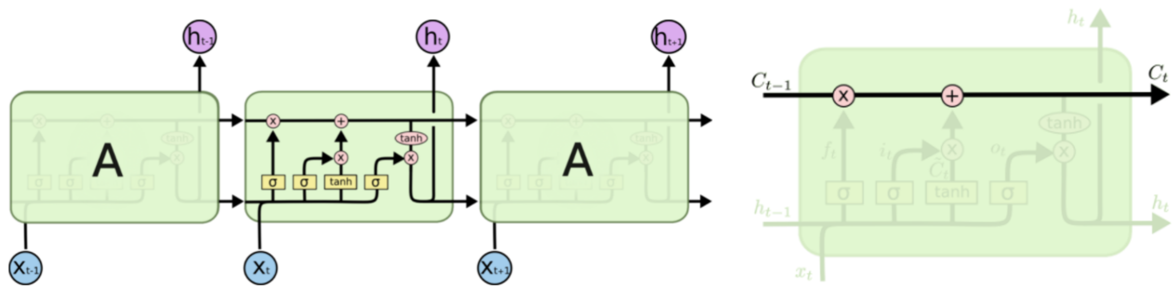


Figure 73: The repeating module in an LSTM and the cell state flow

As shown by Figure 73, LSTMs have a chain-like structure of repeating modules (units or memory cells), each one composed of four neural network layers, interacting in a very special way. A common LSTM unit carries a cell, an input gate, an output gate and a forget gate. The key to LSTMs singularity is the cell state (horizontal line running through the top of the diagram on the right hand side of the above figure) and the information flowing along it. The cell state acts as a conveyor belt that runs straight down the entire chain, undergoing only some minor linear interactions inside each unit. A given cell remembers values over arbitrary time intervals and its three gates regulate the flow of information into and out of it. To sum up, LSTM constitution does have the ability to remove or add information to the cell states, carefully regulated by structures called gates, that are a way to optionally let information through.

Implementation

Let us keep in mind, that the features here are past variations of technical reserves (over chosen past periods) plus (facultatively) carefully selected technical and contractual categorical variables, and the targets are amounts of cash flows (from not yet existing balances) aggregated over chosen future periods (future 4 quarters in the example illustrated by Figures 70 and 71).

When training a neural network, such as a Long Short-Term Memory recurrent neural network, the data for a given sequence prediction problem generally needs to be scaled, if not, large inputs can possibly slow down the learning and convergence of the network and in some cases prevent it from effectively learning the problem. Features values, as well as targets ones are commonly normalized or standardized across all observation points. This does not seem an appropriate scaling procedure for input and output variables in our case, since reserves and aggregated cash flows amounts are not directly comparable between risks or risk groups, and can differ by several orders of magnitude. A way around this complication is to divide each sequence of amounts (past reserves variations along with future aggregate cash flows) by its components' highest absolute value, storing the information of this scaling factor specific to each row of the observation database. With the exception of grafted technical and contractual variables, the proposed rescaling process would throw the time sequences' amounts from their original ranges within the range of 0 to 1.

The rescaled data set is then partitioned into two (or three) subsets, refered as the training set, (the validation set,) and the testing set. This splitting will be done in our special case on the basis of the as-of-date variable: samples linked to older as-of-dates will comprise the training set, the ones with intermediary as-of-dates will constitute the validation set, and finally the samples relative to most recent as-of-dates will be put in the testing set. The dates thresholds are to be determined according to common splitting rules (say for example to build subsets comprising 70% for training, 15% for validation, and 15% for testing, of the initial data set). All those subsets consist of pairs of an input vector (features) and the corresponding output vector (labels or targets).

The LSTM model parameters (weights and biases of connections between neurons among others) are initially fit on the training data set. The model is trained on the training data set using a supervised learning method, for example using optimization methods such as gradient descent or stochastic gradient descent. Running the LSTM with the training data set produces a result, which is compared with the target, and this for each input vector in the training data set. Based on the result of the comparison and the specific learning algorithm being used, the parameters of the model are adjusted. The fitted model is then used to predict the responses for the observation samples in the validation data set, which provides an unbiased evaluation of the model fit on the training data set, potentially used to tune the model's hyperparameters (number of hidden units and layers, optimal early stopping time, dropout coefficients, ...). Finally, the testing data set, never involved in the training of the LSTM network, is used to provide a final unbiased evaluation of the model performance.

The LSTM structure can be built with Keras deep learning API. Python is recommended, but this model can also be implemented on R thanks to a virtual environment and a specific foregoing step-by-step procedure. One of the tricky part in coding an LSTM architecture is that it requires three-dimensional inputs, whose dimensions are:

1. The sample: one sequence (i.e. one row in our data set) is one sample. A batch is comprised of one or more samples, for example ten rows of the data set displayed in Figure 71;

2. The time steps: one time step is one point of observation(s) in the sample. Keeping the same example, the time steps for the model inputs are Y_{-max}, ..., S_{-1}, and for the outputs Q_1, ..., Q_4.

3. The features: one feature is one observation at a given time step. There could be several observations values at each time step. In our particular case, there is only one feature at each time step, namely the rescaled reserves amount variations corresponding to past time steps for inputs or the rescaled cash flows amounts corresponding to future time steps for outputs.

This means that the LSTM input layer expects a 3D array of data when fitting the model and when making predictions, even if specific dimensions of the array contain a single value, like one sample or one feature. Hence, reshaping the training and testing data to enter LSTM keras sequential model is indispensable. We will then reshape our input data from [# rows, # past steps = 11 in this example] to [# rows, # past steps, 1], and output data from [# rows, # future steps = 4 in this example] to [# rows, # future steps, 1].

Below we show a possible and simple LSTM architecture suited to our current example, that is a LSTM composed of an input layer (dense layer) of 5 neurons, taking as input batches of 10 samples showing 1 feature over 11 time steps (Y_{-max}, ..., S_{-1}), followed by a LSTM layer made up of 5 hidden units, and finally an output layer (dense layer) formed of 4 neurons.

```
Layer (type)                 Output Shape              Param #
=================================================================
dense_1 (Dense)              (10, 11, 5)               10
_________________________________________________________________
lstm (LSTM)                  (10, 5)                   220
_________________________________________________________________
dense (Dense)                (10, 4)                   24
=================================================================
Total params: 254
Trainable params: 254
Non-trainable params: 0
_________________________________________________________________
```

Figure 74: LSTM possible architecture and parameters

Other architectures, with more hidden layers and more hidden neurons, can obviously be used. Likewise, different loss functions (mean_squared_error, mean_absolute_error, mean_squared_logarithmic_error, mean_absolute_percentage_error) and different activation functions (LeakyRelu, relu, elu, sigmoid, tanh, linear) can be tested, as well as different optimizer (Adam, RMSprop, Adamax, ...). Futhermore, a dropout regularization within the recurrent layer(s) can be applied. It involves a defined fraction of the input units being randomly dropped at each update during training time, both at the input gates and the recurrent connections, resulting in reduced risk of overfitting and better generalization. Early stopping can also be employed, through a customized callback function, as a further mechanism to prevent overfitting. Early stopping means that, after each epoch, the network predicts the unseen samples from the validation set, leading to the computation of a validation loss. Once the validation loss does not decrease by a predefined value for a certain number of patience periods, the training stops and the model weights that reached the lowest validation loss are restored.

One last issue, is the integration of technical and contractual categorical variables into the learning and predicting process of a LSTM neural network. There is several manners to incorporate auxiliary categorical features in time series data:

- Integer encoding: each unique class is mapped to an integer;
- One hot encoding: each class is mapped to a binary vector;
- Learned embedding: a distributed representation of the categories is learned.

Passing through one of those 3 methods, the categorical variables are encoded and preprocessed the same way as other temporal data. The time steps are not affected by this new data insertion, only the number of features is. One can either concatenate the newly encoded complementary categorical features with the output of the last RNN layer (post-RNN adjustment) or initialize the RNN states with a learned representation of the categories in question.

The ability to predict accurately Q_1, Q_2, Q_3, and Q_4 could be assessed and compared between different configurations of LSTM structure (number and kinds of layers and neurons, loss and activation functions types, dropout proportions, parameter search optimization algorithm i.e. optimizer, call back function, ...), and with other families of models presented in this chapter, thanks to the metrics introduced in subsection 5.4.3, namely the NMSE, NRMSE, NMAE, and MAPE.

6.2.4 To go further

One could be incited to deeper explore the LSTM neural network as a candidate model to link past variations (or levels) of predefined quantities, either technical (amounts of different types of reserves for example) or financial (amounts of past financial transactions classified by nature for example), to future cash flows in currencies.

In previous subsection we examined the simplest application of a LSTM since we took into consideration only one unique feature by time step, gathering at a given time period either various kinds of reserves (past time steps as input of the LSTM) or various sorts of cash transactions (future time steps as output of the LSTM) into a single amount observation. By aggregating different varieties of postings we would inevitably loose probable relevant pieces of information provided by the original database, namely the nature of each recorded statement.

There are dozens of such categories of entries nature, some of which could be considered as interesting features to insert in the LSTM input as predictors, and others to be predicted as constituents of future cash flows, comprising the LSTM output. Let us make use once again of Figure 71 for illustrative purposes, and consider inside each time step several observations of distinct quantities:

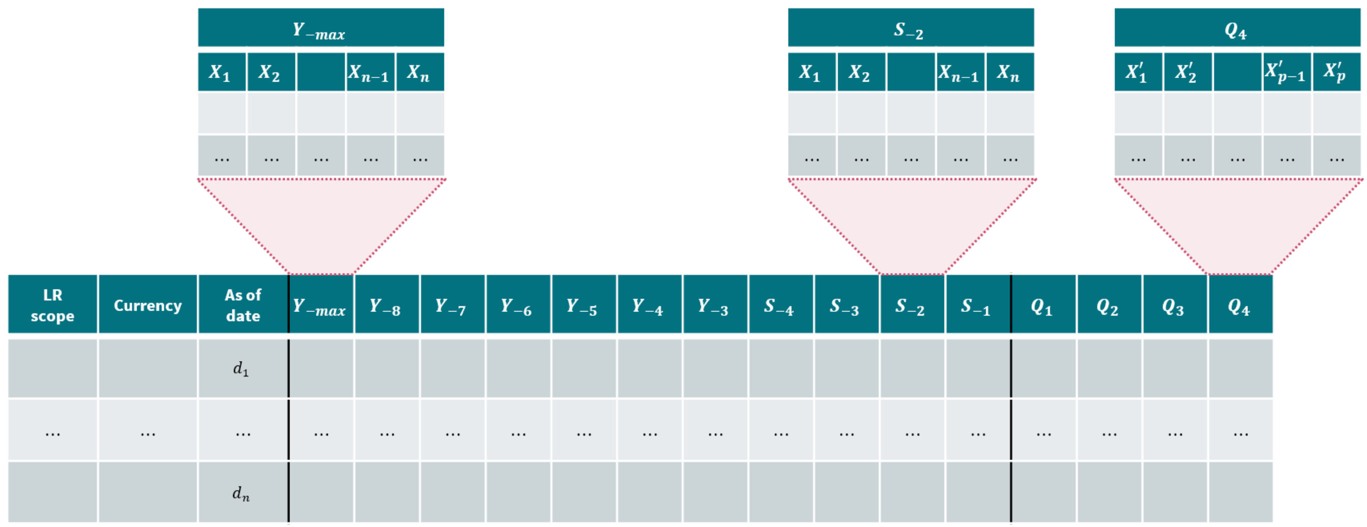


Figure 75: Data structuring to connect past variations or levels of relevant technical and financial quantities to future settlements - data aggregation at LR SCOPE X CURRENCY level - multiple features by time step

Instead of 1 feature observation value over 11 time steps (Y_{-max}, ..., S_{-1}), we have in this more general configuration, n features observations values (X_1, ..., X_n) over the same number of time steps. Likewise, instead of 1 target by future period (Q_1, ..., Q_4), we contemplate here p targets ($X_1^{'}$, ..., $X_p^{'}$). This time, the data entering the LSTM should have the 3 dimensions shape [# rows, # past steps, n], and the output data (to predict) the shape [# rows, # future steps, p].

The multiple quantities incorporated in this model as predictors variables, and whose values are observed over all past periods, could be:

- X_1: rescaled variation (or last stock level) of case reserves amount;
- X_2: rescaled variation (or last stock level) of IBNR reserves amount;
- X_3: case reserves refund rescaled aggregated amount over each past period;
- X_4: IBNR reserves refund rescaled aggregated amount over each past period;
- X_5: cash deposits (linked to premiums) rescaled aggregated amount over each past period;

- X_6: cash deposits (linked to claims) rescaled aggregated amount over each past period;
- X_7: reinsurance commissions rescaled aggregated amount over each past period;
- X_8: brokerage commissions rescaled aggregated amount over each past period;
- X_9: reinstatement premiums rescaled aggregated amount over each past period;
- X_{10}: paid losses rescaled aggregated amount over each past period;
- X_{11}: cash calls rescaled aggregated amount over each past period;
- ....

As for the quantities making up the LSTM output variables as well as the future monthly technical cash flows, and whose values are observed / to be predicted over all future periods, we could consider:

- X'_1: cash deposits (linked to premiums) rescaled aggregated amount over each future period;
- X'_2: cash deposits (linked to claims) rescaled aggregated amount over each future period;
- X'_3: reinsurance commissions rescaled aggregated amount over each future period;
- X'_4: brokerage commissions rescaled aggregated amount over each future period;
- X'_5: reinstatement premiums rescaled aggregated amount over each future period;
- X'_6: adjustment premiums rescaled aggregated amount over each future period;
- X'_7: paid losses rescaled aggregated amount over each future period;
- X'_8: cash calls rescaled aggregated amount over each future period;
- ....

On one hand, resorting to this more complex model structure, one can avail oneself of more information supplied by the original historical-statements-record-keeping database, which could potentially improve the predictions accuracy and prove to yield better outcomes.

On the other hand, the connection between this newly structured data set (Figure 75) and the historical balances database is no more straightforward. In other words, we are no more able to substract the future cash flows coming from EBNYS balances, since we do not have access to their underlying composition (transaction elements of different natures adding up to a given net balance). This means that the two models (Duration model calibrated on all historical balances and applied to EBNYS balances, and LSTM with various features at each time step) are not complementary anymore. The LSTM neural network alone could be employed to predict all near future monthly cash flows without even differentiating EBNYS balances settlements from not yet existing balances settlements. In this case, we would maybe lose the precious available information provided by EBNYS balances at the date of prediction.

Part III
Run-off liabilities cash flows long-term prediction

We have just seen a methodology to anticipate better cash needs in currencies for a short-term horizon, that is, we devised a procedure that aims to predict the next 12-month monthly technical cash flows. Now we will try to reach a solution to forecast long-term yearly cash flows in currencies proceeding from AGRe's existing reserves. As we stressed in chapter 4, it is necessary to have a good idea of our own liabilities and how those liabilities will evolve, in order to quantify our future FX risk exposure and be able to optimally mitigate FX risk along with the creation of investment opportunities.

7 Credibility theory applied to claims development pattern and IBNR allocation

As already mentioned, to be able to establish a good representation of future potential technical outflows and inflows in various currencies, and thus an adequate depiction of future exposures to FX risk, it is first required to dispose of a reliable estimate of current reserves and a trustworthy view on how those reserves will liquidate.

Observed claims paid and reported claims incurred development figures and patterns can be used to predict the ultimate claim amount for different categories of risks as well as for a whole line of business or insurance portfolio. For a given individual class of risk, we can also base ourselves on corresponding observed development pattern to infer a liquidation pattern for reserves associated to that set of risk, that is, the rate of clearance of those reserves or the pace at which those reserves will transform into payments.

To correctly assess AGRe's future FX exposures, having well built and sound estimates of IBNR, file-to-file and other reserves, at the finer grain level of currencies inside a given loss ratio segment, is not sufficient. We need, as well, to work out valid and steady individual development patterns for each of the considered sub-categories of risk.

When going down this path of estimating development patterns for finer granularity of risk, one is generally confronted with the problem that the observed development figures, within the related loss development triangle, heavily fluctuate due to random fluctuations and scarce data. The finer is the grain, the smaller are the individual subsets of risks considered, the fewer historical observations referring to each of those subsets are available, and the less reliable become the corresponding individual development patterns.

In order to tackle this issue, we can rely on more stable and trustworthy data and development patterns based on broader sets of risks, rather than only on the observed individual data. Besides the individual (finer-grained level) claims triangles, collective source of information such as the development pattern of wider scope data, can indeed be tapped into for the purpose of providing useful knowledge on the future individual developments in question. This happens to be a perfect application opportunity for credibility theory, and seems to meet all the prerequisites to use such a mathematical tool. The reader may refer to a synthesis of the theory of credibility (mainly Bühlmann-Straub model and hierarchical credibility model) in appendix F, which was written on the basis of the book "A Course in Credibility Theory and its Applications" by Hans Bühlmann and Alois Gisler [3]. This summary of carefully chosen domains of the credibility theory we develop in appendix, allows us to know exactly to what extent we should rely on those broader observed data and risk scopes. In particular, the Bühlmann-Straub model seems especially well adapted to our current concern, namely, to combine individual and collective claims information to get a best possible estimate of individual development patterns, through a best possible linear combination of individual and collective development factors' estimators.

7.1 Application of the Bühlmann-Straub model to credibilized development factors

7.1.1 Bayesian Chain-Ladder model

Let us first make use of the same notations introduced in subsection 3.2 ("IBNR assessment methodologies"), and define new ones:

- $C_{i,j}$: total cumulative claims payments or claims incurred of underwriting year $i \in [\![1,N]\!]$ observed at the end of development period $j \in [\![1,N]\!]$.
- $F_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}$: corresponding individual development factors.
- $C_{i,N}$: ultimate total claim amount related to underwriting year i (we assume that the claims development ends at development year N).
- $\mathcal{D}_N = \{C_{i,j} : 1 \leq i \leq N, 1 \leq j \leq N, i+j \leq N+1\}$: all available observations at present time $(i+j = N+1)$.
- $\forall i \in [\![1,N]\!], \quad R_i = C_{i,N} - C_{i,N-i+1}$: outstanding claims reserves or IBNR (depending on the studied quantities $C_{i,j}$'s).
- $\forall j \in [\![1,N]\!], \quad \mathbf{C}_j = (C_{1,j}, C_{2,j}, \ldots, C_{N-j+1,j})'$: column vectors of trapezoid $\mathcal{D}_N$.
- $\forall j \in [\![1,N]\!], \quad \mathcal{B}_j = \{C_{i,k} : i+k \leq N+1, k \leq j\} = \{\mathbf{C}_1, \mathbf{C}_1, \ldots, \mathbf{C}_j\} \subset \mathcal{D}_N$: set of observations up to development period j at present time.
- $\forall j \in [\![1,N-1]\!], \quad \mathbf{F}_j = (F_{1,j}, F_{2,j}, \ldots, F_{N-j,j})'$: column vectors of the observed F-trapezoid.
- $\forall j \in [\![1,N]\!]$ and $\forall k \in [\![1,N-j+1]\!], \quad S_j^{[k]} = \sum_{i=1}^{k} C_{i,j}$
- $\forall (i,j) \in [\![1,N]\!]^2$, and $\forall k \in [\![N-i+2,N]\!], \quad C_{i,k}^{CL} = C_{i,N-i+1} \prod_{j=N-i+1}^{k-1} \widehat{f}_j$ with $\widehat{f}_j = \frac{S_{j+1}^{[N-j]}}{S_j^{[N-j]}}$ the Chain-Ladder or age-to-age factors.

Thus the ultimate claim $C_{i,N}$ is predicted by $C_{i,N}^{CL}$ and the Chain-Ladder reserve of underwriting year i at time N is $R_i^{CL} = C_{i,N}^{CL} - C_{i,N-i+1}$.

Mack Chain-Ladder model assumptions can be rewritten:

- Random variables $C_{i,j}$ belonging to different attaching periods (in our case underwriting years) i are independent.
- $\exists \quad f_j > 0$ and $\sigma_j^2 > 0$, such that for all $i \in [\![1,N]\!]$ and $j \in [\![1,N-1]\!]$,

$$E\left[C_{i,j+1} \mid C_{i,j}\right] = f_j C_{i,j} \iff E\left[F_{i,j} \mid C_{i,j}\right] = f_j$$

$$\operatorname{Var}\left[C_{i,j+1} \mid C_{i,j}\right] = \sigma_j^2 C_{i,j} \iff \operatorname{Var}\left[F_{i,j} \mid C_{i,j}\right] = \frac{\sigma_j^2}{C_{i,j}}$$

The Chain-Ladder methodology and the underlying stochastic model of Mack take into account only the individual data of a specific development triangle. To avail oneself of prior or portfolio information (from other "similar" risks), that will certainly enhance our knowledge of the unknown development pattern of the considered individual claims data, we have to consider the Chain-Ladder methodology in a Bayesian set-up.

In the Bayesian Chain-Ladder framework, the unknown Chain-Ladder factors $\mathbf{f} = (f_j)_{j \in [\![1,N-1]\!]}$, are assumed to be realizations of independent, real valued random variables $\mathbf{F} = (F_j)_{j \in [\![1,N-1]\!]}$, and, conditionally given $\mathbf{F}$, the Mack Chain-Ladder model assumptions are fulfilled, that is:

- $(F_j)_{j \in [\![1,N-1]\!]}$ are independent.
- Conditionally, given $\mathbf{F}$, the random variables $C_{i,j}$ belonging to different underwriting years i are independent.
- Conditionally, given $\mathbf{F}$ and $\mathcal{B}_j$, the conditional distribution of $F_{i,j}$ depends only on $C_{i,j}$ and
$$E\left[F_{i,j} \mid \mathbf{F}, \mathcal{B}_j\right] = F_j, \quad \text{and} \quad \operatorname{Var}\left[F_{i,j} \mid \mathbf{F}, \mathcal{B}_j\right] = \frac{\sigma_j^2\left(F_j\right)}{C_{i,j}}$$

Defining $\widehat{F}_j = \frac{S_{j+1}^{[N-j]}}{S_j^{[N-j]}}$ the estimator of the Chain-Ladder factor f_j in the classical Chain-Ladder model, it follows that

$$E\left[\widehat{F}_j \mid \mathbf{F}, \mathcal{B}_j\right] = F_j, \quad \text{and} \quad \operatorname{Var}\left[\widehat{F}_j \mid \mathbf{F}, \mathcal{B}_j\right] = \frac{\sigma_j^2\left(F_j\right)}{S_j^{[N-j-1]}}$$

It can be shown that a posteriori, given the observations $\mathcal{D}_N$, the random variables $(F_j)_{j \in [\![1,N-1]\!]}$ are independent with a known posterior distribution (see proof of Theorem 3.2 in [4]).

Let us explicit the Bayes estimator of the development factors and the ultimate claim.
First, if Z is an unknown random variable and $\boldsymbol{X}$ a random vector of observations, then the best (using the expected quadratic loss as optimality criterion) estimator of Z is $Z^{\text{Bayes}} = E[Z \mid \mathbf{X}]$ and Z^{Bayes} also minimizes the conditional quadratic loss, i.e.

$$Z^{\text{Bayes}} = \underset{\widehat{Z}}{\arg\min} E\left[(\widehat{Z} - Z)^2 \mid \mathbf{X}\right]$$

Second, let us define the conditional mean square error of $\widehat{C}_{i,N}$, a predictor of the ultimate claim $C_{i,N}$ based on the observations $\mathcal{D}_N$:

$$\operatorname{MSE}\left(\widehat{C}_{i,N}\right) = E\left[\left(\widehat{C}_{i,N} - C_{i,N}\right)^2 \mid \mathcal{D}_N\right]$$

Denoting $\widehat{R}_i = \widehat{C}_{i,N} - C_{i,N-i+1}$ the corresponding reserve estimate, we have

$$\operatorname{MSE}\left(\widehat{C}_{i,N}\right) = \operatorname{MSE}\left(\widehat{R}_i\right) = E\left[\left(\widehat{R}_i - R_i\right)^2 \mid \mathcal{D}_N\right]$$

In this particular case, the Bayes estimator is given by $C_{i,N}^{\text{Bayes}} = E\left[C_{i,N} \mid \mathcal{D}_N\right]$, and correspond to the best estimator of $C_{i,N}$ minimizing the conditional mean square error.
Furthermore, we can prove that

$$C_{i,N}^{Bayes} = C_{i,N-i+1} \prod_{j=N-i+1}^{N-1} F_j^{Bayes} \quad \text{where} \quad F_j^{Bayes} \quad \text{is the Bayes estimator of} \quad F_j$$

It can also be demonstrated that the conditional mean square error of the Bayes reserve of attaching period i is given by

$$\text{MSE}\left(R_i^{\text{Bayes}}\right) = E\left[\left(C_{i,N}^{\text{Bayes}} - C_{i,N}\right)^2 \mid \mathcal{D}_N\right] = C_{i,N-i+1}\Gamma_{N-i+1} + C_{i,N-i+1}^2\Delta_{N-i+1}^B$$

where

$$\begin{aligned}
\Gamma_{N-i+1} &= \sum_{k=N-i+1}^{N-1} \left\{ \prod_{m=N-i+1}^{k-1} F_m^{\text{Bayes}} \cdot E\left[\sigma^2\left(F_k\right) \mid \mathcal{D}_N\right] \cdot \prod_{n=k+1}^{N-1} E\left[F_n^2 \mid \mathcal{D}_N\right] \right\} \\
\Delta_{N-i+1}^B &= \text{Var}\left(\prod_{j=N-i+1}^{N-1} F_j \mid \mathcal{D}_N\right) \\
R_i^{\text{Bayes}} &= C_{i,N}^{\text{Bayes}} - C_{i,N-i+1} \quad \text{the associated reserve estimate}
\end{aligned}$$

7.1.2 Chain-Ladder credibility model

$C_{i,N}^{Bayes} = C_{i,N-i+1} \prod_{j=N-i+1}^{N-1} F_j^{Bayes}$ may be the best estimator of $C_{i,N}$, minimizing the conditional mean square error, still to calculate F_j^{Bayes} one needs to know the (usually unknown) distributions of the F_j as well as the conditional distributions of the $C_{i,j}$'s, given $\mathbf{F}$. The edge of credibility theory resides in the fact that only the first and second moments are required (assumed to exist and be finite). These moments can be estimated from the considered portfolio data.

The credibility based predictor of the ultimate claim $C_{i,N}$ given $\mathcal{D}_N$ is defined by $C_{i,N}^{(cred)} = C_{i,N-i+1} \prod_{j=N-i+1}^{N-1} F_j^{cred}$, with $R_i^{(cred)} = C_{i,N}^{(cred)} - C_{i,N-i+1}$ the corresponding reserve estimate.

As a reminder, a credibility estimator based on observations (or statistics) $\mathbf{X}$ is the best (under expected quadratic loss criterion) linear (in $\mathbf{X}$'s components) estimators. Thus, for estimating F_j, we base our estimator on the only observations of the F-trapezoid containing information on F_j, namely the column vector $\mathbf{F}_j = (F_{i,j})_{i \in [\![1,N-j]\!]}$:

$$F_j^{cred} = \text{Pro}(F_j \mid L(\mathcal{B}_\mathbf{j}, 1)) = \underset{\left\{\widehat{F}_j : \widehat{F}_j = a_0^{(j)} + \sum_{i=1}^{N-j} a_i^{(j)} F_{i,j}\right\}}{\arg\min} E\left[\left(\widehat{F}_j - F_j\right)^2 \mid \mathcal{B}_j\right]$$

or equivalently (Markov property)

$$F_j^{cred} = \text{Pro}(F_j \mid L(\mathbf{C_j}, 1)) = \underset{\left\{\widehat{F}_j : \widehat{F}_j = a_0^{(j)} + \sum_{i=1}^{N-j} a_i^{(j)} F_{i,j}\right\}}{\arg\min} E\left[\left(\widehat{F}_j - F_j\right)^2 \mid \mathbf{C}_j\right]$$

Finally, since conditionally on $\mathcal{B}_j$, the random variables $(F_{i,j})_{i \in [\![1,N-j]\!]}$ fulfill the assumptions of the Bühlmann-Straub model, the credibility estimators of the unknown Chain-Ladder factors $(F_j)_{j \in [\![1,N-1]\!]}$, are

$$F_j^{cred} = \alpha_j \widehat{F}_j + (1 - \alpha_j) f_j$$

where $\widehat{F}_j = \dfrac{S_{j+1}^{[N-j]}}{S_j^{[N-j]}}$, $\alpha_j = \dfrac{S_j^{[N-j]}}{S_j^{[N-j]} + \frac{\sigma_j^2}{\tau_j^2}}$, $f_j = E\left[F_j\right]$, $\sigma_j^2 = E\left[\sigma_j^2\left(F_j\right)\right]$, and $\tau_j^2 = \mathrm{Var}\left[F_j\right]$

It follows immediately that

$$\mathrm{MSE}\left(F_j^{Cred}\right) = E\left[\left(F_j^{cred} - F_j\right)^2 \mid \mathcal{B}_j\right] = \alpha_j \frac{\sigma_j^2}{S_j^{[N-j]}} = \left(1-\alpha_j\right)\tau_j^2$$

F_j^{cred} is then a credibility weighted average between the classical Chain-Ladder estimator $\widehat{F}_j$, and the a priori expected value f_j. The structural parameters f_j, σ_j^2 and τ_j^2 can be estimated from portfolio data by using standard estimators presented in the Bühlmann-Straub model subsection F.3 of this thesis appendix.

Within this credibility framework, the following interesting result can be shown:

$$\mathrm{MSE}\left(R_i^{(Cred)}\right) \cong C_{i,N-i+1}\Gamma^*_{N-i+1} + C^2_{i,N-i+1}\Delta^*_{I-i+1}$$

where

$$\Gamma^*_{N-i+1} = \sum_{k=N-i+1}^{N-1}\left\{\prod_{m=N-i+1}^{k-1} F_m^{Cred}\cdot\sigma_k^2 \prod_{n=k+1}^{N-1}\left(\left(F_n^{Cred}\right)^2 + \alpha_n \frac{\sigma_n^2}{S_n^{[N-n-1]}}\right)\right\}$$

$$\Delta^*_{N-i+1} = \prod_{j=N-i+1}^{N-1}\left(\left(F_j^{Cred}\right)^2 + \alpha_j\frac{\sigma_j^2}{S_j^{[N-j-1]}}\right) - \prod_{j=N-i+1}^{N-1}\left(F_j^{Cred}\right)^2$$

and, regarding the total credibility reserve $R^{(cred)} = \sum_i R_i^{(cred)}$:

$$\mathrm{MSE}\left(R^{(cred)}\right) \simeq \sum_i \mathrm{MSE}\left(R_i^{(cred)}\right) + 2\sum_{i=1}^{N}\sum_{k=i+1}^{N} C_{i,N-i+1}C^{(cred)}_{k,N-i+1}\Delta^*_{N-i+1}$$

To close this theoretical section, one could be interested in proving that if conditionally on $\mathbf{F}$ and $\mathcal{B}_j$, the random variables $\left(F_{i,j}\right)_{i\in[\![1,N]\!]}$ are independent with a distribution belonging to the one-parameter exponential dispersion family and that if the a priori distribution of F_j belongs to the family of the natural conjugate priors, then $F_j^{Bayes} = F_j^{cred}$. In other words, the credibility estimators are exact Bayesian in the case of the exponential family with its natural conjugate priors.

7.1.3 Bühlmann-Straub model's algorithm implementation to claims (paid or incurred) development factors

We just saw a procedure that tells us how much one should rely either on the claims experience of an individual subgroup of risks in question or on the observations of other similar risk lines making up an homogeneous collective, i.e. a broader group of risks. Namely, we have considered the Chain-Ladder reserving method in a Bayesian set up, which allows for combining individual claims development data with portfolio information, and derived the Bayes estimators as well as the credibility estimators within this Bayesian framework.

Let us formalize ones and for all the correspondences between the mathematical notations in the Bühlmann-Straub model as presented in F.3, and the ones used in a claims development analysis context.

To apply directly the Bühlmann-Straub model to claims development factors through an algorithm, we must consider the following indices correspondence.

Index	**Bühlmann-Straub model** ($X_{i,j}$)	**Applied to development factors** ($F_{i,j}^{ind}$)
i	individual	underwriting year
j	j th observation for a given individual	development year
ind	NA	individual

Table 26: Indices correspondence between the Bühlmann-Straub model and its application to credibilized development factors

For a fixed development period $j \in [\![1, N-1]\!]$, we thus have the following equivalences:

$$(X_{i,j})_{i\in[\![1,I]\!],j\in[\![1,n=N-j]\!]} \iff \left(\mathbf{F}_j^{ind} = \left(F_{i,j}^{ind}\right)_{i\in[\![1,N-j]\!]}\right)_{ind\in[\![1,I]\!]}$$

$$(w_{i,j})_{i\in[\![1,I]\!],j\in[\![1,n=N-j]\!]} \iff \left(C_{i,j}^{ind}\right)_{ind\in[\![1,I]\!],i\in[\![1,N-j]\!]}$$

$$\left(X_i = \sum_{j=1}^{n} \frac{w_{ij}}{w_{i\bullet}} X_{i,j}\right)_{i\in[\![1,I]\!]} \iff \left(F_j^{ind} = \sum_{i=1}^{N-j} \frac{C_{i,j}^{ind}}{C_{\bullet j}^{ind}} F_{ij}^{ind}\right)_{ind\in[\![1,I]\!]}$$

$$\left(\widehat{\mu(\Theta_i)}^{cred} = \alpha_i X_i + (1-\alpha_i)\mu_0\right)_{i\in[\![1,I]\!]} \iff \left(\widehat{F_j(\Theta_{ind})}^{cred} = \alpha_j^{ind} F_j^{ind} + (1-\alpha_j^{ind}) f_j\right)_{ind\in[\![1,I]\!]}$$

$$\left(\widehat{\mu(\Theta_i)}^{emp} = \widehat{\alpha}_i X_i + (1-\widehat{\alpha}_i)\widehat{\mu}_0\right)_{i\in[\![1,I]\!]} \iff \left(\widehat{F_j(\Theta_{ind})}^{emp} = \widehat{\alpha}_j^{ind} F_j^{ind} + (1-\widehat{\alpha}_j^{ind})\widehat{f}_j\right)_{ind\in[\![1,I]\!]}$$

$$\text{where} \quad \left(\widehat{\alpha}_i = \frac{w_{i\bullet}}{w_{i\bullet} + \frac{\widehat{\sigma^2}}{\widehat{\tau^2}}}\right)_{i\in[\![1,I]\!]} \iff \left(\widehat{\alpha}_j^{ind} = \frac{C_{\bullet j}^{ind}}{C_{\bullet j}^{ind} + \frac{\widehat{\sigma_j^2}}{\widehat{\tau_j^2}}}\right)_{ind\in[\![1,I]\!]}$$

$$\text{and} \quad \widehat{\mu}_0 = \frac{\sum_i \widehat{\alpha}_i X_i}{\sum_i \widehat{\alpha}_i} \iff \widehat{f}_j = \frac{\sum_{ind} \widehat{\alpha}_j^{ind} F_j^{ind}}{\sum_{ind} \widehat{\alpha}_j^{ind}}$$

The above right-hand side equations with the displayed notations reflects exactly what was implemented in R algorithms in order to obtain credibilized payment and development patterns at different aggregation levels, including fine grained levels for which estimated claims development factors generally are extremely volatile and untrustworthy in a classical framework (without resorting to credibility theory).
Figures 76 and 77 below, represent in a visual way the equivalence between the structures of the Bühlmann-Straub model in a general set-up (see F.3 in appendix or [3]) and in a claims (paid or incurred) development one.

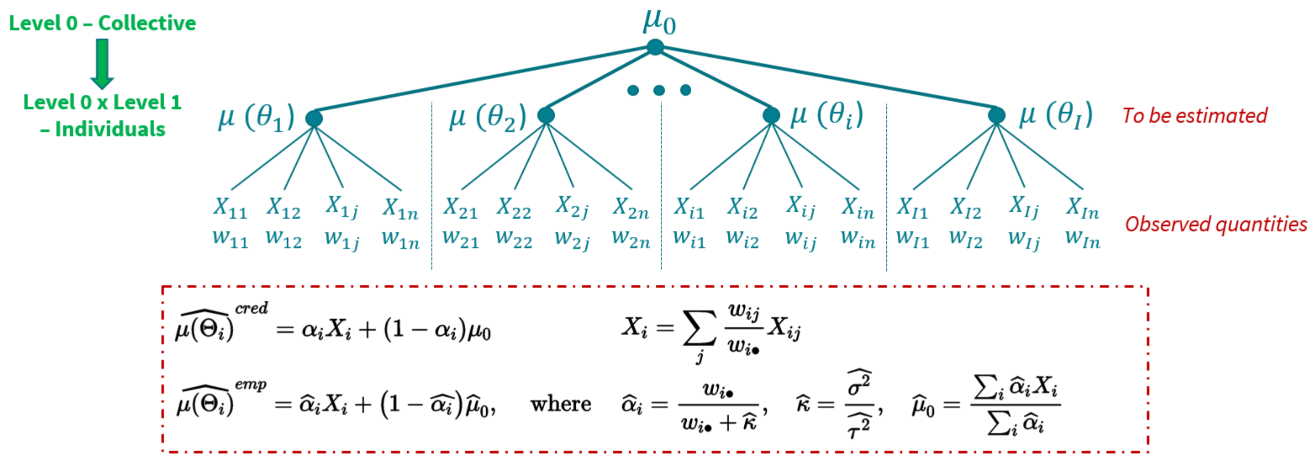


Figure 76: Bühlmann-Straub model - structure and notation

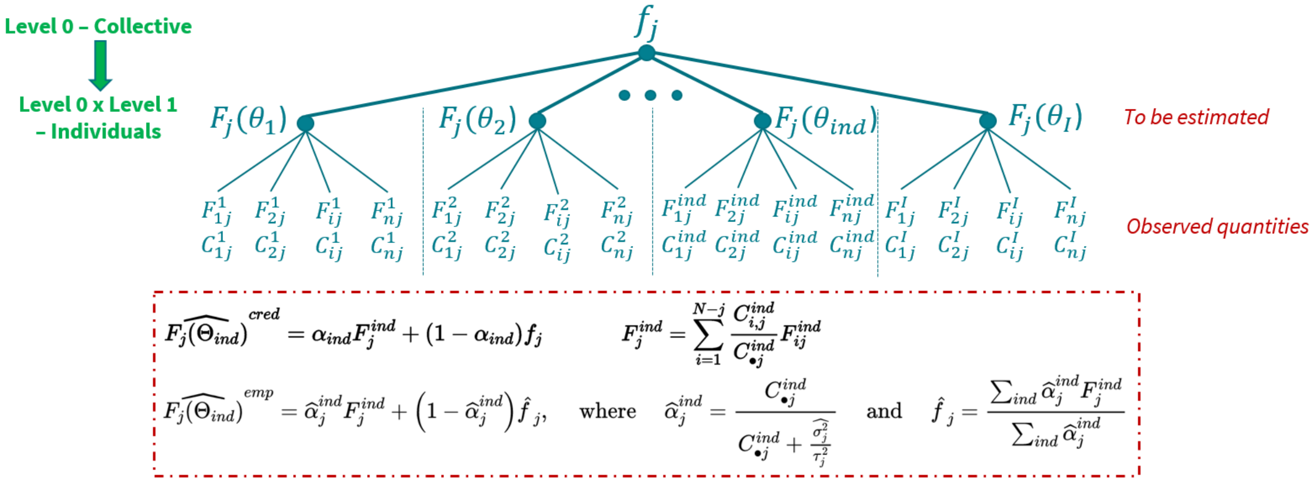


Figure 77: Bühlmann-Straub model applied to development factors - structure and notation

7.2 AGRe's portfolio multiple aggregation levels and corresponding credibilized development patterns

7.2.1 Claims triangle databases

There exist two claims triangle databases mirroring the two AGRe's business scopes: Life and P&C. Those triangle databases contain the history of accepted risk incremental technical positions by reinsurance treaty, underwriting year, accounting year (development period), cedent, ceding period, etc. All historical incremental paid claims and case reserves amounts are valued in euro equivalent according to the accounting period's closing exchange rate. In particular, one can retrieve recorded claims incurred (and paid) increments from one accounting year to the other, and thus derive development patterns on any portfolio data aggregation level, from the whole P&C or Life portfolio down to individual treaties.

It is worth underlying that the considered claims amounts (paid or incurred) in those triangle data bases are neither from ground up (FGU) nor net of retrocession recoverable. Instead, they correspond to claims amounts covered by the reinsurance treaties between AGRe and the cedent AXA entities.

Furthermore, let us highlight once again that AGRe covers multi-currencies risks, and thus settles claims or constitutes reserves in several currencies. The corresponding development triangles handled by the reserving team of AGRe are in fact multi-currencies triangles. To single out clearly the sole technical variations of claims involved in those triangles and remove noises due to historical FX rates fluctuations, all recorded amounts in foreign currencies are systematically converted into euro at each closing / triangle cut-off date on the basis of spot FX rates (used also for reserves evaluation). Amounts are always valued in euro equivalent of last year closing. This way, one does not need to take into consideration future FX fluctuations when projecting claims developments, and estimating IBNR reserves. In other words, the historical amounts (quantified in euro unit) of AGRe's triangles (at arbitrary level of aggregation) are dynamical and are reevaluated each year, in order to be measured in the same "actualized" euro value. No inflation effect is taken into account.

7.2.2 AGRe portfolio hierarchical subdivision

We can subdivide AGRe's portfolio, and consequently all related data, statistics, performed studies and analysis, according to different hierarchical scope levels.

From now on, we will denote by "Level 0 x Level 1" the set of all individual risk classes resulting from combinations of "Level 1" risk subgroups within "Level 0" risk groups. In this set-up, we will refer to the "Level 1" as the "lower level". The same applies for a more-than-two-level configuration, swapping comparative to superlative adjectives. In addition, to create connections with the credibility theory vocabulary, in a two-level hierarchical structure, elements of "Level 0" can be seen as "collectives", and elements of "Level 0 x Level 1" as "individuals".

The following figure represents the hierarchical nature of AGRe's portfolio structure, through a conceivable breakdown.

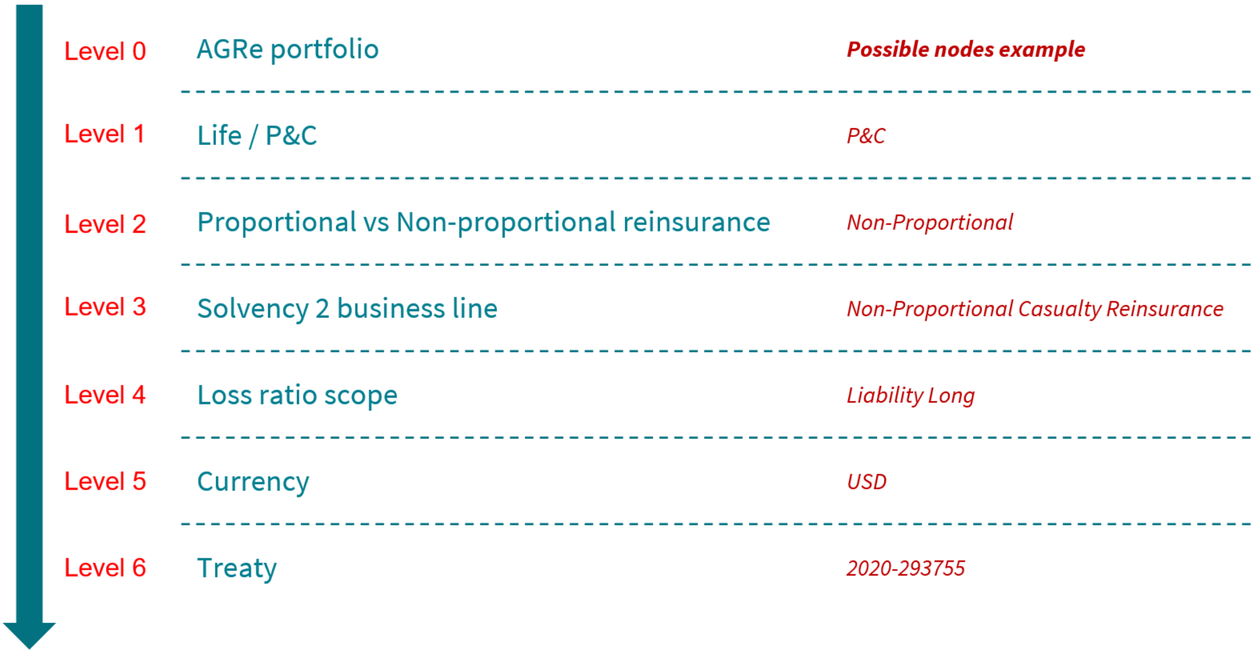


Figure 78: AGRe portfolio's hierarchical structure example

Figure 78 displays only one possible hierarchical configuration, and others with different number of levels could be considered, such as:

- Level 0 x Level 1 = Loss ratio scope x Currency
- Level 0 x Level 1 = Solvency 2 business line x Loss ratio scope
- Level 0 x Level 1 x Level 2 = Proportional vs Non proportional reinsurance x Solvency 2 business line x Currency

This hierarchical subdivision of AGRe's portfolio can be seen as tree structure with parent / child relationships analogous to the collective / individual ones underlying the credibility theory and introduced in this chapter. All paths from the lower level to the higher one, in a given hierarchical configuration, uniquely define an individual risk subclass as an element among others inside broader encompassing risk groups. Many claims development triangles can then be built on top of those established subgroups of risks. The nodes at the "Solvency 2 business line" and "Loss ratio scope" levels have been listed in subsection 2.3 of this thesis ("AGRe segmentation").

Both the Bühlmann-Straub model (for a 2-level consideration) and the hierarchical credibility model (Chapter 6 of [3] summarized in appendix part F.4) (when the considered number of levels is higher than 2), are particularly well suited for estimating quantities in this hierarchical structure set-up. The quantities of interest here, being the development factors of individual (lowest level in the tree, i.e. smallest risk scope) claims triangles, from which we seek credibility estimators.

7.2.3 Numerical examples

We display here some illustrating outputs of the R algorithm implementing the Chain-Ladder credibility model (i.e. applying the Bühlmann-Straub model to loss or payment development factors) on AGRe portfolio historical claims data.

The following two tables show the computed values of the development factors F_j's estimated by classical Chain-Ladder and by the credibility estimators of previous subsections 7.1.1 and 7.1.3. The structural parameters f_j's (named here F_j_collective), σ_j^2's and τ_j^2's have been estimated by the "standard estimators" exhibited in the appendix (part "Estimation of the structural parameters σ^2 and τ^2" of subsection F.3.5).

First, let us see the explicit numerical values taken by the individual, collective and credibilized development factors (F_j's) with their corresponding development patterns (Z_j's) for a combination of two data aggregation levels (collective vs individual), namely Level 0 (SOLVENCY II BUSINESS LINE) x Level 1 (LR SCOPE) = Non-Proportional Casualty Reinsurance x Miscellaneous.

development_period	Fj_individual	Fj_collective	alphaj	sigma2	Fj_credibility	Zj_individual	Zj_collective	Zj_credibility
1	4.581	3.585	0.818	15142467.172	4.399	0.148	0.134	0.121
2	1.213	1.363	0.718	2355133.004	1.255	0.677	0.480	0.530
3	1.107	1.073	0.752	595423.243	1.098	0.822	0.655	0.666
4	1.040	1.068	0.000	2052834.077	1.068	0.909	0.702	0.731
5	0.998	1.074	0.000	2288832.907	1.074	0.946	0.750	0.780
6	1.023	1.012	0.000	271702.983	1.012	0.944	0.805	0.838
7	1.061	1.054	0.000	734232.983	1.054	0.966	0.815	0.848
8	0.975	1.006	0.599	371466.231	0.988	1.025	0.859	0.894
9	1.071	1.025	0.000	1788595.260	1.025	1.000	0.864	0.883
10	1.003	1.018	0.616	140790.876	1.009	1.071	0.885	0.905
11	0.928	0.962	0.522	882940.649	0.944	1.074	0.901	0.913
12	0.979	1.049	0.000	638687.402	1.049	0.997	0.867	0.862
13	0.996	1.001	0.000	67968.927	1.001	0.976	0.910	0.904
14	0.993	1.005	0.000	54877.323	1.005	0.972	0.910	0.905
15	1.000	1.026	0.045	349639.202	1.024	0.965	0.915	0.909
16	0.997	1.010	0.000	68136.910	1.010	0.965	0.938	0.931
17	0.998	0.979	0.833	59990.081	0.995	0.962	0.948	0.941
18	1.001	1.026	0.000	1735109.769	1.026	0.961	0.928	0.936
19	0.984	1.013	0.533	14978.587	0.997	0.962	0.952	0.961
20	1.009	1.044	0.000	90617.136	1.044	0.947	0.964	0.958
21	0.999	0.966	0.824	43827.385	0.994	0.956	1.006	1.000
22	1.009	0.991	0.578	43016.101	1.001	0.955	0.971	0.994
23	0.999	0.995	0.000	11239.575	0.995	0.963	0.963	0.995
24	1.039	1.004	0.537	5437.554	1.023	0.962	0.958	0.990
25	1.000	1.012	0.854	4316.500	1.002	1.000	0.962	1.013
26	1.000	0.989	0.000	11204.280	0.989	1.000	0.974	1.014
27	1.000	0.982	0.784	5220.626	0.996	1.000	0.963	1.004
28	1.000	1.059	0.965	10601.086	1.002	1.000	0.946	1.000
29	1.000	0.998	NA	NA	0.998	1.000	1.002	1.002
30	1.000	1.000	NA	NA	1.000	1.000	1.000	1.000

Table 27: Individual, collective and credibilized development factors (Fj's) and corresponding development patterns (Zj's) – Level 0 (SOLVENCY II BUSINESS LINE) value X Level 1 (LR SCOPE) value = Non-Proportional Casualty Reinsurance X Miscellaneous

The values of the above result table can be visualized by looking at the following graph showing the resulting loss development patterns.

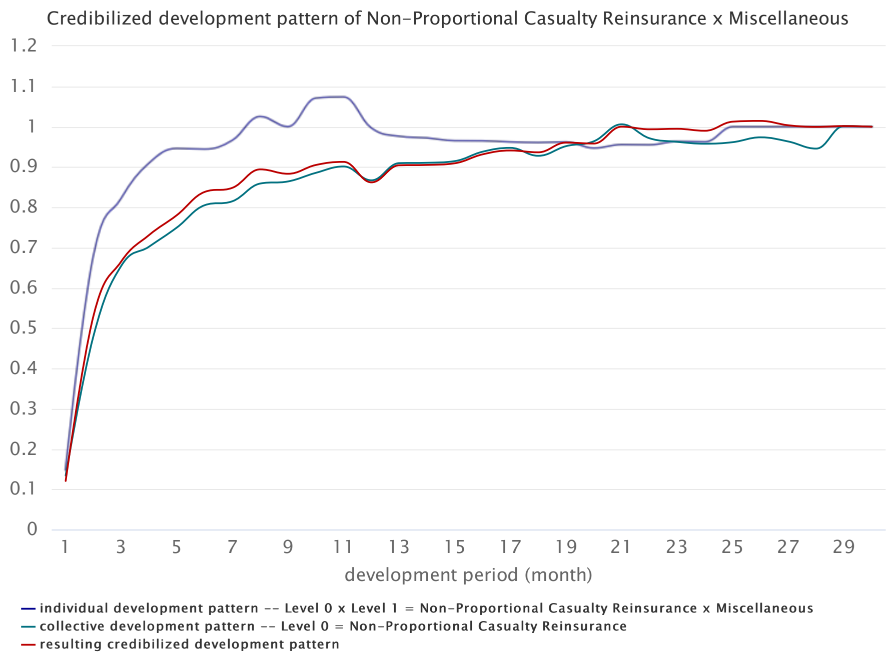


Figure 79: Developments of loss incurred (development patterns) issued from individual, collective and credibilized development factors – Level 0 (SOLVENCY II BUSINESS LINE) value X Level 1 (LR SCOPE) value = Non-Proportional Property Reinsurance X Miscellaneous

In the table below are presented the individual, collective and credibilized development factors (F_j's) with their corresponding development patterns (Z_j's) for Level 0 (LR SCOPE) x Level 1

(CURRENCY) = Transport x USD. The subsequent graph draws those particular development patterns curves.

development_period	Fj_individual	Fj_collective	alphaj	sigma2	Fj_credibility	Zj_individual	Zj_collective	Zj_credibility
1	5.186	7.790	0.000	5290295.000	7.790	0.018	0.033	0.020
2	2.427	1.963	0.000	1236591.000	1.963	0.092	0.257	0.159
3	2.041	1.315	0.971	535752.500	2.020	0.223	0.504	0.311
4	1.496	1.233	0.000	1614782.000	1.233	0.456	0.663	0.629
5	1.306	1.138	0.000	1284199.000	1.138	0.682	0.817	0.776
6	1.116	1.059	0.937	25435.730	1.112	0.890	0.929	0.882
7	0.986	0.997	0.000	7449.302	0.997	0.993	0.984	0.982
8	0.999	1.007	0.000	3289.940	1.007	0.980	0.982	0.979
9	1.001	1.006	0.797	897.573	1.002	0.979	0.988	0.986
10	1.002	1.002	0.000	2009.585	1.002	0.980	0.994	0.987
11	1.017	1.010	0.000	5551.178	1.010	0.982	0.996	0.989
12	1.000	0.999	0.000	771.351	0.999	0.999	1.006	0.999
13	1.000	0.990	0.937	921.153	0.999	0.999	1.005	0.998
14	1.001	1.002	0.000	208.249	1.002	0.999	0.995	0.997
15	1.000	1.002	0.942	77.431	1.000	1.000	0.997	0.999
16	1.000	1.000	0.000	0.008	1.000	1.000	0.999	0.999
17	1.000	1.001	0.000	14.682	1.001	1.000	0.999	0.999
18	1.000	1.000	0.000	0.025	1.000	1.000	1.000	1.000
19	1.000	1.000	0.423	0.001	1.000	1.000	1.000	1.000
20	1.000	1.000	NA	0.000	1.000	1.000	1.000	1.000
21	1.000	1.000	0.701	0.005	1.000	1.000	1.000	1.000
22	1.000	1.000	0.000	1.349	1.000	1.000	1.000	1.000
23	1.000	1.000	0.988	0.000	1.000	1.000	1.000	1.000
24	1.000	1.000	NA	0.000	1.000	1.000	1.000	1.000
25	1.000	1.000	NA	0.000	1.000	1.000	1.000	1.000
26	1.000	1.000	NA	0.000	1.000	1.000	1.000	1.000
27	1.000	1.000	NA	0.000	1.000	1.000	1.000	1.000
28	1.000	1.000	NA	0.000	1.000	1.000	1.000	1.000
29	1.000	1.000	NA	NA	1.000	1.000	1.000	1.000
30	1.000	1.000	NA	NA	1.000	1.000	1.000	1.000

Table 28: Individual, collective and credibilized development factors (Fj's) and corresponding development patterns (Zj's) – Level 0 (LR SCOPE) value X Level 1 (CURRENCY) value = Transport X USD

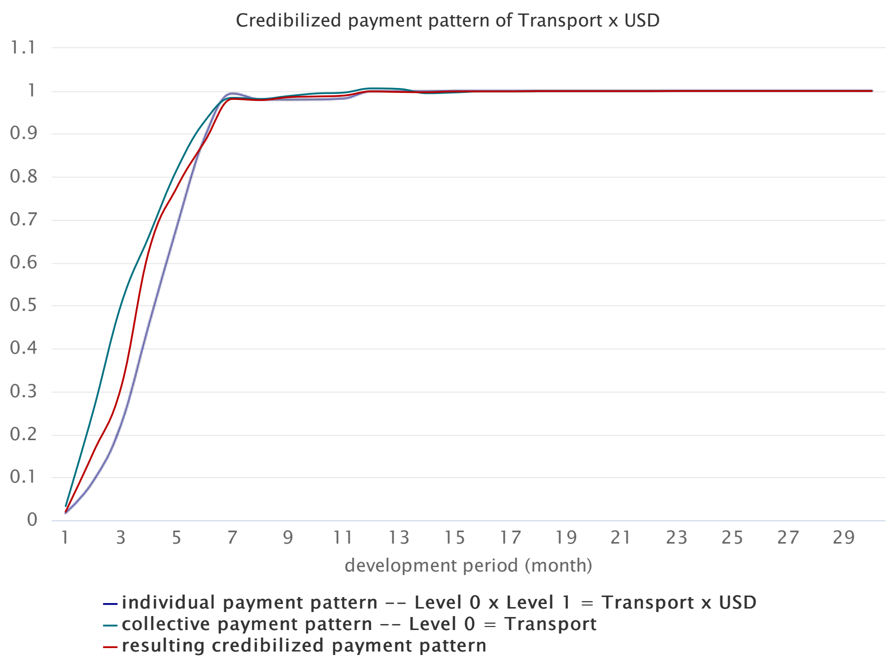


Figure 80: Developments of claims payments issued from individual, collective and credibilized development factors – Level 0 (LR SCOPE) value X Level 1 (CURRENCY) = Transport X USD

From the above tables we can see that the Chain-Ladder (F_j_individual) and the credibility (F_j_credibility) estimates can differ quite substantially. The estimates of the variance components τ_j^2 are most of the times negative leading to corresponding α_j^2 being equal to zero (cf

F.3.5 in appendix), in which cases F_j_credibility is identical to F_j_collective for all individuals in the same collective.

Claims incurred and claims payment development patterns Z_j's (individual, collective and credibilized) resulting from the application of the Bühlmann-Straub model to loss and payment development factors of various combinations of individual versus collective groups of reinsurance treaties, are graphically rendered in the figures below and in section G of the appendix.

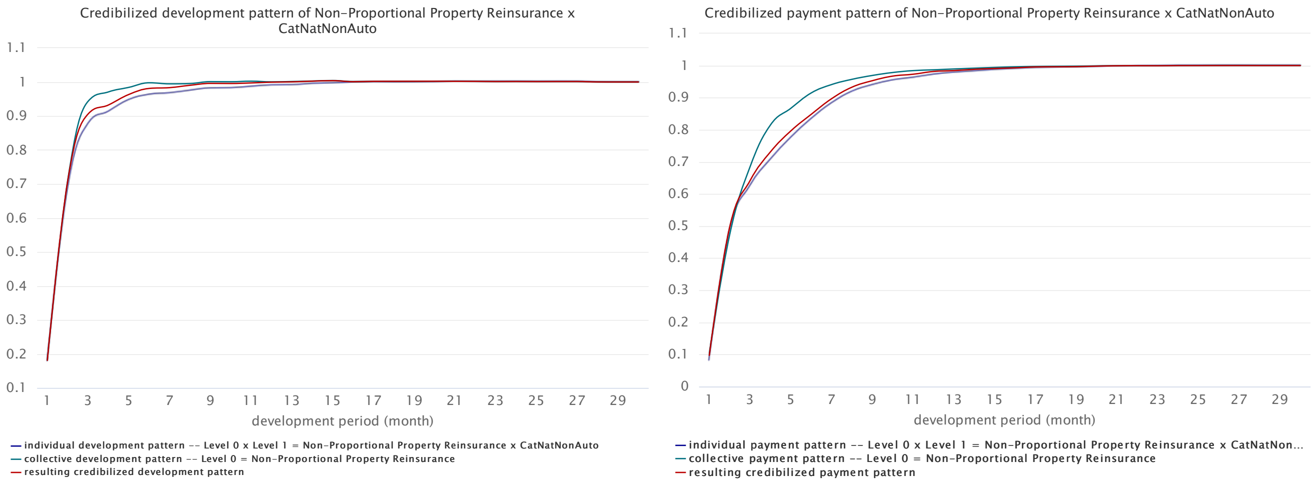


Figure 81: Developments of loss incurred (on the left) and claims payments (on the right) issued from individual, collective and credibilized development factors – Level 0 (SOLVENCY II BUSINESS LINE) value X Level 1 (LR SCOPE) value = Non-Proportional Property Reinsurance X CatNatNonAuto

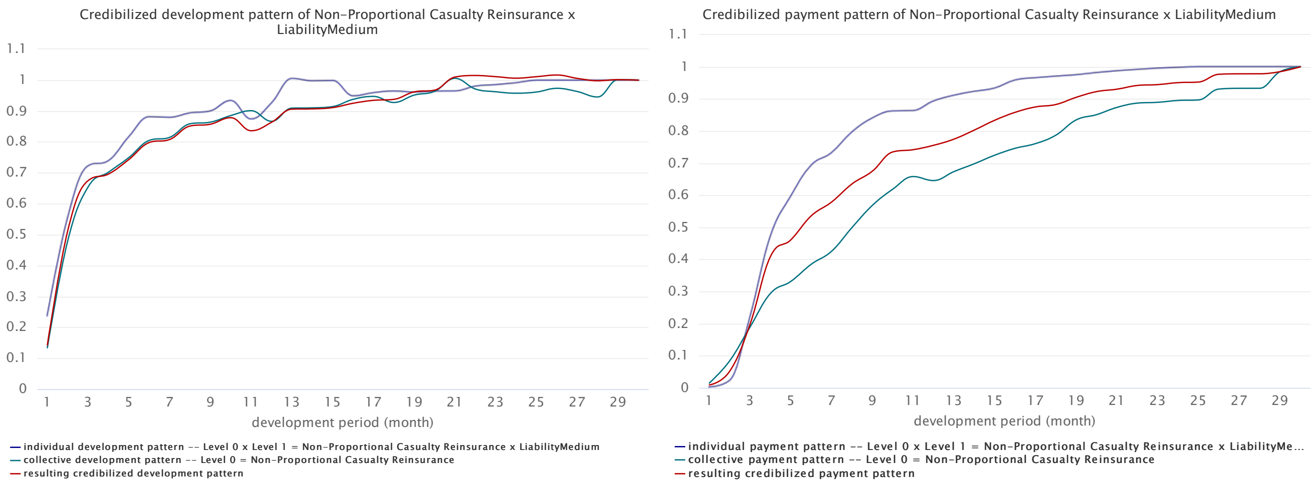


Figure 82: Developments of loss incurred (on the left) and claims payments (on the right) issued from individual, collective and credibilized development factors – Level 0 (SOLVENCY II BUSINESS LINE) value X Level 1 (LR SCOPE) value = Non-Proportional Casualty Reinsurance X LiabilityMedium

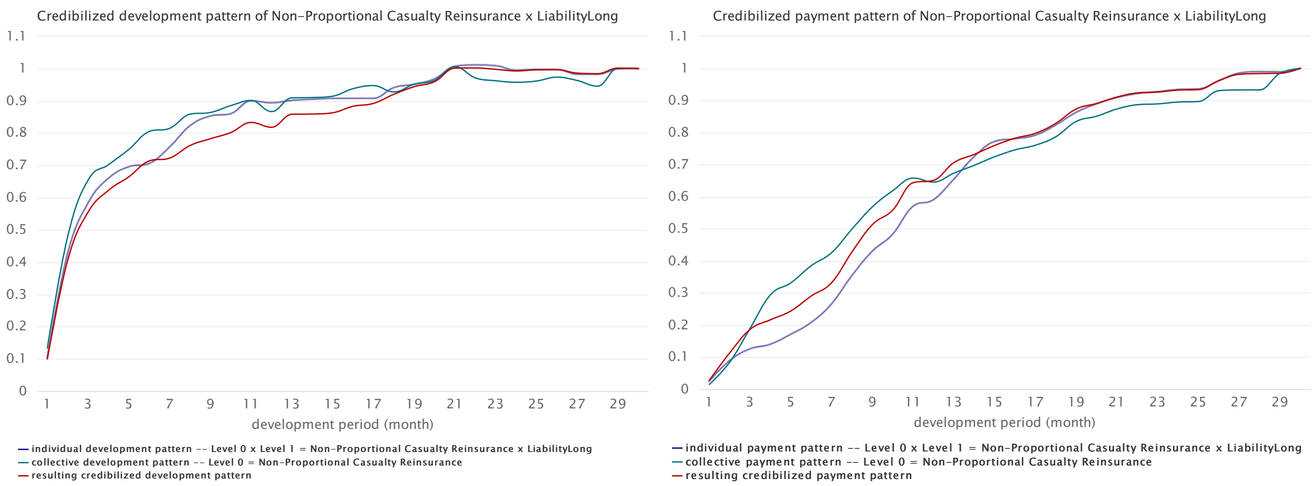


Figure 83: Developments of loss incurred (on the left) and claims payments (on the right) issued from individual, collective and credibilized development factors – Level 0 (SOLVENCY II BUSINESS LINE) value X Level 1 (LR SCOPE) value = Non-Proportional Casualty Reinsurance X LiabilityLong

From those graphs, we can observe several interesting facts:

- First, the credibilized development pattern tends to be smoother than the individual one (smoothing effect of the Chain-Ladder credibility procedure on the estimates);
- Second, higher the weigh of the individual node within the collective and steadier is its individual development pattern, closer to the individual development pattern will be the credibilized one. And conversely, if the lower node does not have much weight within the collective and/or has erratic development factors, the corresponding credibilized development pattern will be close to the collective one. As a clear illustration, the credibilized curve for the currency EUR inside a given LR SCOPE generally follows closely its individual pattern (most of the claims paid and claims incurred amounts within a given LR SCOPE are labelled in EUR), whereas the credibilized development pattern of a minor currency (whose amounts and development factors are not much representative inside a given LR SCOPE) tends to track nearly the collective one (of the encompassing LR SCOPE).
- Finally, the credibilized development pattern is not necessarily localized entirely between the individual and the collective ones. This point is a bit counter-intuitive, but what really are (and must be) encapsulated between individual and collective figures, are in fact development factors and not patterns (F_j's and not Z_j's).

An attempt have been made at an algorithm implementation of the hierarchical credibility model applied to development factors, resorting to AGRe portfolio's multi-levels. Figure 84 below presents one particular output of this algorithm. Nevertheless, when applying the hierarchical credibility model on AGRe portfolio claims data (scarce in certain branches and sub-branches), several special cases appear and have to be handled leading to skewed outputs. For this, we decided to restrain ourselves to a two-level credibility model.

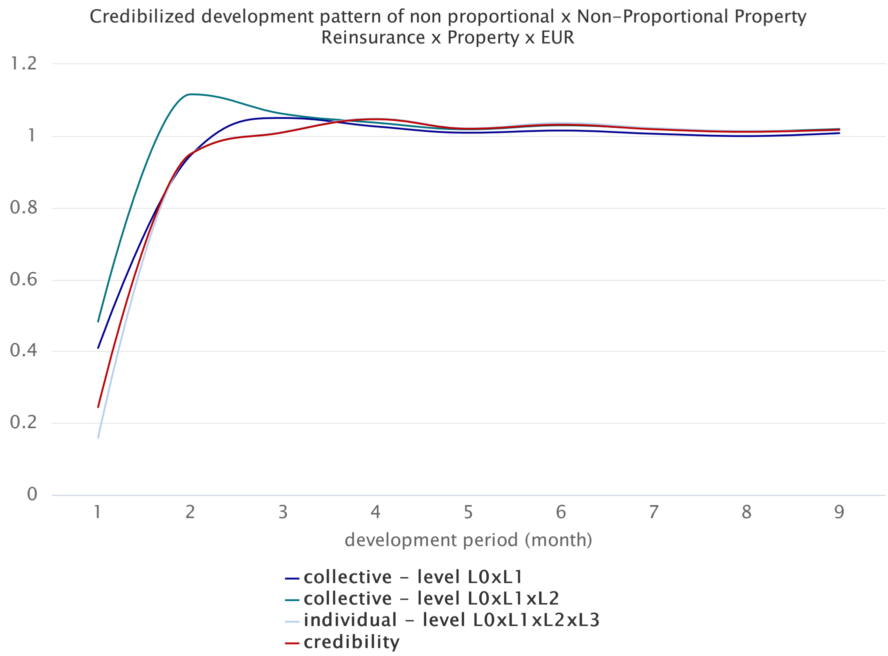


Figure 84: Development patterns issued from hierarchical credibility model – Level 0 (PROPORTIONAL vs NON-PROPORTIONAL) X Level 1 (SOLVENCY II BUSINESS LINE) X Level 2 (LR SCOPE) X Level 3 (CURRENCY) = Non-Proportional X Non-Proportional Property Reinsurance X Property X EUR

7.3 Allocating IBNR reserves at the LR scope x currency level thanks to a Bornhuetter-Ferguson approach and credibilized loss development factors

7.3.1 Relevance of a good IBNR representation at a finer level and current method

As discussed in chapter 4 and in the introductory part of this chapter, having reliable technical reserves estimates at the currency level inside each loss reserving scope is essential to get to grips with the intricate issue of currencies' cash flows prediction in a reinsurance business.

It is well known that in non-life (property and casualty) loss reserving, relevant reserve segments must result from a compromise between the conflicting goals of obtaining homogeneous risk groupings and achieving a sufficient volume of data for each group. Already laid out AGRe's Loss Ratio segments, combining many accepted risks from ceding entities with similar characteristics and claims development behavior, were designed with this exact aim of achieving a proper combination of volume and homogeneity. That reserving segmentation, currently in force at AGRe, thus consists of homogeneous and sufficiently populated groups of risks, on the basis of which actuarial estimates are established, among which ultimate loss and IBNR reserves estimates.

Each LR scope's IBNR reserves are estimated by specific methods, according to their distinctive and individual features. Some evaluation methodologies (see section 3.2) being more suitable than others for a given LR scope's properties. Independently of the chosen reserving method, the resulting IBNR reserves estimated at the macro level (LR scope x UDWY: by LR scope and underwriting year) need to be allocated down to a more detailed level: by LR scope, underwriting year and currency. Indeed, as stated previously, to address the complex issue of anticipating future FX exposure, one is required to at least have a good grasp of one's current liabilities in separate currencies. To achieve this goal, it is then a precondition to devise IBNR estimates at a finer level of detail than the currently defined reserves segments.

In addition, underwriting year results (including IBNR) at the individual program, account or reinsurance treaty level can be required in order to properly manage the business. Those underwriting year results at a low aggregation level can then be compared with the corresponding prior estimates produced by the pricing actuaries.

AGRe reserving actuaries currently split LR scope x UDWY's IBNR by treaty:

- at the prorata of earned premium for treaty with a young maturity (below or equal to 8 years old from their subscription year);
- at the prorata of case reserves for treaty with an old maturity (above 8 years old from their subscription year).

To legitimate this treaty's maturity threshold of 8 years, the time lag, for each treaty x section, between the UDWY january 1st and the first reported claims assessments has been analysed. 8 years thus corresponds to the quantile 97% of this time interval (UDWY to first claim report) distribution, and seemed an appropriate threshold to consider by reserving actuaries.

For a given underwriting year and a given adequate risk breakdown structure (such as LR scope x Currency), there exist two common methods for allocating IBNR:

- **the earned premium method:** assigns IBNR for each risk subset in proportion to the earned premium of that subset in view of the aggregate premium earned by the collective set. This method does not take into account the possibility that certain individual risk classes may experience greater claim frequency, paid loss ratio, and case-incurred loss ratio, and thereby justify a larger proportion of the aggregate IBNR reserve, than others within the same encompassing set.
- **the case-incurred loss method:** allocates IBNR in proportion to the underlying individual case-incurred loss amount, which is equivalent to applying an identical cumulative loss development factor to the case-incurred losses for each sub components of the collective considered. This method may induce unstable and unreliable allocations, all the more for recent underwriting years and long-tailed reserve segments.

Each of these two methods therefore have drawbacks:

- for immature UDWY, the emergence of sporadic claims invalidate the case-incurred based allocation in favor of the earned premium one. Those early case-incurred loss may still contain information we would like to take into account when sharing out IBNR down to finer level of detail.
- for older UDWY, case-incurred loss allocation method will obviously be favoured. Nevertheless, it wouldn't be wise to totally discard prior knowledge provided by the earned premium.

The new IBNR allocation approach presented in next subsection has been developed to answer those limitations and take heed of the exposed considerations.

7.3.2 New allocation methods

First of all, let us stress that we did not seek to challenge procedures producing current amounts of aggregate IBNR reserves. As developed in this chapter, one could have resorted to credibility theory through the Chain-Ladder credibility model, not only to determine credible claims development (as we did) but also to specify credibility based reserves estimates for any wished aggregation level. We did not attempt to offer new estimates of IBNR at the LR scope x UDWY level, for the simple reason that we did not look for changing existing procedures and methods, which are based on complex processes, years of practice and experts' judgments that could not possibly be reflected by a direct application of the Chain-Ladder credibility model. We therefore assume that the projected ultimate loss amounts for each LR scope x UDWY (and the corresponding IBNR reserves) are issued from sound and specifically adapted loss reserving methodologies, which are not relevant to the IBNR allocation procedure.

With the only ambition to suggest a new allocating approach of already defined IBNR aggregate amounts, we decided to adopt the methodology proposal imparted in the paper "The modified Bornhuetter-Ferguson approach to IBNR allocation" [9]. In addition, we provide an add-on to this approach, exploiting the outputs of previous section, that is, credibilized development patterns.

The "Modified Bornhuetter-Ferguson (BF)" allocation method offers a more reasonable and stable alternative to the earned premium and case-incurred methods, described previously. This procedure involves a credibility-weighted average of the earned premium and case-incurred allocation bases. This combined allocation, as a compromise between allocating IBNR solely based on either earned premium or case-incurred, has thus an edge on both methods. The relative weights assigned to each of the two methods depends on the treaty underwriting year: for most recent UDWY, most of the weight will be attributed to the earned premium allocation, and as the maturity of treaty increases, more weight is transferred to the case-incurred allocation.

As in the case of the traditional BF reserving method, the proper weighting between the earned premium and case-incurred based allocations depends on the loss development factor (LDF: meant to adjust claims to their ultimate projected level) proper to the risk class (Level 0) and the maturity (current accounting year - UDWY) considered. The modified BF approach calculates an "implied LDF" for each underwriting year and risk class, defined as the ratio of projected aggregate ultimate losses to aggregate case incurred losses. According to traditional BF reserving formula, for each Level 0 x UDWY risk class, the weight assigned to the case-incurred based allocation is equal to the reciprocal of the corresponding implied LDF. In other words, the weight $\alpha(UDW, Level\_0)$ is equal to the claims development rate

$$Z_{current\_year-UDWY+1,Level\_0} = \frac{1}{LDF_{current\_year-UDWY+1,Level\_0}}$$

The weight given to the earned premium allocation method is then equal to the complement of the case-incurred one.

An alternative consists in extracting LDF's from the underlying Level 0 case incurred loss triangles: one LDF for each Level 0 x UDWY risk class. It represents an excellent opportunity to avail ourselves of credibilized development factors: as best linear estimates of true individual development factors given Level 0 x Level 1 as well as Level 0 loss triangles' observations, they can lead to reasonable LDF's and development patterns coefficients estimates for each Level 0 x Level 1 x UDWY risk class. On one hand, using those "credibilized weights" grants proper individualized weighing (at Level 0 x Level 1 granularity) which better reflects specific claims development behaviors at the lower level. On the other hand, those weights are no more constant at the Level 0 granularity and a rescaling of allocated IBNR must be performed.

Let us consider one individual element of Level 0 x Level 1 = LR scope x Currency (for example LiabilityLong x USD), and a given underwriting year $UDWY_i$. We denote:

- $Premium_{i,0}$: aggregate premium earned by all the reinsurance treaties written in $UDWY_i$ and comprised in the considered Level 0 element's scope (LiabilityLong LR scope to continue with our example);
- $Premium_{i,1}$: premium earned by all the reinsurance treaties written during year $UDWY_i$ and contained in the considered Level 0 x Level 1 element's scope (LiabilityLong x USD);
- $Case\_Incurred_{i,0}$: aggregate case incurred loss suffered by all the reinsurance treaties written in $UDWY_i$ and composing the considered level 0 element's scope;
- $Case\_Incurred_{i,1}$: case incurred loss related to Level 0 x Level 1 x $UDWY_i$ element's scope;
- $IBNR_{i,0}$: aggregate IBNR reserves estimated for the considered Level 0 x $UDWY_i$ element's scope.
- $IBNR_{i,1}$: IBNR reserves at the finer level (Level 0 x Level 1 x $UDWY_i$), to be determined.

The "Modified Bornhuetter-Ferguson" allocation approach boils down to:

1. Chose an appropriate hierarchical structure breakdown (in our case LR scope x Currency);
2. Allocate IBNR in proportion to the earned premium for each $UDWY_i$ and each lower level element:
$$IBNR_{i,1}^{premium} = \frac{Premium_{i,1}}{Premium_{i,0}} \cdot IBNR_{i,0}$$
3. Allocate IBNR in accordance with the case-incurred loss for each $UDWY_i$ and each lower level element:
$$IBNR_{i,1}^{case\_incurred} = \frac{Case\_Incurred_{i,1}}{Case\_Incurred_{i,0}} \cdot IBNR_{i,0}$$
4. Calculate the implied (or credibility based) LDF and the respective relative weights
$$\alpha_{i,0}^{implied} = \frac{1}{LDF_{i,0}^{implied}} = \frac{Case\_Incurred_{i,0}}{Case\_Incurred_{i,0} + IBNR_{i,0}} = Z_{i,0}^{implied}$$
or
$$\alpha_{i,1}^{cred} = \frac{1}{LDF_{i,1}^{cred}} = Z_{i,1}^{cred}$$
5. calculate the weighted-average IBNR allocation for each $UDWY_i$ and each lower level element
$$IBNR_{i,1}^{BF\_implied} = \alpha_{i,0}^{implied} \cdot IBNR_{i,1}^{case\_incurred} + (1 - \alpha_{i,0}^{implied}) \cdot IBNR_{i,1}^{premium}$$
or
$$IBNR_{i,1}^{BF\_cred} = \alpha_{i,1}^{cred} \cdot IBNR_{i,1}^{case\_incurred} + (1 - \alpha_{i,1}^{cred}) \cdot IBNR_{i,1}^{premium}$$

This allocation method combines two elements of information by considering both experience-based estimate (relative underwriting results to date via the case-incurred loss allocation) and an a priori estimate (size of the underlying risk class via the earned premium allocation). It can further be improved by incorporating individualized credibility based development patterns information through the the weights calculation.

Adjustment of the an priori loss ratio by Level 0 x Level 1 elements' scopes is not allowed by the modified BF approach. For a recent UDWY of a long-tailed Level 0 reserve segment, and a predetermined ultimate loss ratio (and aggregate IBNR), the case incurred loss amount should be low and the modified BF method would allocate IBNR largely in proportion to earned premium, which would be equivalent to apply the same loss ratio to all underlying Level 0 x Level 1 elements. In the case that for the considered Level 0 risk class (LR scope = LiabilityLong for example), the corresponding Level 0 x Level 1 risk classes (LR scope x Currency = LiabilityLong x EUR, LiabilityLong x USD, LiabilityLong x CHF, ...) have very different expected levels of profitability, then one could replace the earned premium portion of the allocation with an "expected loss" allocation.

Another possible pitfall of the BF allocating approach concerns old UDWY, for which aggregate IBNR will largely be assigned in accordance with case incurred losses. For these old UDWY, certain underlying individual risk classes will show paid losses close or equal to case-incurred losses (i.e. file-to-file reserve close or equal to zero) and most or all claims closed. Even so, the modified BF method may allocate a large proportion of the remaining aggregate IBNR to these lower level risk classes. If they are liable to be subject to very late-reported claims, or reopened claims, this allocation may still be appropriate. If not, one must handle these particular cases separately. Concretely, while allocating AGRe's estimated aggregate IBNR down to LR Scope x Currency granularity level, we decided to distribute no IBNR to all (but a few long-tailed exceptions) individual classes of accepted risk with a maturity older than 8 years and no existing case reserves.

7.3.3 Net IBNR allocation and algorithm

As we will see in the last part of this thesis, replicating portfolios, representing Best Estimate Liabilities in each currency and their duration, are used at AGRe to assess and compute Market risks. They are in fact future cash flows emanating from AGRe net reserves, i.e. they represent the future net loss payments.

To project long-term futures cash flows, we will then mainly consider net cash flows originating from today net reserves (net of retrocession). We thus have to transform the new gross IBNR allocation obtained earlier into net IBNR amounts by LR scope x Currency.

The steps of the implemented IBNR allocation algorithm are as follow:

INPUT: Accepted and Retroceded risks' portfolios data, containing, among others, variables giving the earned premium, paid claims, case and IBNR reserves by reinsurance treaties in acceptation as well as in retrocession. Other accounting and contractual variables displayed in 5.1.2 are also present. Some reinsurance treaties with clearly identified characteristics are maintained out of this study's risk perimeter.

1. Computation of the acceptation side's earned premium, paid claims, case reserves and IBNR by LR scope x Currency x UDWY together with the computation of the same quantities at the collective LR scope x UDWY granularity level.

2. Application of the chosen IBNR allocation methodology (implied or credibilized LDF) as developed previously (IBNR amounts estimated at the LR scope x UDWY level by the reserving team, allocated down to the LR scope x Currency x UDWY granularity level).

3. Post-treatments to manage peculiar and exceptional cases.

4. IBNR reserves of LR scope x Currency x UDWY older than 8 years classes for which there are no case reserves, are set to 0. Other individual classes' assigned IBNR amounts are then rescaled to keep the same overall IBNR amounts at the LR scope x UDWY collective level.

5. The new gross IBNR allocation thus derived is further splited by reinsurance treaties in proportion to earned premium for recent treaties, and according to case reserves for old treaties (same threshold of 8 years). Reinsurers' signed retrocession share to each acceptation treaties can then be taken into account to get IBNR amounts net of specific retrocession but gross of Group covers. Those amounts are aggregated back to LR scope x Currency x UDWY level.

6. Steps 1, 2, and 3 are performed again, this time to the ceded risks portfolio, in particular to treaties subject to group cover only, in order to quantify the IBNR reserves retroceded in the scope of Group covers in accordance with the new chosen allocation method.

7. Group covers' ceded IBNR are added (values signed according to accounting conventions) to IBNR amounts net of specific retrocession and gross of Group covers, to finally achieve net IBNR amounts by LR scope x Currency x UDWY proceeding from the new allocation procedure.

8. Net file-to-file reserves amounts are more easily retrieved, just adding accepted risks' case reserves and ceded ones at the LR scope x Currency x UDWY granularity level.

OUTPUT: Net reserves amounts (newly allocated net IBNR + net case reserves) by LR scope x Currency x UDWY.

8 Technical reserves projection into cash flows and AGRe's intern FX risk measure model under Solva II

We now have developed a procedure based on a mix of credibility theory and Bornhuetter-Ferguson method allowing us to not only retrieve loss development rhythm but also to assign IBNR at the finer than currently considered granularity, namely LR SCOPE x CURRENCY in place of just LR SCOPE. Having a better view on AGRe's liabilities distributed over its range of reserving segments and currencies, and being more confident with the patterns of loss and payments development (and thus the patterns of reserves extinction) for each of those currencies inside each LR scope, we can finally try to project the future yearly cash flows issued from technical reserves as of date.

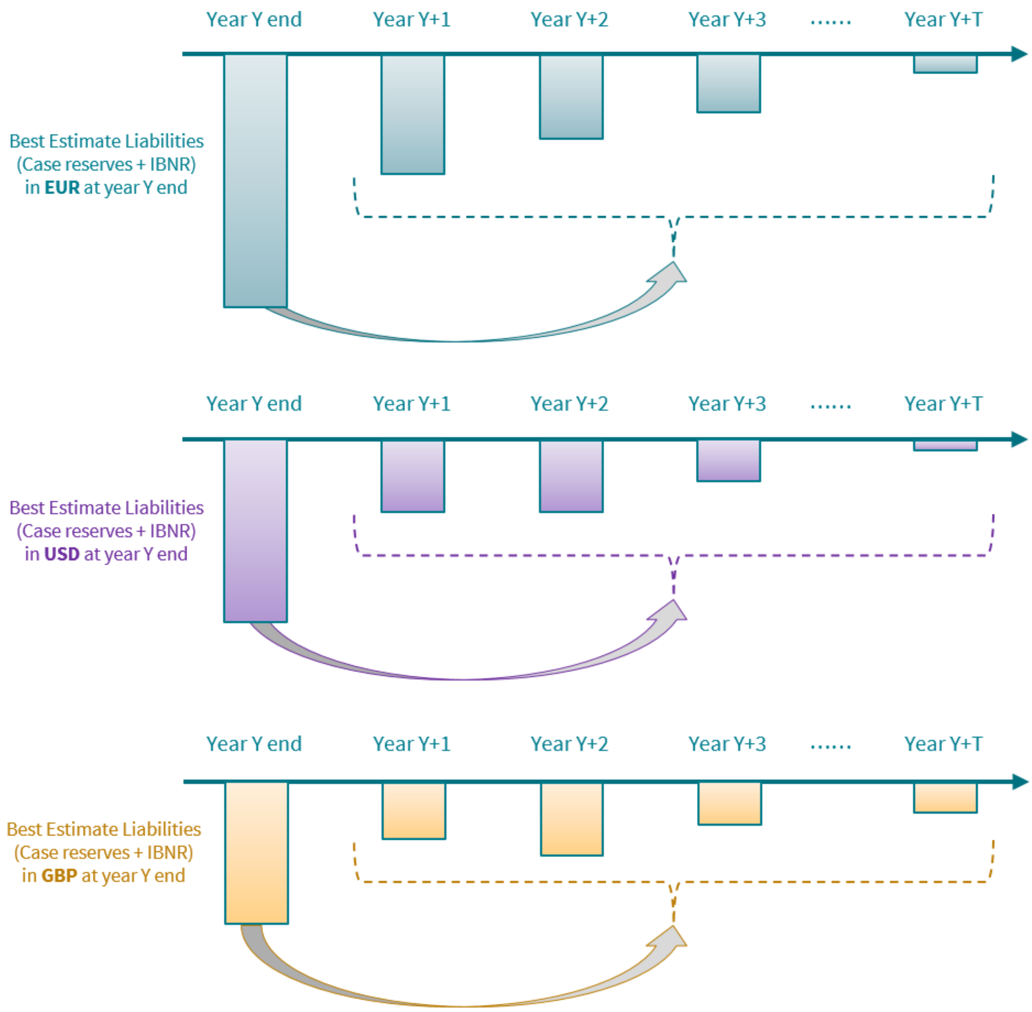


Figure 85: Projection illustration of Best Estimate Liabilities in currencies into yearly cash flows

8.1 Market risk simulation and liability modeling at AGRe

A projection of future yearly cash flows in currencies emanating from existing technical reserves as of date is already carried out by the Risk Management team of AGRe, mainly to figure out the market risk linked to AGRe's run-off liabilities.

8.1.1 STEC (Short-Term Economic Capital) methodology for market risk

The STEC, equivalent to the Solvency Capital Requirement (SCR), is issued from the internal model of AXA and is based on a one-year horizon value-at-risk (VaR) at a 99.5% confidence level. It is computed through an instantaneous shock on the Solvency II balance sheet, where the scenarios are calibrated to reflect extreme events that could potentially occur over a one-year time horizon. Market risk considers the risks of AXA's economic balance sheet (assets minus liabilities) towards movements in the financial markets, both in liquid markets (equity, interest rates, spread, FX, implied volatility and inflation) and in illiquid markets (real estate, hedge fund and private equity). Actual defaults for fixed income instruments, mortgages, receivables and reinsurance counterparties, are not considered in the market risk scope but in the credit one (Credit STEC). To sum up, the STEC computed at year Y end is the amount of capital which should ensure (in 99.5% of cases) the solvency of the undertaking at year Y+1.

To facilitate the assessment of the STEC associated to the stochastic components of the Best Estimate Liabilities (BEL), the dynamics of the liability cash flows are approximated using a set of standard financial instruments. The crucial point is that this portfolio of basics financial instruments has the same market value and sensitivities as the liabilities in local stochastic projection model. Therefore the replicating portfolio reacts in the same way to changes in the underlying risk factors. This portfolio is called a replicating portfolio as it replicates the dynamics of the market value of the liabilities. In other words, a replicating portfolio is a description of the liabilities (values, durations and sensitiveness) with standard financial instruments such as zero coupon bonds, equities, equity indexes and derivatives (caps, receiver and payer swaptions, equity options, inflation swaps, etc.).

In general, liabilities are thus modeled as financial instruments, which can either be a set of short zero coupon bonds or a more complex replicating portfolios containing options. In the P&C business it is normally sufficient to model the liabilities as a set of short zero coupon and zero coupon inflation linked bonds. For the life business a replicating portfolio with more complex financial instruments is used in order to allow for the options embedded in the liabilities.

In the AGRe's process of market risk measurement, the swap curve plus an adjustment for the volatility adjuster, the credit risk adjustment and the ultimate forward rate are then used to discount the liabilities. In a Monte Carlo simulation, the level of interest rates and the volatility adjuster are shocked, after what the BEL is re-assessed using the shocked Solvency II discount curve.

We will once again pass over the Life business perimeter of AGRe and focus only on its P&C business and corresponding liabilities replicating portfolios.

8.1.2 Creation of P&C replicating portfolios as currently performed at AGRe

We describe hereafter the current methodology for the creation of P&C liabilities replicating portfolios (denoted RP) used to compute market risks in the STEC framework of AXA Global Re.

As we saw, the P&C replicating portfolios stand for AGRe's P&C business BEL in each currency, integrating their durations. The P&C RP's are in fact a translation of the future cash flows in currencies stemming from AGRe net multi-currency reserves, i.e. they represent future net loss payments. They also include the modeled BE of the Pools and Group covers future losses and premiums by currency for the renewal year, as well as future cash-flows relative to ULAE reserves and cash deposits. Other cash flows from renewals and from new business are not considered.

Sensitivity to inflation assimilation

For each cash flow one can specify, through a factor, how sensitive the liabilities are linked to movements in inflation (0%, 25%, 50%, 75%, 100%). Depending on this factor, the expected cash flow is partially modeled as zero coupon (ZCB) bond and partially as inflation linked zero coupon bond (ILZCB).

For example, let's assume a liability cash flow, taking place in future year T, is strongly but not fully linked to inflation. In such a case a factor of 75% can be seen as adequate for modeling the inflation sensitivity of the P&C liabilities. 25% of the resulting cash flow would then be modelled as zero coupon bond and 75% as inflation linked zero coupon bond.

First, the cash flow in year T is partially modelled as ZCB with maturity T and with a notional equal to 25% of the expected cash flow amount. The initial market value of the ZCB is then derived in the following way:

$$MV_0 = \frac{25\% \cdot CF_T}{(1 + s_{T,0})^T}$$

In the Monte Carlo procedure, the Solvency II reference rate ($s_{T,0}$) is shocked and the price of the zero coupon bond is re-evaluated in each scenarios.

The remainder of the cash flow is modelled as ILZCB with maturity T and with a notional equal to

$$N = \frac{75\% \cdot CF_T}{(1 + i_{T,0})^T}$$

where $i_{T,0}$ is the current inflation expectation for term T.

As the expected cash flow of a ILZCB amounts to $N \cdot (1 + i_{T,0})^T$, the expected cash flow in year T amounts to 75% of the initial expected cash flow. This means that the expected cash flow remains unchanged when modelling it with an ILZCB compared to a ZCB. The same is the case for the initial market value, which is computed as

$$MV_0 = N \cdot \frac{(1 + i_{T,0})^T}{(1 + s_{T,0})^T} = \frac{75\% \cdot CF_T}{(1 + s_{T,0})^T}$$

In the Monte Carlo scenarios the inflation expectation curves ($i_{T,k}$), and the Solvency II reference rate ($s_{T,k}$) are shocked and the market value of the ILZCB is re-evaluated as:

$$MV_k = N \cdot \frac{(1 + i_{T,k})^T}{(1 + s_{T,k})^T}$$

A multi-step process

Currently, AGRe's P&C replicating portfolios are constructed (through the execution of a R algorithm) on the basis of the following step-by-step process:

INPUTS: AGRe's NET (Acceptance + Retrocession views) contractual and general accounting database providing a run-off net (of retrocession) state of AGRe's business as of date (date of cash flows projection and Replicating portfolios creation), and particularly the fields needed to compute the net BE reserves (net case reserves + net IBNR) by reserving segment, currency and underwriting year (LR SCOPE x CURRENCY x UDWY); Aggregate cumulative and strictly increasing payment patterns (PP) by LR scope; FX rates as of date to be used for the conversion of BE reserves EUR amounts into local currencies amounts (value of 1 € in the local currencies); Pool/Group Cover modeled next year losses and premiums by currency and LR scope.

1. *(Cash flows from run-off losses)* Specification of the NET database's variables which compose the Best Estimate reserves and sum their values (NET_PROV_SAP_CLO + PROV_DD + NET_PROV_IBNR_CLO + PROV_IBNR_BE i.e. at-previous-closing estimated and additional case reserves and IBNR) to retrieve the BEL valued in EUR aggregated by LR SCOPE x CURRENCY x UDWY.
2. *(Cash flows from run-off losses)* Computation of the ultimate loss by LR SCOPE x CURRENCY x UDWY on the basis of corresponding reserves' amount and maturity together with the related LR SCOPE payment pattern.

3. *(Cash flows from run-off losses)* Cash flows projection (i.e the amounts paid along the time needed to cover the full reserve) by LR SCOPE x CURRENCY x UDWY on the basis of the related LR scope PP increments (taking reserves maturity as a starting point) and previously calculated ultimate loss (exception of maturities higher than the length of related PP in which case the whole reserve is considered to turn into effective claims payments next year).

4. *(Cash flows from expected premiums and losses on the pools / Group covers)* Addition to the first projected year already calculated cash flows of expected (modeled) net premiums earned under the pools / Group covers for the following year. Projection of the expected (modeled) losses borne by the pools / Group covers for the following year in the same manner as the run-off losses (same extinction pattern as the technical reserves).

5. *(Cash flows from ULAE reserves)* Quantification of future cash flows issued from ULAE reserves. To be consistent with the fact that both run-off losses and pools / Group covers expected future losses and premiums have been projected, both ULAE relative to run-off losses (computed by the reserving department) and Future ULAE (computed by Risk Management as an adjustment for Contract Recognition in AFR) also have to be projected. The sum of those two elements, $ULAE_{TOT} = ULAE_{run-off} + ULAE_{future}$, is then projected for year $Y + i$ as follow:

$$ULAE_{TOT_i} = \frac{|\mathrm{CF}_i|}{\sum_j |\mathrm{CF}_j|} \cdot ULAE_{TOT}$$

   Where CF_j represents the sum of cash flows valued in EUR for all currencies in year $Y+i$.

6. (Cash flows from cash deposits) To have a better picture for STEC Market calculation of AGRe's Economic Balance Sheet, the net cash deposits (collaterals) are incorporated in the replicating portfolios, assuming they have the same extinction pattern as the technical reserves.

7. Aggregation of all future yearly cash flows (valued in EUR) by currency.

8. Conversion of cash flows amounts valued in EUR into local currencies in accordance with end of current year spot FX parities.

9. Reconciliation at an aggregated level discussed and validated with the Group Risk Management and the Chief Risk Officer, to ensure that there is a match between RP's figures and the total amounts (extracted from row data input files) of net technical reserves, ULAE reserves, and expected losses and premiums for next year (on pools / Group covers).

OUTPUT: File used for RP computation, containing future projected yearly net cash flows amounts by currency valued in local currency.

We can consider the following numerical dummy example to better understand the current process of predicting cash flows from run-off losses (i.e. constituted technical reserves). Let us say that we are at year 2021 end (closure) and let us take the particular risk group LR SCOPE x CURRENCY x UDWY = Engineering x USD x 2019. We estimated that the total net of retrocession technical reserves (Case + IBNR) to put aside for this perimeter is 1M USD, that is around 880k EUR evaluated at 2021 end. Summing over all relevant fields, the AGRe's NET contractual and accounting database will give this amount of 880k EUR aggregated over the scope Engineering x USD x 2019. We seek to project over the years to come and according to already established methodology (above process) the cash flows in USD issued from this risk subgroup reserves as of 2021 closure (run-off losses part). The maturity of those reserves is $2021 - 2019 + 1 = 3$ years. We suppose that the payment pattern relative to the Engineering LR SCOPE is given by Table 29 below. Since the cumulative PP at the third development period is equal to 0.55 and the reserves with a maturity of 3 years have been evaluated at about 880k EUR, we deduce that the estimated ultimate loss for this perimeter is

$$\text{Ultimate run-off loss} = 880\text{k EUR} \cdot \frac{1}{1-0.55} \simeq 1.95\text{M EUR}$$

We then multiply this ultimate loss by PP increments starting at development year 3 + 1. We therefore obtain the future yearly cash flows valued in EUR (we check that those projected outflows of cash with negative accounting sign sum in absolute value to the initial reserves stock, i.e. here -880k EUR). Converting into the local currency (here USD) in keeping with the spot EUR/USD FX rate at 2021 end ($\simeq 1.14$), we finally get future cash flows in USD emerging from Engineering x USD x 2019 technical run-off reserves as of year 2021 closing.

Development Period	Payment Pattern	PP increments	Year of projection	Projected cash flows (in k EUR)	Projected cash flows (in k USD)
1	0.15	0.15			
2	0.4	0.25			
3	**0.55**	0.15	Y (2021)		
4	0.7	0.15	Y+1	-293.25	-333.43
5	0.8	0.1	Y+2	-195.50	-222.28
6	0.85	0.05	Y+3	-97.75	-111.14
7	0.9	0.05	Y+4	-97.75	-111.14
8	0.93	0.03	Y+5	-58.65	-66.69
9	0.95	0.02	Y+6	-39.10	-44.46
10	0.97	0.02	Y+7	-39.10	-44.46
11	0.98	0.01	Y+8	-19.55	-22.23
12	0.99	0.01	Y+9	-19.55	-22.23
13	0.995	0.005	Y+10	-9.78	-11.11
14	1	0.005	Y+11	-9.78	-11.11
			TOTAL	**-879.75**	**-1000.28**

Table 29: Numerical dummy example of currently carried out projection method of technical run-off reserves into yearly cash flows, as of year 2021 end, for LR SCOPE X CURRENCY X UDWY = Engineering X USD X 2019

To synthesize, in this framework the RP's are built from future yearly cash flows in currencies originating from run-off liabilities (eventually turning into claims payments) and complementary components (ULAE reserves, Pools / Group covers modeled future losses and premiums, and net cash deposits).

Let us assume that sticking to this tread, we hypothetically come to the following projection of aggregated cash flows in USD valued in USD: $CF_{USD,Y+1} = -3M$, $CF_{USD,Y+2} = -2.7M$, $CF_{USD,Y+3} = -2.2M$, $CF_{USD,Y+4} = -1.8M$, ..., $CF_{USD,Y+T} = -0.3M$. Disregarding inflation effect we could then set up the following combination of ZCB as a replicating (i.e. approximately equivalent in terms of amounts, durations and sensitiveness) portfolio for those futures cash flows: $-3M \cdot ZCB_{USD,6\text{ months}} - 2.7M \cdot ZCB_{USD,1\text{ year and 6 months}} - 2.2M \cdot ZCB_{USD,2\text{ years and 6 months}} - 1.8M \cdot ZCB_{USD,3\text{ years and 6 months}} - ..... - 0.3M \cdot ZCB_{USD,\text{T-1 years and 6 months}}$. $ZCB_{USD,\text{T}}$ corresponding to an inflow of 1 USD at maturity T. Middle years maturities are considered here, as is done in practice at AGRe. We indeed assume that projected cash flows are homogeneous within a given future year of projection, and a ZCB with a middle year maturity is a satisfactory way to reflect a homogeneously distributed cash flow aggregate amount over that year.

8.2 New approach implementation and historical backtest

8.2.1 Technical reserves new projection methods

In this subsection we challenge the projection methodology just presented and currently employed in the context of the replicating portfolios, and offer alternative methods of case and IBNR run-off reserves projection. We seek to implement possible projection method substitutes in order to take into account the a priory different behaviors in the liquidation patterns of file-to-file reserves and IBNR ones. Those new methods endeavor to take advantage of the information contained in both claims payment and claims incurred patterns at a chosen level of data aggregation.

First of all, let us have a quick look on the reserves liquidation patterns, that is the manner in which reserves (file-to-file and IBNR) transform into effective claims payments over time. For a given risk group, case and IBNR reserves liquidation patterns are easily recovered from payment (claims paid) and loss (claims/case incurred) development patterns:

- case (file-to-file) reserves liquidation pattern = loss development pattern - payment pattern,

- IBNR reserves liquidation pattern = 1 - loss development pattern

Those patterns can be credibilized with respect to a broader risk group (higher level of data aggregation) as we saw in subsection 7.2, or not.

Below are displayed the credibilized development patterns of incurred losses, claims payments, file-to-file reserves as well as IBNR reserves, for some example classes of Level 0 (SOLVENCY II BUSINESS LINE) x Level 1 (LR SCOPE) (could have been LR SCOPE x CURRENCY).

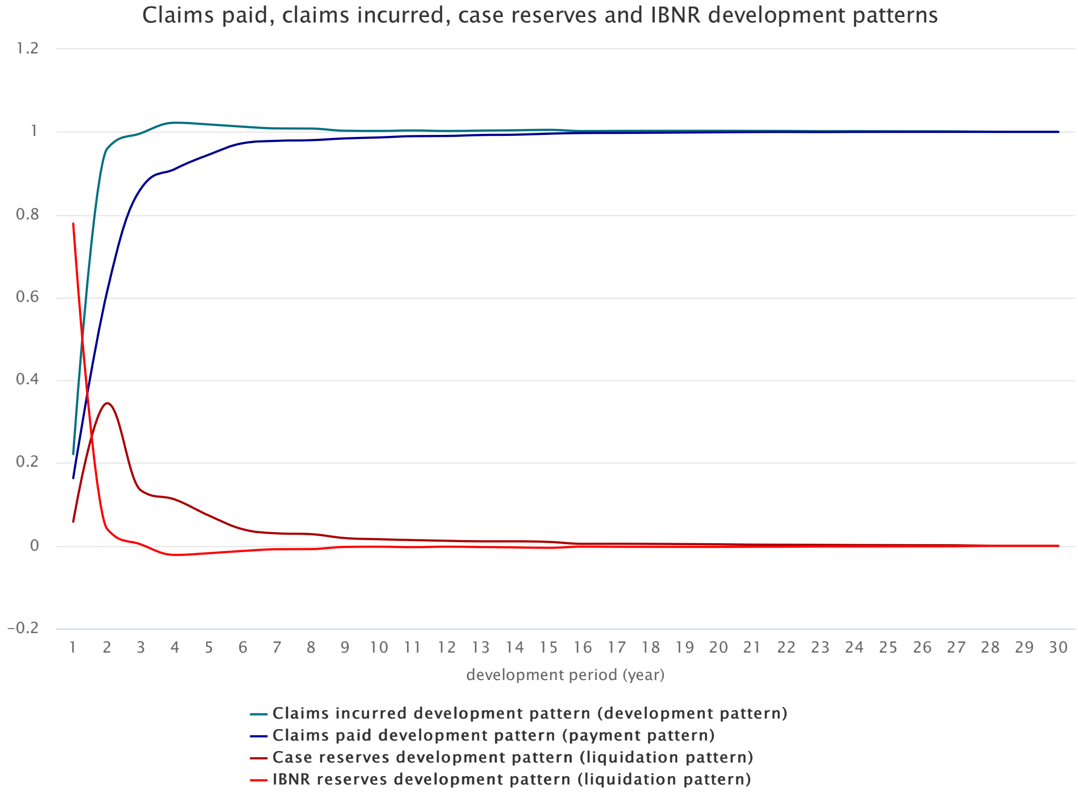


Figure 86: Credibilized developments of claims incurred, claims paid, case reserves, and IBNR – Level 0 (SOLVENCY II BUSINESS LINE) value X Level 1 (LR SCOPE) value = Non-Proportional Property Reinsurance X PoolPropertyParEvt

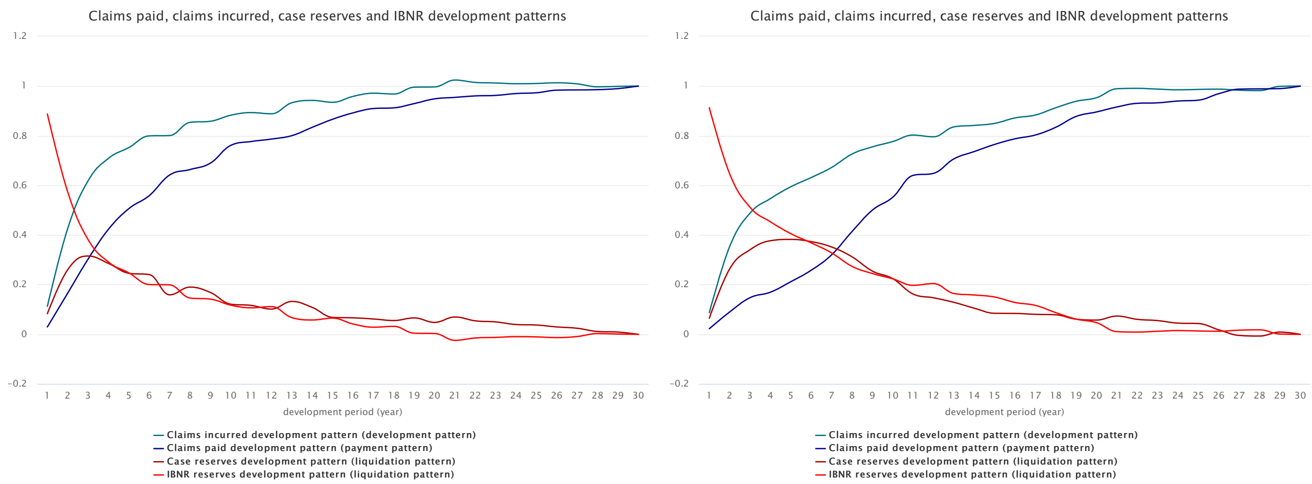
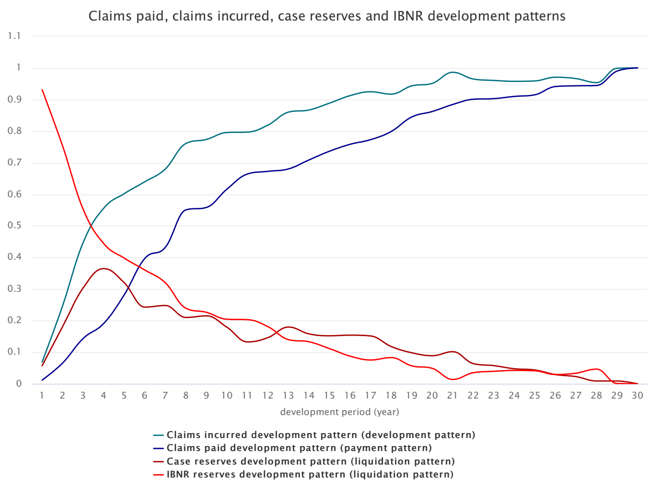


Figure 87: Credibilized developments of claims incurred, claims paid, case reserves, and IBNR – Level 0 (SOLVENCY II BUSINESS LINE) value X Level 1 (LR SCOPE) value = Non-Proportional Casualty Reinsurance X Engineering (on the left) and Non-Proportional Casualty Reinsurance X LiabilityLong (on the right)



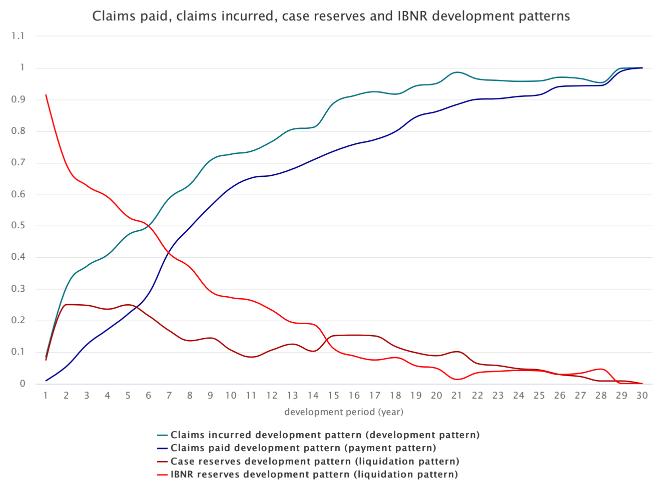


Figure 88: Credibilized developments of claims incurred, claims paid, case reserves, and IBNR – Level 0 (SOLVENCY II BUSINESS LINE) value X Level 1 (LR SCOPE) value = Non-Proportional Casualty Reinsurance X PoolLiability (on the left) and Non-Proportional Casualty Reinsurance X PoolMotor (on the right)

In addition to the existing one, three reserves projection functions have been designed as an attempt to take into consideration a more subtle and reliable approximation of technical reserves extinction rhythms, whether file-to-file or IBNR. All three proposals of projection method take as arguments:

- the accounting year of reference, i.e accounting closure at which we place ourselves to perform the projection (we inform 2021 if we are at year 2021 end and we seek to anticipate future run-off cash flows from year 2021 closing reserves, prior years can be considered for historical backtesting purpose);
- a data table giving the file-to-file and IBNR reserves amounts by seniority (UDWY) for a given risk group (the data structure is then the risk group, a specific value of LR SCOPE x CURRENCY for example, the underwriting year, and the corresponding aggregate amounts of case reserves and IBNR reserves);
- the corresponding loss incurred development pattern for that same risk group (credibilized or not);
- the corresponding payment pattern for that same risk group (credibilized or not);

The three projection procedures are based on the assumption that IBNR run-off incremental amounts (over each development year) transform entirely into case reserves. They differ on the way they handle the transformation of case reserves into cash flows (claims payments). Let us describe briefly the first method, the two others only adding layers of complexity on top of it.

Projection method 1:

⇒Applied on a particular homogeneous risk group (can be looped over several LR SCOPE or combinations of LR SCOPE x CURRENCY, or any other predefined sets of reinsurance treaty).

- *Hypothesis 1: All IBNR reserves turn into FF (file-to-file) reserves, according to their extinction pattern.*

- *Hypothesis 2: FF reserves maturity is based on UDWY.*
- *Hypothesis 3: FF reserves liquidate uniformly and totally over a soon to be defined liquidation duration.*

1. Extraction of IBNR development pattern from the loss development pattern passed as an input. IBNR development pattern capped at 0 (to remove negative values) or not;
2. FF reserves liquidation duration computed as a barycentre of the input payment pattern increments (if the payment pattern of the considered risk group is 0.1, 0.4, 0.6, 0.75, 0.85, 0.9, 0.95, 0.98, 1, the related increments are 0.1, 0.3, 0.2, 0.15, 0.1, 0.05, 0.05, 0.03, 0.02 and the liquidation duration will be equal to 0.1 x 1 + 0.3 x 2 + 0.2 x 3 + 0.15 x 4 + 0.1 x 5 + 0.05 x 6 + 0.05 x 7 + 0.03 x 8 + 0.02 x 9 = 3.47 which will be rounded to 3 years.
3. For each UDWY, the maturity of the associated reserves is defined as: accounting year of reference - UDWY + 1;
4. According to their maturity and their rescaled development pattern, IBNR reserves are first projected into FF reserves (special cases must be handled), i.e. they are distributed over the years succeeding the accounting year of reference;
5. Once the IBNR have been splited into FF reserves springing out over future years, the existing FF reserves as well as those projected ones are then spread out homogeneously over the liquidation duration previously calculated, which finally gives us projected cash flows amounts.

The numerical example in the table below illustrates this procedure for one hypothetical risk group with the displayed dummy IBNR development pattern, and a liquidation duration of 3 years (reserves and cash flows amounts are expressed in the unit of choice).

Development period (year)		1	2	3	4	5	6	7	8	9	10
IBNR dev pattern		1	0.9	0.6	0.4	0.25	0.15	0.1	0.05	0.02	0
IBNR dev pattern increments		0.1	0.3	0.2	0.15	0.1	0.05	0.05	0.03	0.02	0
UDWY = 2020 i.e maturity = 2 years											
IBNR dev pattern (maturity 2 years)		0.9	0.6	0.4	0.25	0.15	0.1	0.05	0.02	0	0
IBNR dev pattern rescaled (maturity 2 years)		1.000	0.667	0.444	0.278	0.167	0.111	0.056	0.022	0.000	0.000
IBNR dev pattern increments rescaled (maturity 2 years)		0.333	0.222	0.167	0.111	0.056	0.056	0.033	0.022	0.000	0.000
Year of projection	Y closure	Y+1	Y+2	Y+3	Y+4	Y+5	Y+6	Y+7	Y+8	Y+9	Y+10
Projection of IBNR into FF reserves	**-1000**	-333	-222	-167	-111	-56	-56	-33	-22	0	0
FF reserves (existing as of year Y end + projected from IBNR)	**-2000**	-333	-222	-167	-111	-56	-56	-33	-22	0	0
Projection of FF reserves into cash flows (payments of claims)		**-778**	**-852**	**-907**	**-167**	**-111**	**-74**	**-48**	**-37**	**-19**	**-7**

Table 30: Numerical dummy example of projection method 1 procedure

To pass from the penultimate line (FF reserves existing as of year Y end + FF reserves projected from IBNR) to the last one (cash flows), the amounts are divided by the liquidation duration (in this example 3 years) and spread evenly over 3 (liquidation duration) periods. For instance the cash flow amount -852 comes from -2000 x 1/3 - 333 x 1/3 - 222 x 1/3, and -111 = -167 x 1/3 - 111 x 1/3 - 56 x 1/3.

Table 30 shows solely the projection of reserves into cash flows for only one row of the hypothetical input data table (maturity of the reserves = 2 years). All reserves within that input data table, presenting different maturities, are projected according to that process within a R algorithm through matrix manipulations.

We will not detail the two other projection methods that have been tested, since they are built on the same foundations as the first one. They just integrate further subtleties and nuances in the manner the FF reserves are converted into effective cash flows, as attest the following assumptions shifts:

Projection method 2:

- *Hypothesis 1: All IBNR reserves turn into FF reserves, according to their extinction pattern.*
- *Hypothesis 2: FF reserves maturity is based on UDWY.*
- *Hypothesis 3: FF reserves liquidate entirely within the liquidation duration following the corresponding payment pattern.*

Projection method 3:

- *Hypothesis 1: All IBNR reserves turn into FF reserves, according to their extinction pattern.*
- *Hypothesis 2: FF reserves maturity is based on both UDWY and IBNR/UC (Ultimate Charge) ratio.*
- *Hypothesis 3: FF reserves liquidate entirely within the liquidation duration following the corresponding payment pattern.*

8.2.2 Performance assessment of future yearly cash flows prediction issued from the new procedure

In this subsection, we will appraise whether the new proposed procedure allows for a better prediction of cash flows in currencies over the years to come. For that, we will focus only on run-off technical reserves originated cash flows and we will confront the projections issued from the new procedure (as of a past accounting year closure) to the actually observed historical cash flows. We will resort to the currently applied procedure projections, as a reference point to rate the new strategy put in place in order to forecast long-term cash flows labelled in currencies.

As a reminder:

- The "old" (currently applied) approach to turn run-off technical reserves into future cash flows and build RP for market risk measure, consists in aggregating the amounts of reserves Best Estimate (FF + IBNR) by reserving segment and UDWY, and cast them forward according to their corresponding payment patterns (at the level LR SCOPE) and maturity. A theoretical ultimate loss by LR SCOPE x UDWY is computed as a proxy for this projection, and no differentiation in the behaviors of distinct currencies' development patterns is taken into account.
- The new stratagem dives into an underlying realm of details, trying to handle and to capture the particularities of homogeneous subgroups of risks valued in the same currencies. Basically, instead of the more general perspective adopted by the old projection (LR SCOPE), the new approach takes into consideration an additional sublevel (LR SCOPE x CURRENCY), and endeavors to squeeze from it relevant pieces of information for the

challenge of interest. It relies, first of all, on an a priori smarter allocation of IBNR reserves down to each currency inside each reserving segment, and exploits the theory of credibility to tailor related individualized development patterns (see chapter 7). Having erected those foundations, the new procedure can then take advantage of one of the three newly designed alternative projection methods to throw run-off and newly allocated reserves into cash transactions (based on newly individualized and credibilized development patterns).

While reserves net of retrocession are projected into cash flows, and will still be whether with the old or new procedure, the historical backtest is carried out with gross reserves. That is, for backtest purpose we restrict ourselves to the acceptance side of AGRe business without considering retroceded reserves, only accepted ones.
The data used for this backtest is comprised of a series of historical pictures of the treaty acceptance technical and contractual database described in subsection 5.1.2. To put it another way, it corresponds to a concatenation of past time slices between 2014 and 2020, each slice representing the entire contractual and general accounting state of AGRe's accepted risks at a particular time. From all those slices, we only kept year closings ones. Therefore, we have at disposition closure states (same variables as listed in 5.1.2, i.e. contractual and accounting elements: premiums, claims paid, case reserves, IBNR Best Estimate, ...) of all the reinsurance treaties (x sections) taken out by AGRe, and this for each year end from 2014 to 2020.

From this historical database it is possible to retrieve precious pieces of information such as the yearly aggregate amounts of claims paid proceeding from a given group (arbitrary level of aggregation) of risks subscribed before a given underwriting year.

Let us set forth the broad lines of this historical back-test algorithm:

1. Credibilization of payment and loss development patterns on the basis of the triangle databases and pre-established functions (at a Level 0 x Level 1 to be chosen, for instance SOLVENCY II BUSINESS LINE x LR SCOPE or LR SCOPE x CURRENCY);

2. Historical risk acceptation database filtering and pretreatment (granularity: TREATY x SECTION ID (x UDWy) x HISTORICAL EXERCISE CLOSURE);

3. Actualization of all historical amounts (valued in EUR as of each historical exercise year end FX rates) into EUR equivalent value as of 2020, in order to be able to compare on the same ground reserves stocks to claims payments, both being valued at different times on the basis of different FX rates;

4. New allocation of IBNR BE (granularity: LR SCOPE x CURRENCY x UDWY x HISTORICAL EXERCISE CLOSURE);

5. Projection of case reserves as well as newly allocated IBNR ones thanks to newly defined projection methodologies (granularity: LR SCOPE x CURRENCY x UDWY x HISTORICAL EXERCISE CLOSURE).

6. Projection of case reserves and IBNR (as estimated and allocated by reserving team) following the old procedure detailed in subsection 8.1.2;

7. We then obtain cash flows (valued in 2020 euro equivalent) by (SOLVENCY II BUSINESS LINE x) LR SCOPE x CURRENCY x UDWY x HISTORICAL EXERCISE CLOSURE (i.e. past year of reference), projected according to both the new and old strategies. We can afterwards decide to aggregate those projected future (with regard to the historical

exercise of reference) cash flows, to the extent we wish (by CURRENCY or LR SCOPE x CURRENCY for example);

8. The projected cash flows are compared to the actual amounts of claims settlements relative to underwriting years anterior or equal to the historical exercise of projection.

To evaluate the projected yearly cash flows predictions in relation with real ones, we made use of the same performance (or proximity) criteria as in subsection 5.4.3, namely the NMSE, NRMSE, NMAE and MAPE of projected versus actually observed yearly cash flows valued in 2020 euro equivalent.

Say, we place ourselves at year 2014 accounting closure period and consider a particular scope (group of treaties labelled in a given currency for instance). From the historical database we have at hand the actual settlements amounts over 2015, 2016, 2017, 2018, 2019 and 2020 coming from prior to 2014 end subscribed risks in the considered scope. Here we denote those claims payments $CF_{2015}\_observed$, ..., $CF_{2020}\_observed$. We moreover have at disposal the run-off technical reserves (newly allocated or not) as of 2014 end, linked to the studied group of risks and connected to the claims payments we just introduced. Those reserves (newly allocated or not) are projected in line with the old or new (say we opt for projection method 1) projection methods into

$CF_{2015}\_predicted\_\_old\_allocation\_\_old\_projection$, ...,
$CF_{2020}\_predicted\_\_old\_allocation\_\_old\_projection$, ...,
$CF_{2030}\_predicted\_\_old\_allocation\_\_old\_projection$, ...,
or
$CF_{2015}\_predicted\_\_new\_allocation\_\_old\_projection$, ...,
$CF_{2020}\_predicted\_\_new\_allocation\_\_old\_projection$, ...,
$CF_{2030}\_predicted\_\_new\_allocation\_\_old\_projection$, ...,
or
$CF_{2015}\_predicted\_\_old\_allocation\_\_proj1$, ...,
$CF_{2020}\_predicted\_\_old\_allocation\_\_proj1$, ...,
$CF_{2030}\_predicted\_\_old\_allocation\_\_proj1$, ...,
or
$CF_{2015}\_predicted\_\_new\_allocation\_\_proj1$, ...,
$CF_{2020}\_predicted\_\_new\_allocation\_\_proj1$, ...,
$CF_{2030}\_predicted\_\_new\_allocation\_\_proj1$, ...,

Hence, the performance of 4 model configurations (old vs new IBNR allocation and old vs new reserves projection methodology) have to be measured.

Since for the picture of time slice "as of 2014" we only have 6 real observation points of future yearly cash flows, we are only able to appraise the projection models (old and new) on the basis of those 6 points (if we place ourselves as of 2016 closure for example, it goes down to solely 4 possible comparison points). As for the NMSE metrics, we subsequently confront the quantities

$$\text{NMSE(model)}_{\text{as of 2014}} = \frac{\sum_{i=2015}^{2020} (CF_i\_observed - CF_i\_predicted\_model)^2}{\sum_{i=2015}^{2020} (CF_i\_observed)^2}$$

We boil those NMSE's (as of 2014 to as of 2018) down to a single performance indicator, computing their weighed average:

$$\text{NMSE(model)} = \frac{\sum_{i=2014}^{2018}\Big((2020-i)\cdot \text{NMSE(model)}_{\text{as of year i}}\Big)}{\sum_{i=2014}^{2018}(2020-i)}$$

The same is done with metrics NRMSE, NMAE and MAPE, and for each group of risks of interest. Furthermore, as in the rating of short-term cash flows prediction models in chapter 5, we are likewise interested in assessing our models prediction power on the cumulative yearly cash flows.

Appendix H displays the backtest results for historical reference exercises 2014 to 2018 and for main AGRe business currencies, namely EUR, USD, GBP and CHF. Those results take the form of clusters of tables, each cluster representing a particular assessment metrics (NMSE, NRMSE, NMAE and MAPE). The historical exercise of reference (and of projection) is displayed along with the currency (scope of all treaties labelled in that currency), as well as the corresponding numerical values of the given performance measure for the different model configurations. Backtest results for all currencies aggregated (i.e. all reinsurance treaties considered within the base scope) are also presented.

From Table 31 in appendix section H, looking at the metrics values weighted averages (weighted by the number of actual observation points, i.e. of observed yearly claims paid transaction amounts following the exercise of reference and related to prior treaties) from a global point of view (all currencies mixed up), we note that:

- there is no significant improvement passing from the old IBNR allocation to the new one: for a given projection method, old and new allocations yield the same results regardless of the metrics considered;
- projection method 1 gives better global results than the old projection with respect to the prediction of individual yearly cash flows as well as to the prediction of cumulative yearly cash flows: this is most obvious for quadratic distances, where the NMSE and NRMSE of projection method 1 are twice as low as the old projection ones;
- projection methods 2 and 3 perform well regarding cumulative cash flows, with all the performance criteria, but MAPE, taking the smallest values.

Looking at the weighted mean of each currency individually (Tables 32 to 35 in appendix section H), we observe that:

- The new IBNR allocation leads to better results only for the prediction of CHF and GBP individual and cumulative cash flows;
- Projection method 1 surpass the old projection method in the case of EUR and GBP cash flows, is comparable for USD cash flows and falls behind as for CHF cash flows;
- Projection methods 2 and 3 stand out as the best ones in the prediction of GBP future yearly cash flows, and in the prediction of cumulative claims paid amounts in GBP and EUR.

A few examples setting side by side reserves projections into yearly cash flows and actually observed claims payment yearly cash flows (with their cumulative counterparts) can be visualized on the graphs belows as well as in appendix H. They represent only a fraction of possible scopes and backtest projection exercises and the predictions are truncated at 2030.

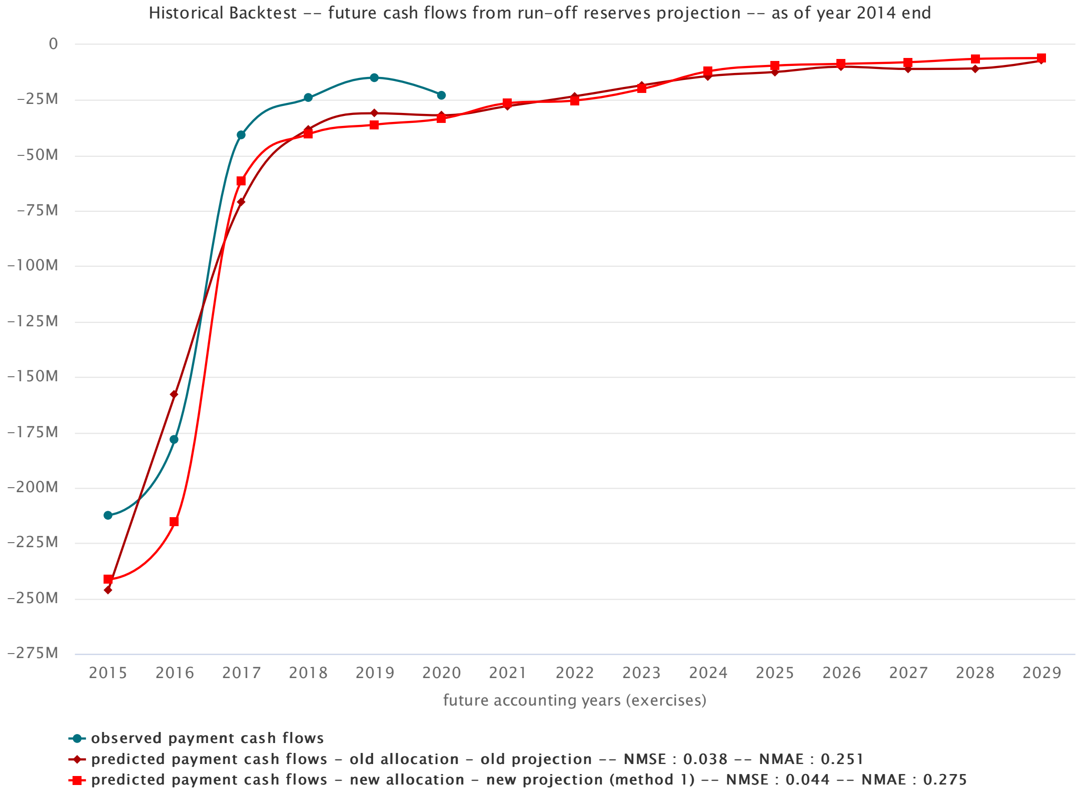


Figure 89: Reserves projections into yearly cash flows vs actually observed claims payment yearly cash flows – as of 2014 – all currencies – pooled business

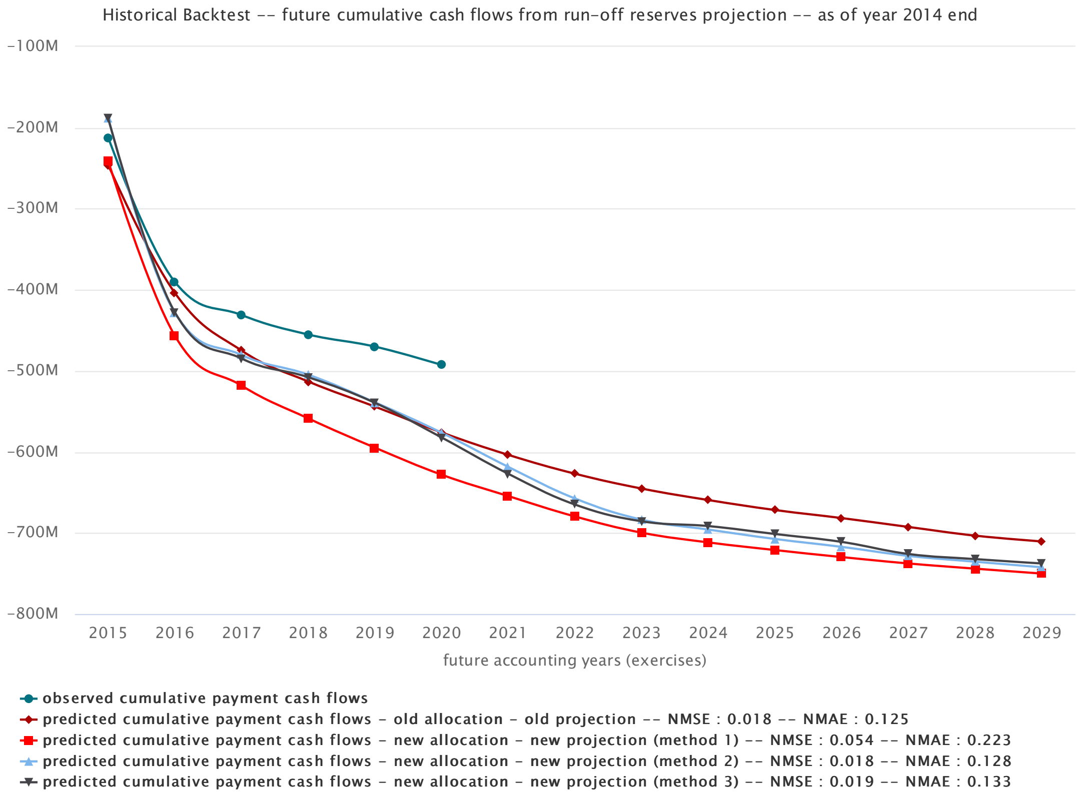


Figure 90: Reserves projections into cumulative yearly cash flows vs actually observed claims payment cumulative cash flows – as of 2014 – all currencies – pooled business

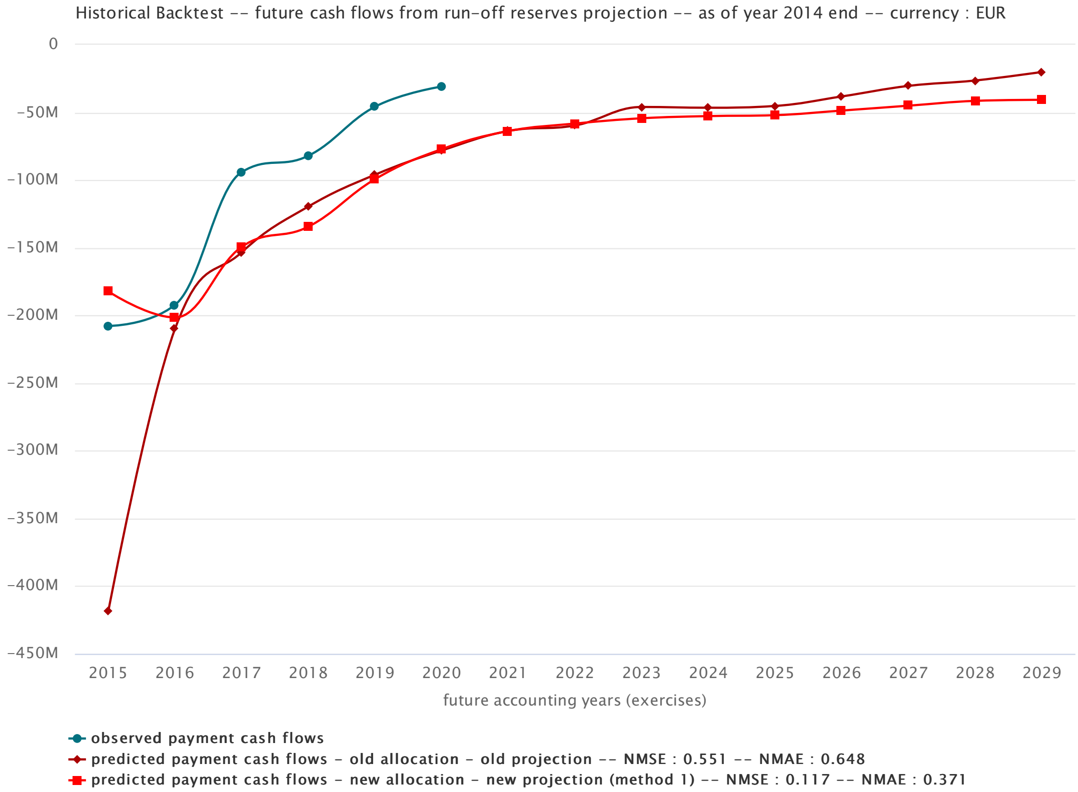


Figure 91: Reserves projections into yearly cash flows vs actually observed claims payment yearly cash flows – as of 2014 – currency EUR

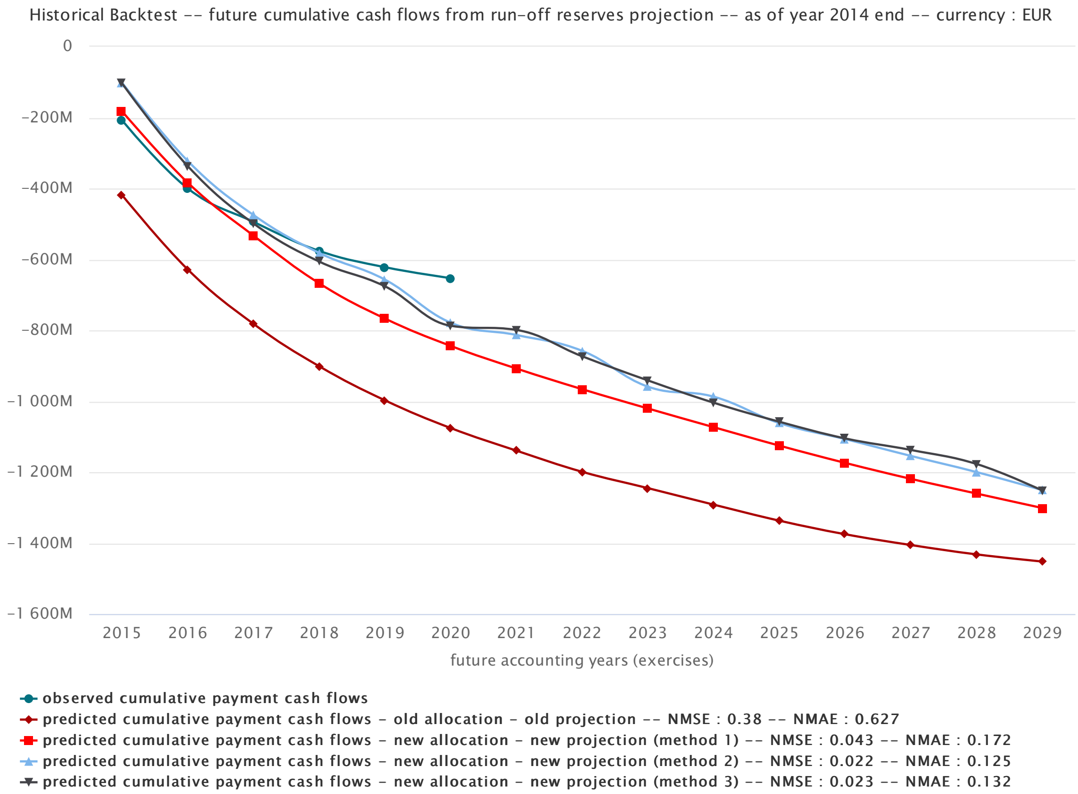


Figure 92: Reserves projections into cumulative yearly cash flows vs actually observed claims payment cumulative cash flows – as of 2014 – currency EUR

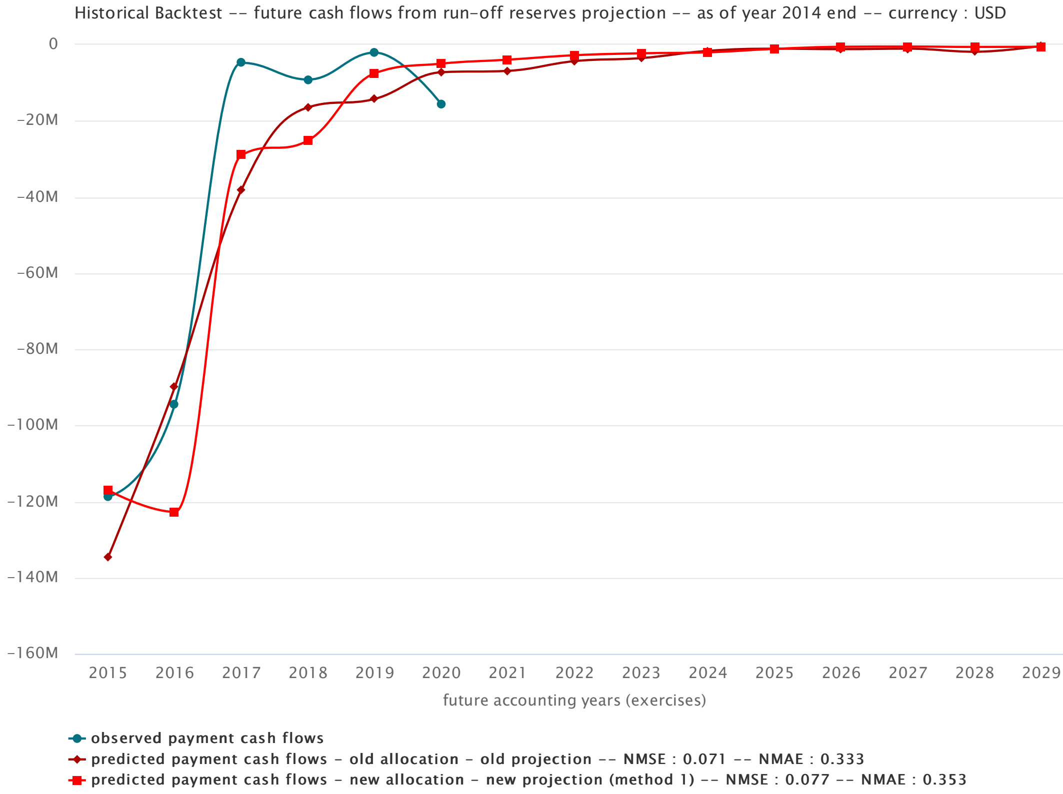


Figure 93: Reserves projections into yearly cash flows vs actually observed claims payment yearly cash flows – as of 2014 – currency USD

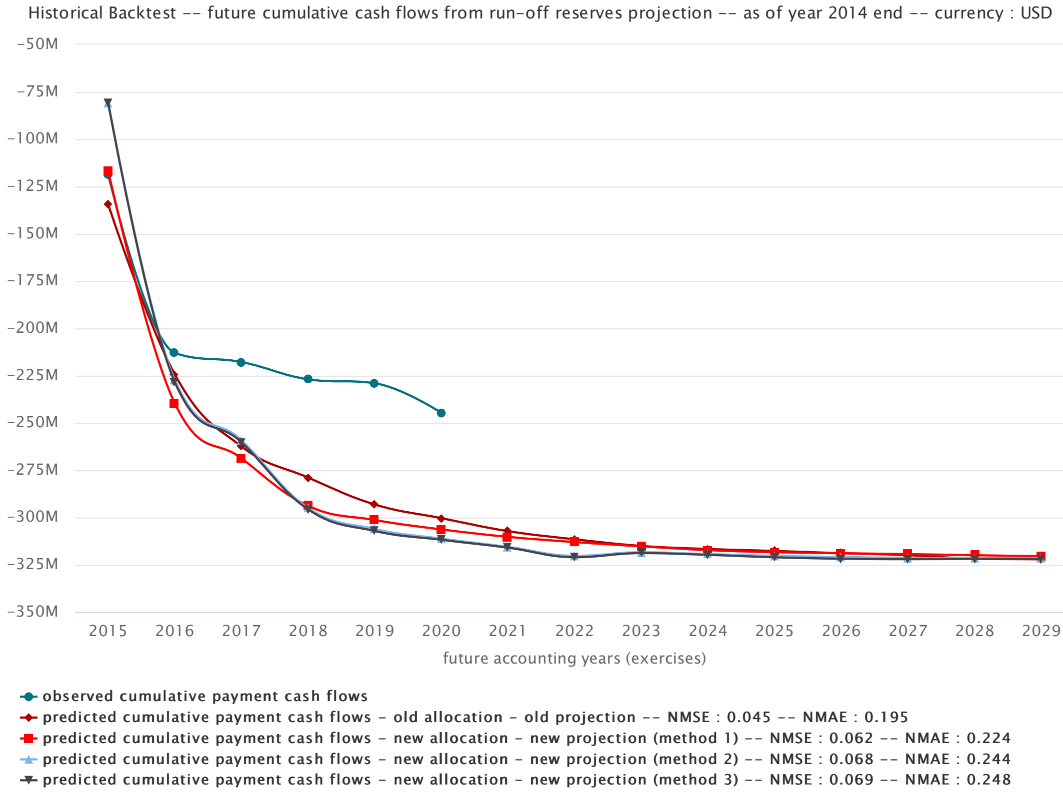


Figure 94: Reserves projections into cumulative yearly cash flows vs actually observed claims payment cumulative cash flows – as of 2014 – currency USD

We can conclude that the expected positive effect of the new IBNR allocation procedure was not perceived in the backtest implementation, while the new devised projection methods globally proved to generate individual (and cumulative) cash flows closer to the real ones than the old projection method. In addition, we noticed that over backtest projection years 2014 to 2018, we projected higher absolute amounts than what was actually observed. To express it differently, it seems that in taking reserves as a the projection base material, that is, as a starting point for cash flows prediction, one overestimates the future claims paid amounts.

Conclusion

To conclude this research project in actuarial science, that initially intended to lay a possible path to future cash flows prediction in currencies within a reinsurance business, we will address consecutively the challenges we met to specify and acquire some of the necessary underlying data, our work achievements and related derivatives together with its limits, and potential ways to go beyond what we have undertook and accomplished so far.

Let us begin with the data issues. Data linked to AGRe gross or net of retrocession quantities related to each reinsurance treaty's section such as reserves, premiums, claims paid as well as data linked to claims development triangles were already directly available when we began this undertaking. We had only to tap into those data, manipulate, aggregate and format them, to make them talk and provide us with insightful pieces of information. This was not that straightforward to retrieve historical AGRe balances comprising the right scope and on top of which we could reliably build models (data used in chapter 5). It was even less so to achieve and manage the quantity and quality of historical single financial and non-financial statements data desired (process still ongoing : see chapter 6). Similarly, a bit of effort was spent to acquire time pictures of the general accounting and contractual database over many closing periods going back to 2014 (data employed for backtesting purpose in chapter 85). Those databases did not exist, and had to be specified and built together with AGRe's data team. Multiple back and forth, and feedbacks were necessary to produce data meeting our needs in terms of quality, variables and perimeter. After each prototype extraction from the data team, the freshly assembled and delivered data were passed under close scrutiny through numerous carefully realized checks and corrective actions performed from our side. Those newly designed databases are liable to be of interest for other users (actuaries or not) from other departments of AGRe.

Regarding pure results achieved throughout this project, we supplied with systematic and methodical procedures to predict short-term monthly cash flows in currencies, as well as run-off BE liabilities projection into long-term yearly cash transactions. Although, the performance of proposed duration models to forecast monthly cash flows coming from all EBNYS balances as of date (whatever their amounts) is mitigated, those models have proven to be well adapted, yielding satisfactory results in the prediction of cash flows stemming from either positive or negative EBNYS balances. We have furthermore devised data structuring methodologies, and reviewed potentially relevant models (to be tested), with a view to anticipate future monthly cash flows in currencies arising from not yet constituted balances. What is more, thanks to the credibility theory applied to AGRe claims development, along with a Bornhuetter-Ferguson approach for IBNR allocation at lower levels of reinsurance treaties aggregation, and redesigned reserves projection methods, we managed to reach more trustworthy patterns (closer to real ones) of run-off liabilities liquidation in each currency.

Let us recall that what spurred this whole study was the distant and somewhat utopian goal of perfect economic congruence at every time scales and for each transactional currency. This red thread runs through this thesis as an underlying incentive driving all the procedures put in place, the data sourcing and structuring, as well as the models developments carried out in this work. Endeavors towards as accurate as possible cash flows predictions (whether monthly or yearly) are just means to acquire the ability to better hedge FX risk and create more interesting investment opportunities either on a short or long-term basis.

We identified some limitations and unfinished tasks to this research project and the manner we tackled the issue at hand:

- First of all, an entire part of the short-term monthly cash flows prediction is yet to be completed. We indeed laid out several courses of action to avail ourselves of a broader source of information than solely historical balances, in order to anticipate transactions of cash not necessarily proceeding from EBNYS balances. The strategies and models brought forward need nonetheless to be confronted with real business data, what we have not been able to do yet;

- A limit of submitted models we have observed is that they do not capture erratic and sizeable fluctuations in cash transactions patterns for minor (low represented) currencies, generally emanating from one or two single substantial amounts paid or received. The same problem arises for the technical reserves developments into yearly cash flows of those currencies with rough and difficulty predictable patterns;

- It is moreover worth highlighting that a 12-month horizon perfect prediction of cash flows in currencies is, in essence, impossible (at least when referring to insurance and reinsurance activities): how could one possibly foresee the exact occurrence time and financial cost of future claims and catastrophic events defining in part future cash flows? As a striking example, nobody at AGRe (and elsewhere) was expecting the floods in Germany and Belgium in the first half of July of 2021, the resulting claims and subsequently the cash transactions that this event caused. The same is obviously true for longer time scales and further forecast horizons;

- Another restriction is involved in the way we dissociated the short-term prediction problematic from the long-term view, and the corresponding two separate perimeters we adopted (forecast of all technical cash flows versus forecast of claims payments originating from run-off technical reserves at date). We did not try to conciliate the two perspectives, and as such there is a priori no continuities between short-term and long-term cash flows predictions. A reconciliation between the claims settlements amounts components of near monthly cash flows predictions and future yearly claims settlements (above all the first one) obtained from run-off liabilities projections, could be performed by means of constraints in the two sets of models or through a model mixing both time scales and both sources of information;

- At last, we did not quantified the impact of proposed methods and models on AGRe's capacity to protect itself against FX risk. Neither did we address the consequences of our predictions on AGRe's investment strategies and portfolio returns.

All of the above-mentioned limiting points constitute as many pending topics to inquire further into in a potential other actuary thesis project. To add up to this, one could go deeper into exploiting hierarchical credibility model for loss development (but need for more data/treaties or at least more balanced sub categories of reinsurance treaties). In an attempt to go even further, one could also think of, build, explore, and test machine learning algorithms integrating as input of the model all information available at date (provided that this information is accessible through reliable historical data) about every single liabilities elements past stock levels or variations and every single past financial transactions (together with their natures and characteristics), as well as claims first notice and evaluations (and possibly other sources of information), to predict future cash flows (monthly, quarterly, or yearly, at the desired level of aggregation) as output of the model (see as a proposal of such a model 6.2.4).

In the absence of astonishing results and breakthroughs, we provide nonetheless new insights, sound methodologies and mathematically backed models in our attempt to predict future cash flows in currencies. Besides, we believe that what we developed in this thesis can be applied (or at least be a source of inspiration) to other insurance or reinsurance international businesses, with higher volume of cash flows in currencies and potentially better predictive power of the employed models.

References

[1] *Segmentation et provisionnement en réassurance non vie*, Mémoire présenté devant l'Université Paris Dauphine pour l'obtention du diplôme du Master Actuariat et l'admission à l'Institut des Actuaires, Jérôme CRÉTIEN, AXA Global PC, 2013

[2] *DISTRIBUTION-FREE CALCULATION OF THE STANDARD ERROR OF CHAIN LADDER RESERVE ESTIMATES*, Thomas Mack, 1993

[3] *A Course in Credibility Theory and its Applications*, Hans Bühlmann and Alois Gisler

[4] *Credibility for the Chain Ladder Reserving Method*, Alois Gisler and Mario V. Wüthrich

[5] *AGRe - Reserving Process - 20Q4*, Internal documentation, Q4 2020

[6] *Market STEC Methodology - Technical Documentation v3.11*, Internal documentation, November 2019

[7] *ENSAE class Duration models*, Olivier Lopez, CREST ENSAE, 2020

[8] *Reinsurance and Foreign Exchange Risk*, Henri Loubergé, 1979

[9] *THE "MODIFIED BORNHUETTER-FERGUSON" APPROACH TO IBNR ALLOCATION*, TRENT R. VAUGHN, PHOEBE TINNEY

[10] *Multivariate Covariance Generalized Linear Models*, Wagner Hugo Bonat

[11] *Multiple Response Variables Regression Models in R: The mcglm Package*, Journal of Statistical Software, Wagner Hugo Bonat

List of Figures

List of Tables

Appendices

A Mack-Bootstrap of claims triangles

The Mack-Boostrap method is applied by AGRe at aggregated Solvency II Line of Business level for the calculation of risk margins (moderate and stress margin rates). AGRe's relevant lines of business for this method are as follow:

- Credit and Suretyship Insurance and Proportional Reinsurance
- Non-Proportional Casualty Reinsurance
- Non-Proportional Marine Aviation and Transport Reinsurance
- Non-Proportional Property Reinsurance

The calculation at this level of aggregation allows for more diversification within reserving lines that are composed of similar risks (property, casualty, marine...) but the non-diversification between lines of business remains.

Adaptation of Mack-Bootstrap theory to reserving issues offers a relevant alternative to the closed formulas. The bootstrap allows us to simulate future settlements and to construct an empirical distribution of the ultimate claims estimates. To achieve this, several steps are carried out:

1. Calculation of Bootstrap estimated parameters

The average chain-ladder development factors and their respective claims development variance are estimated as follows:

$$\hat{f}_j = \frac{\sum_{i=1}^{N+1-j} f_{i,j} C_{i,j}}{\sum_{i=1}^{N+1-j} C_{i,j}} \quad \text{and} \quad \hat{\sigma}_j = \frac{1}{N} \sum_{i=1}^{N+1-j} C_{i,j} \cdot \left(f_{i,j} - \hat{f}_j \right)^2 \quad \forall j \in [\![1, N]\!]$$

The residuals are then defined as follows:

$$\hat{r}_{i,j} = \sqrt{\frac{N}{N-1}} \cdot \sqrt{C_{i,j}} \cdot \frac{\hat{f}_{i,j} - \hat{f}_j}{\hat{\sigma}_j}$$

2. Resampling (bootstrap)

The bootstrap consists in resampling the triangle and computing for each resampled triangle the Best Estimate reserves via the underlying reserving method (here Chain-Ladder). For each resampled triangle:

a. Calculate the individual development factors

$$f_{i,j}^s = \hat{f}_j + r^s \cdot \frac{\hat{\sigma}_j}{\sqrt{C_{i,j}}}$$

b. Calculate the development factor

$$f_j^s = \frac{\sum_{i=1}^{N+1-j} f_{i,j}^s \cdot C_{i,j}}{\sum_{i=1}^{N+1-j} C_{i,j}}$$

c. Calculate the lower part of the triangle: a process error is added using normal distribution with estimated mean and standard error.

The Mack-Bootstrap process is summarized in the following figure:

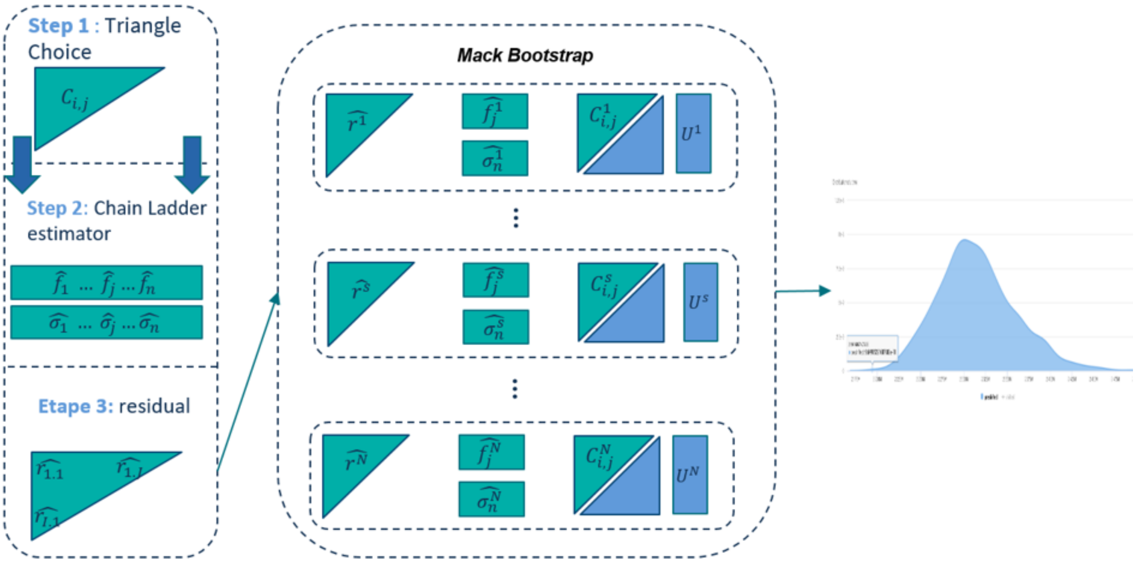


Figure 95: Summary of the Mack-Bootstrap method

B Curve-fitting methods

1. Curve-fitting methods

We may wish to impose a particular form on the development pattern or factors, by performing a regression of the observed coefficients against a predetermined model. The assumption is that there is a relationship between the development factors f_j and the development periods j. There are 4 widely-used fitting functions, involving two real parameters a and b:

- Inverse Power: $f_j = 1 + \frac{a}{j^b}$
- Exponential: $f_j = 1 + a \cdot e^{-b \cdot j}$
- Power: $f_j = a^{b^j}$
- Weibull: $f_j = \frac{1}{1 - e^{-a \cdot b^j}}$

For each fitting function, parameters a, b can be estimated using least square regression. Note that one has to exclude first the $f_j < 1$, if there exist some. For example, for the exponential fitting:

$$\log(f_j - 1) = \log(a) - b \cdot j$$

thus

$$\widehat{b} = -\frac{(N-1) \cdot \sum_{j=1}^{N-1} \mathrm{j} \cdot \log(f_j - 1) - \Sigma_j^{N-1} j \sum_{j=1}^{N-1} \log(f_j - 1)}{(N-1) \cdot \sum_{j=1}^{N-1} \mathrm{j}^2 - \left(\sum_{j=1}^{N-1} \mathrm{j}\right)^2}$$

and

$$\log(\hat{a}) = \frac{\sum_{j=1}^{N-1} \log(f_j - 1)}{N-1} + \hat{b} \cdot \frac{\sum_{j=1}^{N-1} j}{N-1}$$

Once the parameters are calibrated, the tail factor is applied by multiplying the fitted development factors for $j \geq N$ up to $j = J$, where J is the development period after which it is considered that no more claims development occurs (e.g. $J = 30$).

$$F_N = \prod_{j=N}^{J} f_j^{fitted}$$

Those curve-fitting methods can also be seen as a smoothing process on the whole development curve, useful in the case where the observed development factors behave too oddly with respect to what the chain-ladder model expects, namely coefficients greater than 1, decreasing and tending to 1 (or equivalently: development pattern increasing first rapidly and then slowly to 1).

2. Bondy Methods

The principle of Bondy is to repeat the last observed link ratio: $F_N = f_{N-1}$. The age-to-age factors in the tail are obtained with the assumption that $\forall j \geq N, \quad f_j = \sqrt{f_{j-1}}$:

$$F_N = \prod_{j=N}^{\infty} f_j = \prod_{j=N}^{\infty} \sqrt{f_{j-1}} = \prod_{k=1}^{\infty} {f_{N-1}}^{1/2^k} = f_{N-1}^{\sum_{k=1}^{\infty} 1/2^k} = f_{N-1}$$

It is to be noted that f_{N-1} has to be greater than 1 in order to have the age-to-age factor greater than 1. Thus, in practice, it can be required to smooth the development factors before applying the Bondy method. The generalization of Bondy method leads to:

$$F_N = \prod_{k=1}^{\infty} {f_{N-1}}^{B^k} = f_{N-1}^{\frac{B}{1-B}} \quad \text{where } B < 1 \text{ is to estimate.}$$

Let $I_j = \log(f_j)$ where f_j is an estimate of the age-to-age factor for $j \leq N-1$ (for example obtained by the weighted mean of link-ratios). We have: $I_{N+k} = I_N \cdot B^k$ for $k \geq 0$. Thus, one estimates $\hat{B}$ to fit on latest development ratios, say $j \geq i$, i.e. one tries to minimize:

$$\sum_{j=i}^{N} \left(I_j - \hat{I}_i \cdot \hat{B}^{j-i} \right)^2$$

over the parameters $\hat{I}_i = \log\left(\hat{f}_i\right)$ and $\hat{B} < 1$

3. Benchmark methods

The principle of Benchmark methods consists in deriving a tail factor using claims triangle of similar risks for which sufficient claims developments are available.
If one denotes $f_j^b = 1 + v_j^b$, for $j \in [\![1, N_b]\!]$, the age-to-age factors of the claims triangles used as benchmark, N_b being the number of available developments of the Benchmark triangle, the tail factor to apply is:

$$F_N = \prod_{j=N}^{N^b} f_j^b = 1 + V_N$$

We can adjust the estimated tail factor to take into account the differences in the known developments between the triangle under study and the Benchmark triangle. For this, one computes the mean R of the ratios $\frac{v_j}{v_j b}$ over the periods $j = i \ldots N$, and uses it to adjust the estimation:

$${F_N}^{\text{adjusted}} = 1 + V_N \cdot R$$

C Other reserves estimation

C.1 Unearned Premium Reserves (UPR)

The Unearned Premium Reserves are estimated for every treaty underwritten in the current year, proportionally to premiums amounts:

$$UPR = \text{ PREMIUM } \times \max\left(0, 1\text{-rate}_{\text{ EARNED PREMIUM }}\right)$$

with:

$$\text{rate}_{\text{ EARNED PREMIUM }} = \frac{\text{current date } - \text{ treaty section start date}}{\text{treaty section end date } - \text{ treaty section start date}}$$

C.2 Unexpired Risk Reserves (URR)

The Unexpired Risk Reserves (URR) is set to take into account a known insufficiency of the UPR to cover future claims. A non-null URR is usually the consequence of an inadequate pricing, as the estimated claim cost relative to the earned premiums is higher than the earned premiums. The URR is calculated as follows:

$$URR = \min\left((1 - LR_{BE}) \cdot UP + UC + ULAE_{UPR}, 0\right)$$

where LR_{BE} is the ultimate Best Estimate Loss Ratio net of reinsurance, UP is the unearned premiums (> 0) net of reinsurance, UC is the unearned commissions (generally a cost) and $ULAE_{UPR}$ is the related unallocated adjustment expenses.
Remarks:

- The lines of business on which a URR may exist are the pools and retentions (Property, Marine, Motor, Liability) where the net written premium is non-null.
- For AGRe, the URR is potentially visible during intermediate closings. Indeed, at full-year closing, the volume of unearned premiums net of reinsurance is not material. This is due to the fact that almost all the treaties cover the period 01/01 to 31/12 and thus almost all the written premiums are earned at 31/12.
- The unallocated adjustment expenses are often considered as of second order and can be thus negligible compared to commissions or compensated by "pool management commissions".
- According to the Group standard, the URR is not required if it can be shown that the current Global Safety Margin (GSM) can cover both the Risk Margin Moderate and the URR.

C.3 Unallocated Loss Adjustment Expenses Reserves (ULAE)

Unallocated loss adjustment expenses (ULAE) are costs incurred by an insurance company that cannot be attributed to the processing of a specific claim. They are among the expenses for which an insurer has to set aside reserve funds, in addition to allocated loss adjustment expenses and contingent commissions.

Hereafter are described two methodologies tested by AGRe to estimate the ULAE reserves: New York Method vs Duration Method. The ULAE computation relies on a strong input from the controlling department: the annual cost for handling claims, also called loss adjustment expenses (LAE).

Loss Adjustment Expenses (LAE)

The Loss Adjustment Expenses (LAE), used for the estimation of ULAE reserves, is the amount of annual general expenses corresponding to the loss management activity for each department. It's calculated by the financial controlling department, applying the formula:

$$LAE = \begin{array}{c} \text{GRID GE - GRID Average \% for Loss Adjustment} \\ + \text{ LEGAL GE } \cdot \text{ LEGAL Average \% for Loss Adjustment} \end{array}$$

where:

- General Expenses (GE) per department = statutory general expenses amount allocated to LEGAL and GRID departments. It takes into account salaries (for the GRID according to the salaries of the main GRID teams (Claims, Mature Markets, High Growth Markets, Business Support, etc.), and expenses related to rents, IT, salaries, travel, fees, etc. Rent and IT cost are allocated on number of employee's basis. It is assumed that only GRID (underwriting and claims department) and Legal department performs claims management activities.
- Average % for Loss Adjustment = the weighted average % of time spent on loss adjustment activity for each department estimated as a percentage of the time spent per activity for GRID and LEGAL department (on a declaration basis). The weighted average percentage is on a salary basis for each GRID team.

LAE are split on LAE Current and LAE run-off, because some of the LAE is used for opening of the claims files and some for settling open claims. This split is made using a key on time spent by each GRID team on "current" or "run-off" claims management activity.

Methodology based on duration

This methodology was developed in 2015 by AGRE and mostly consists in calculating the average duration for handling claims and applying to it the annual cost.

$$\text{ULAE Reserves } = \text{ Duration } \cdot LAE$$

The annual cost for handling claims LAE is a direct input from the controlling department. An average based on a few years can be preferred so that the value is not too much impacted by exceptional costs. "Duration" being the run-off duration for managing claims, calculated based on the development curves of reported and closed claims, derived from corresponding claims number triangles.
1. Development pattern of reported, closed claims and claims to be managed:

One defines for $j \geq 1$:

$$r_j = \frac{\text{number of reported claims after j years}}{\text{ultimate number of claims}}$$

r_j is obtained by calculating the development curve from the cumulative claims number triangle of reported claims. It corresponds to the % of reported claims after j years of development since the beginning of the UWY.

$$c_j = \frac{\text{number of closed claims after } j \text{ years}}{\text{ultimate number of claims}}$$

c_j is obtained by calculating the development curve from the cumulative claims number triangle of closed claims. It corresponds to the % of closed claims after j years of development since the beginning of the UWY.
The % of claims still to be managed after j years of development is then defined by the difference:

$$m_j = r_j - c_j = \frac{\text{number of claims to be maintained in development year } j}{\text{ultimate number of claims}}$$

The number of claims to be manage first increases with the maturity as the reported claims accumulates and few of them can be closed, until it reaches a maximum. Then it decreases since almost all claims are reported and more and more claims are closed.

2. Run-off duration of the portfolio:

One first computes the expected duration d_j of an open claim with maturity j years (risk underwritten j years ago).

$$d_j = \sum_{i=j}^{N} p_i * (i - j) \text{ where } p_i = \frac{m_i}{\sum_{k=i}^{N} m_k}$$

- m_i represents the part of ultimate number of claims still to be managed in development year i (posterior to development year j).
- p_i is thus the probability (normalization to 1) that the claim is still to be managed in development year i (posterior to development year i)
- $(i - j)$ represents the number of extra years needed to close the claims.

It is assumed that all claims are closed after $N = 30$ years of maturity.

One then computes the run-off duration of the portfolio, being the average of the expected duration weighted by the number $\mathrm{n_j}$ of open claims with maturity j years:

$$\text{Duration } = \frac{\sum_{j=1}^{N} d_j * n_j}{\sum_{j=1}^{N} n_j}$$

New York method

The New York method is simpler and less parametric than the first method. It relies on the amount of reserves and considers that the same cost will be applied in the future as the current cost for managing the claims. It is assumed that twice as less cost will be implied for liquidating the case reserves than for liquidating IBNR reserves since the claims have already been partially managed (assessed, reported, and documented).
The ULAE rate can be estimated over the current year or in a more robust way with an average on the past years:

$$\text{rate }_{ULAE} = \frac{\sum_{i=1}^{N} \text{ annual cost for handling claims in year } i}{\sum_{i=1}^{N} \text{ claims paid during the year } i}$$

ULAE rate is allocated by main line of business P&C and Life and by Pooled/Non Pooled criteria using number of treaties with claims as allocation key.

C.4 Counterparty Risk Provision (CRP)

The counterparty risk provision is calculated with the view to cover the risk of default of a counterparty. Exposure corresponds to reserves likely to be paid by our reinsurers and holds for the time needed to pay the claims. Probability of default are given based on reinsurers' ratings.

Exposure

The exposures on reinsurers are the Best Estimate of ceded reserves (unearned premiums and claims), i.e. the expected recoveries or reinsurance recoverables, minus collaterals.

$$\text{Exposure}_{\ r,l,y} = \text{ BE Claims Reserves}_{\ r,l,y} + \text{ BE Premium Reserves}_{\ r,l,y} - \text{ Collateral}_{\ r,l,y}$$

where r is a reinsurer, y is the projected year, and l is the LOB.

Ratings

A table with updated ratings by reinsurer from AM Best and S&P is provided by the "Comité de Sécurité". The S&P ratings are chosen: in case it is missing or is not available, the AM Best rating is considered and converted to S&P ratings format. If both S&P and AM Best ratings are missing, then the rating "BBB" is attributed.

Probabilities of Default

Probabilities of defaults (PDs) are computed from S&P probabilities of default of a company belonging to a given rating class over different time horizons (1 to 15 years).

From these data, we compute the probability that the reinsurer defaults in each time interval:

$$PD_{\leq i} = \text{Probability of Default before year i}$$
$$PD_{i+1} = \text{ Probability of Default between year i and year i+1}$$
$$PD_{i+1} = PD_{\leq i+1|\bar{\leq} i} = \frac{PD_{\leq i+1} - PD_{\leq i}}{1 - PD_{\leq i}}$$

After year 15, the probabilities of default are considered as flat: $PD_j = PD_{15}$ for $j > 15$.

Payment patterns

Payment patterns are derived, through a classic Chain-Ladder approach, from corresponding LOB cumulative paid claims development triangles. Payment patterns are imposed to be non-decreasing with the condition $f_j \geq 1$ for all j.

Recovery Rate

The recovery rate chosen is constant and the same for all reinsurers: it is equal to 34%, in compliance with the default recovery rate of the AXA Group's internal model.

Computation of expected loss

1[st] *step*: the projected exposure is computed for each reinsurer r, LOB l and projected year y

$$E_{r,l,y} = \frac{E_{r,l,y_0} \cdot (1 - PP_{y-y_0-i_0})}{(1 - PP_{i_0})}$$

Where PP_{i_0} is the percentage of claims already paid, depending on the maturity i_0 of the underwriting year y_0 to which the exposure is attached ($i_0 = 2021 - y_0 + 1$ if the closing year is 2021).

2[nd] *step*: each projected exposure is multiplied by the corresponding probability of default $PD_{r,i}$ depending on the reinsurer's rating and the projected year. The recovery rate is then applied to get each expected loss:

$$EL_{r,l,y} = E_{r,l,y} \times (1 - RR) \times PD_{r,y}$$

3[rd] *step*: the counterparty risk provision is finally obtained by summing over all reinsurers (N_R), all LOB's (N_L) and all projected years (N_Y) to get:

$$CPR_{\text{provision}} = \sum_{y=1}^{N_Y} \sum_{r=1}^{N_R} \sum_{l=1}^{N_L} PD_{r,y} \cdot E_{r,l,y} \cdot (1 - RR)$$

C.5 Global Safety Margin (GSM)

The Global Safety Margin managed by AGRe corresponds to margin over Best Estimate Scenario. Part of this GSM does not belong to AGRe but to other entities participating in the pools. It is estimated according to different risk scenarios and the AGRe risk appetite. AGRe respects Group Standard's recommendation for a GSM superior to Moderate Scenario and the GSM AGRe is placed between the moderate scenario and the stress scenario.

D Short-term monthly cash flows predictions - 12-month curves

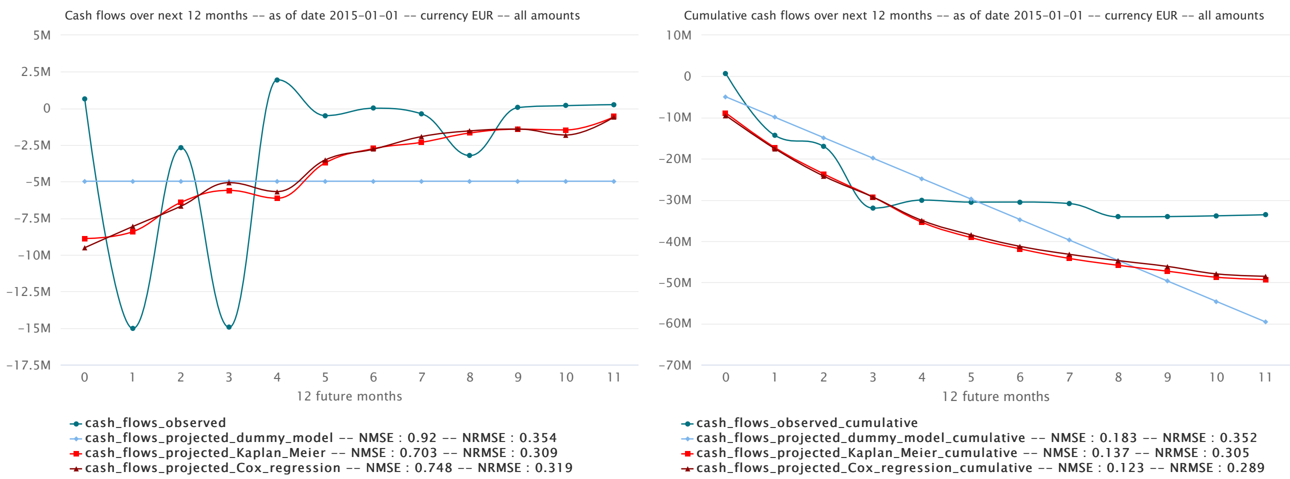


Figure 96: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2015-01-01 – currency EUR – all amounts

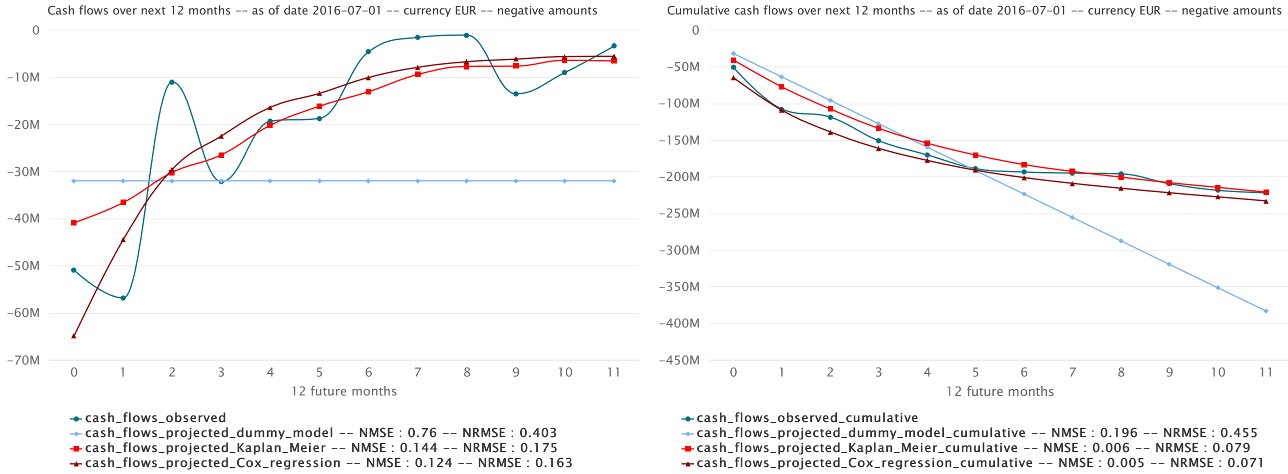


Figure 97: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2016-07-01 – currency EUR – negative amounts

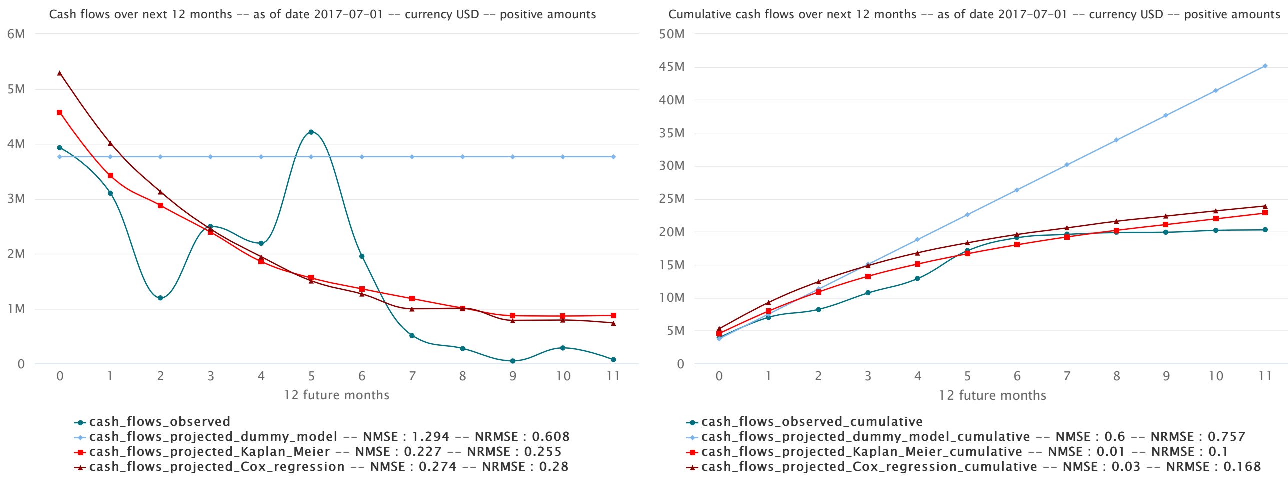


Figure 98: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2017-07-01 – currency USD – positive amounts

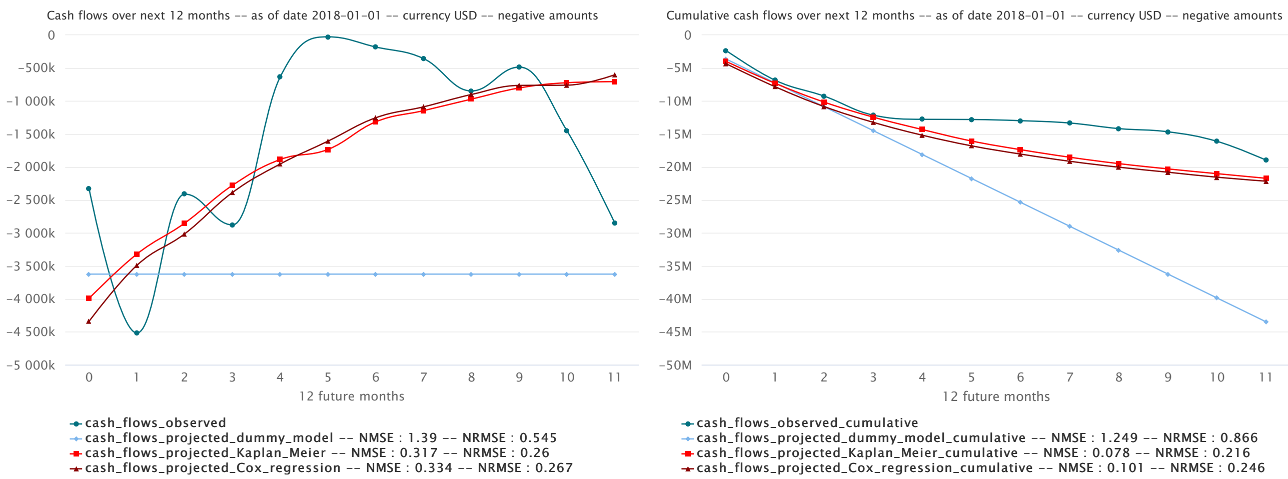


Figure 99: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2018-01-01 – currency USD – negative amounts

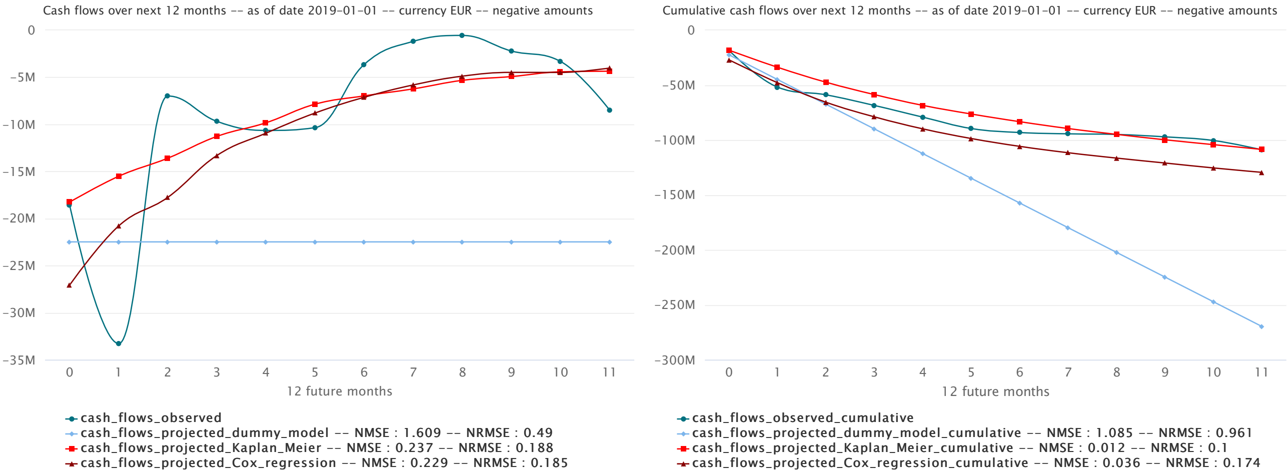


Figure 100: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2019-01-01 – currency EUR – negative amounts

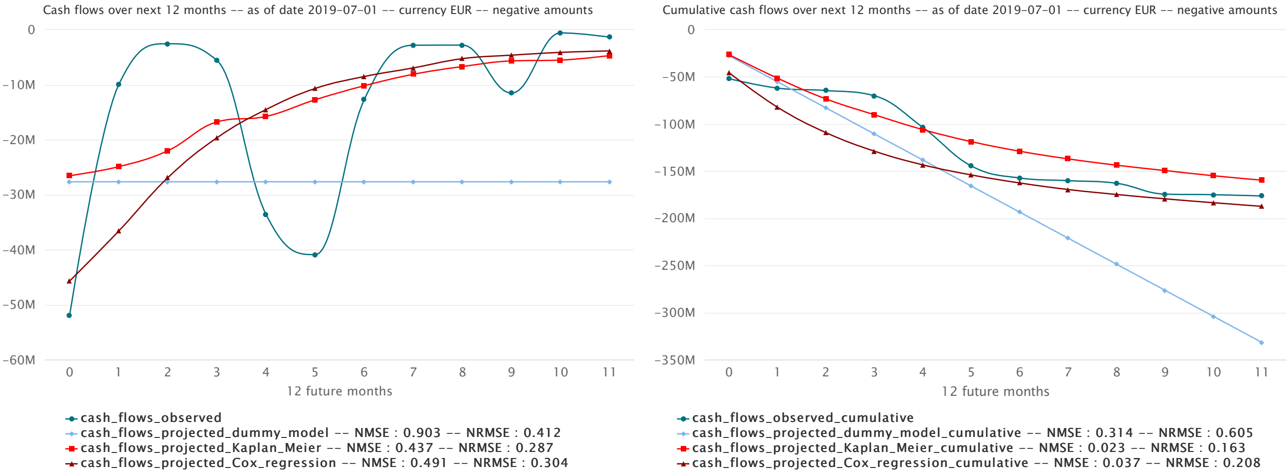


Figure 101: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2019-07-01 – currency EUR – negative amounts

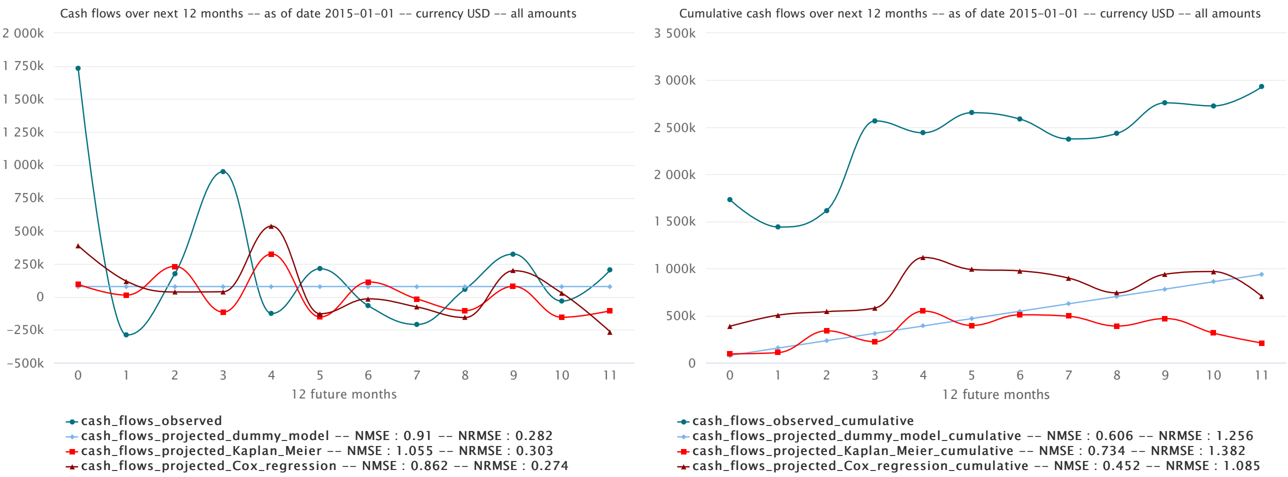


Figure 102: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2015-01-01 – currency USD – all amounts

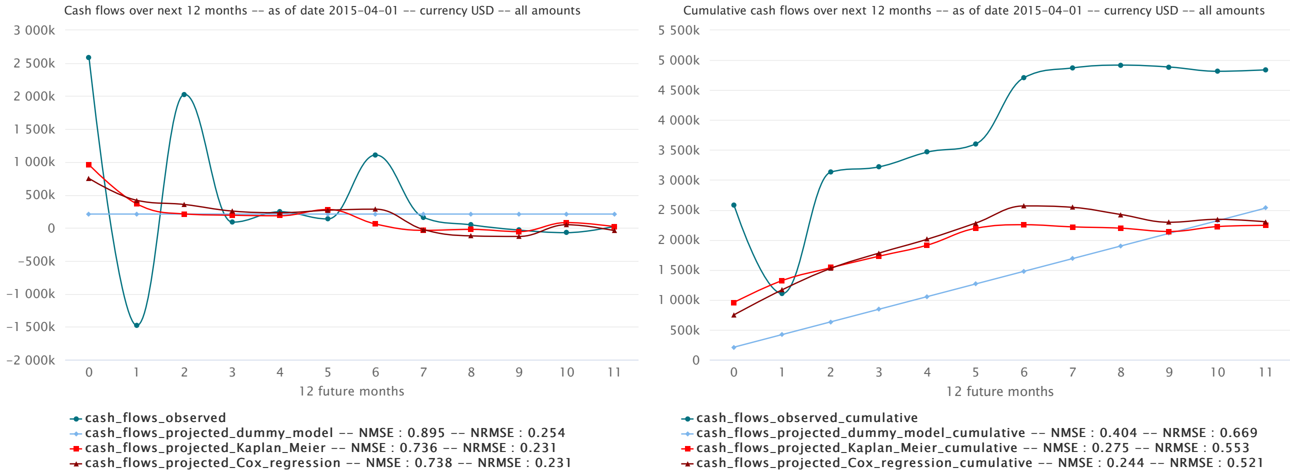


Figure 103: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2015-04-01 – currency USD – all amounts

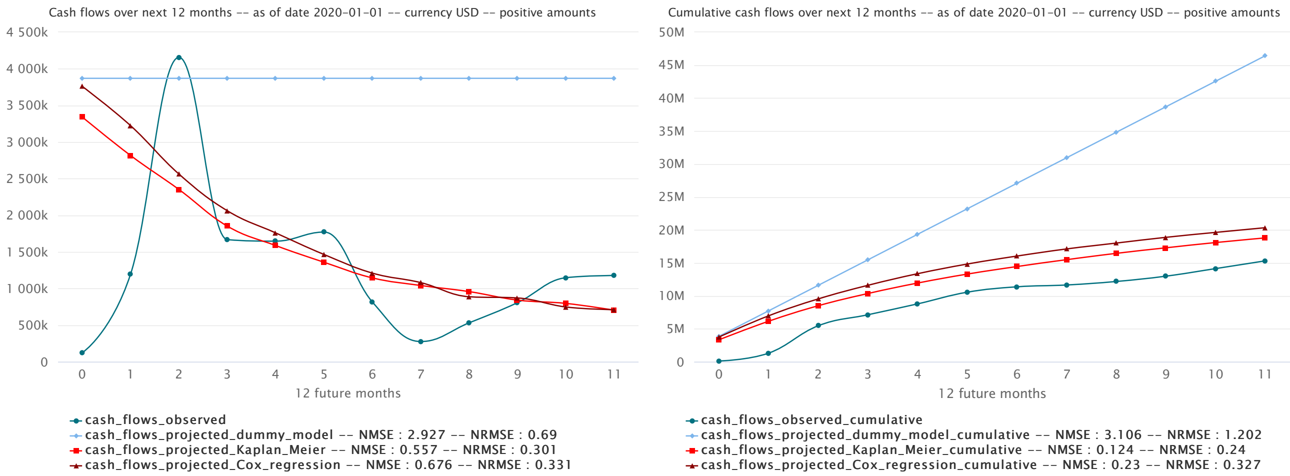


Figure 104: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2020-01-01 – currency USD – positive amounts

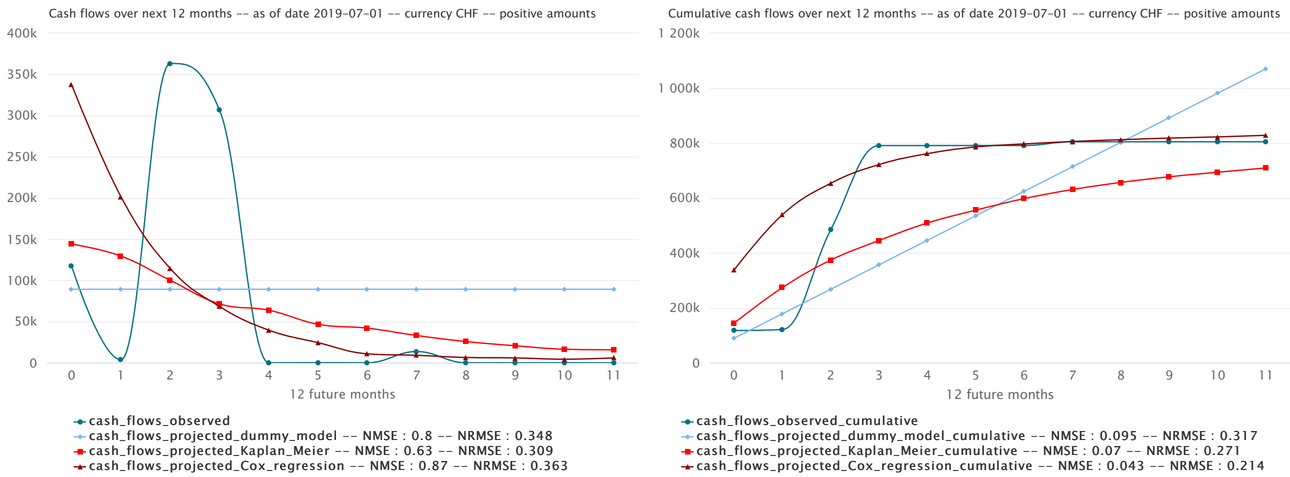


Figure 105: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2019-07-01 – currency CHF – positive amounts

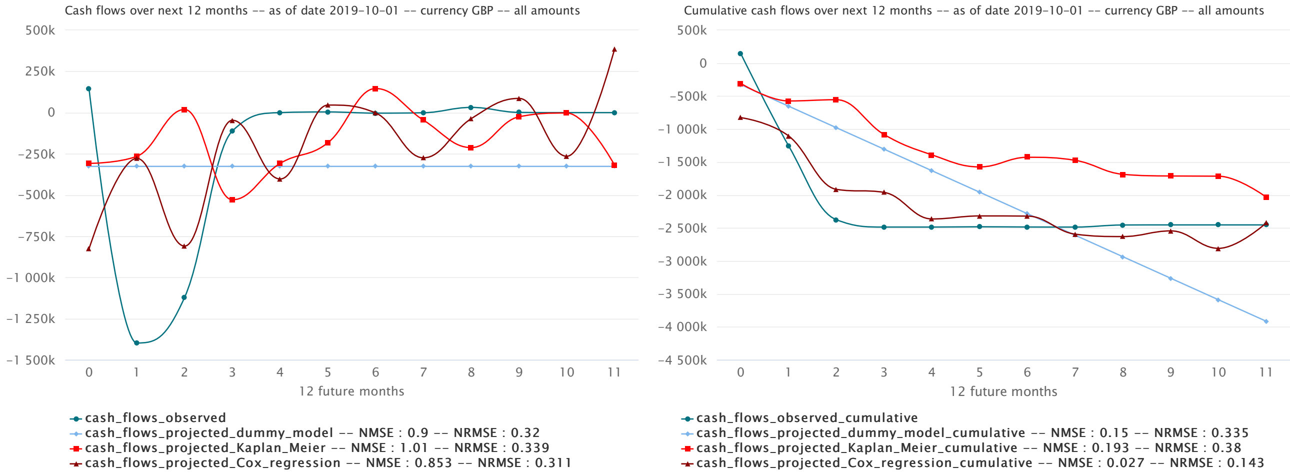


Figure 106: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2019-10-01 – currency GBP – all amounts

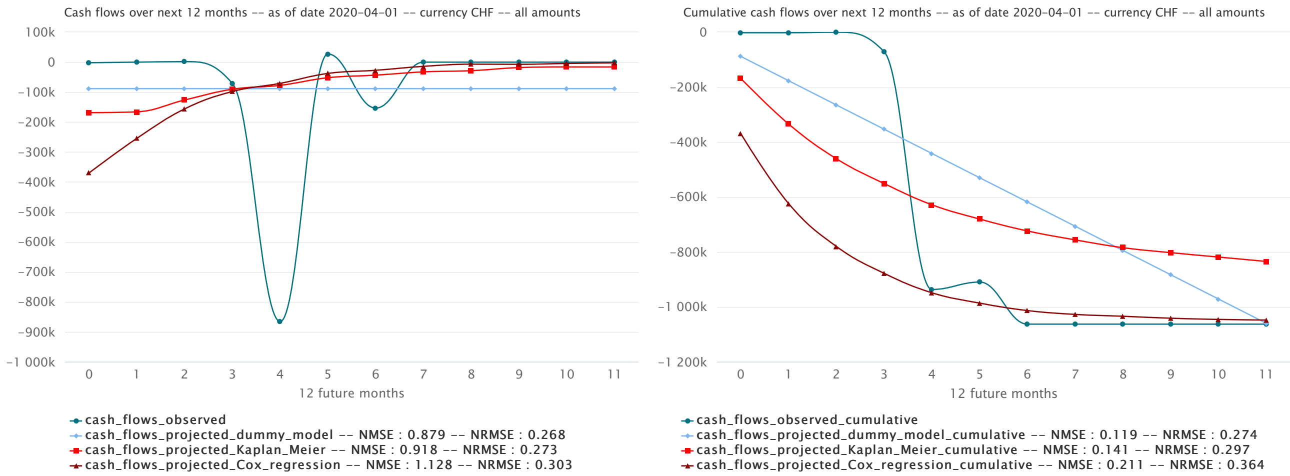


Figure 107: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2020-04-01 – currency CHF – all amounts

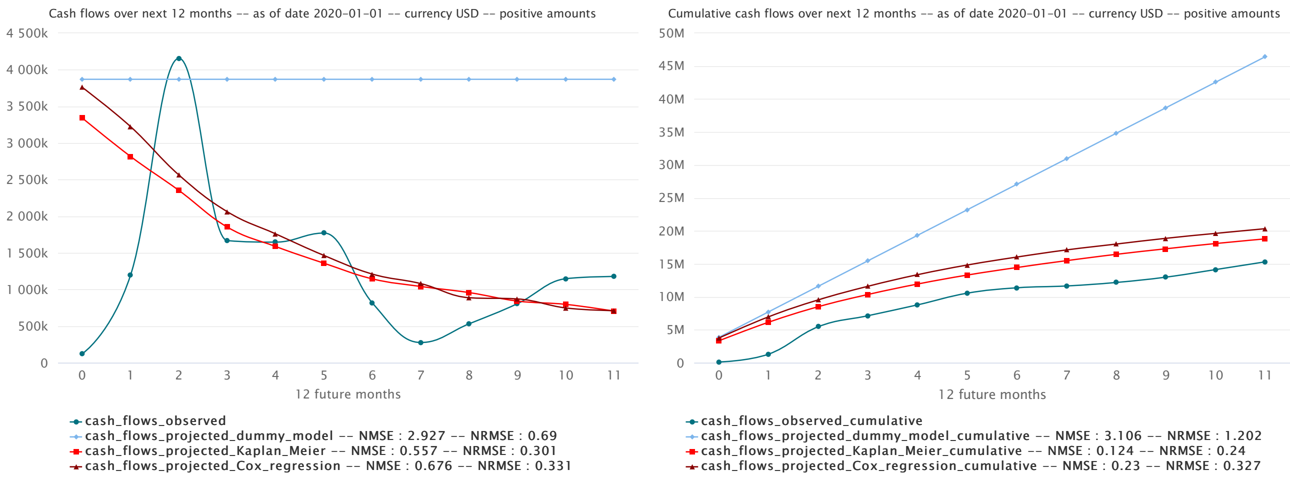


Figure 108: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2020-01-01 – currency USD – positive amounts

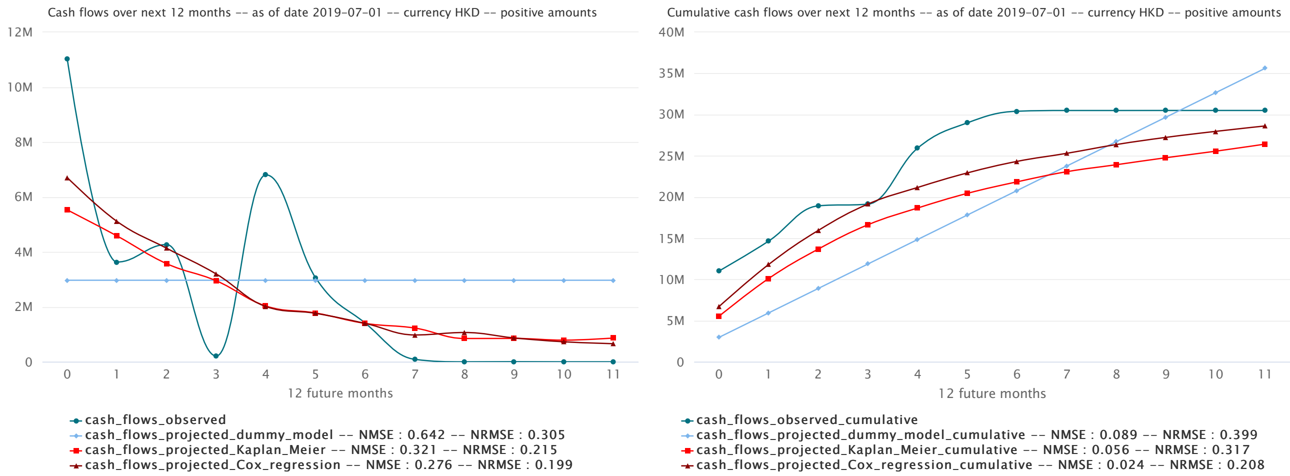


Figure 109: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2019-07-01 – currency HKD – positive amounts

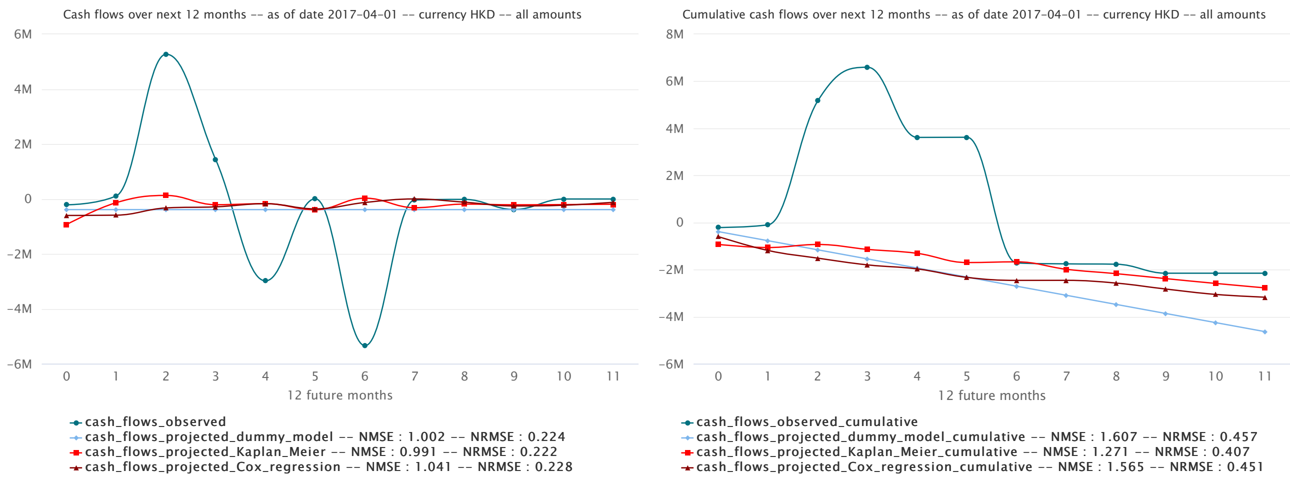


Figure 110: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2017-04-01 – currency HKD – all amounts

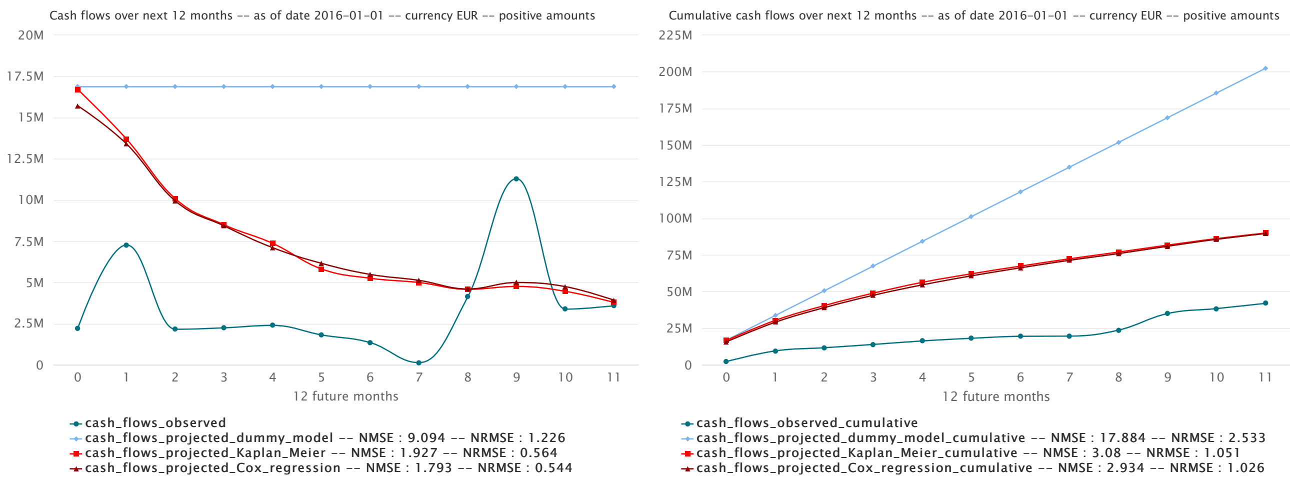


Figure 111: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2016-01-01 – currency EUR – positive amounts

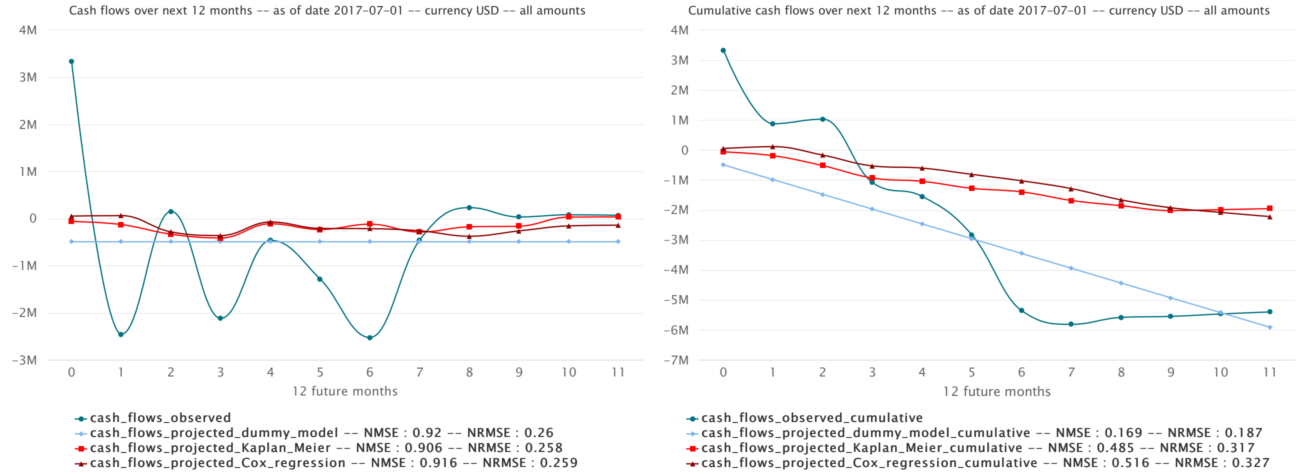


Figure 112: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2017-07-01 – currency USD – all amounts

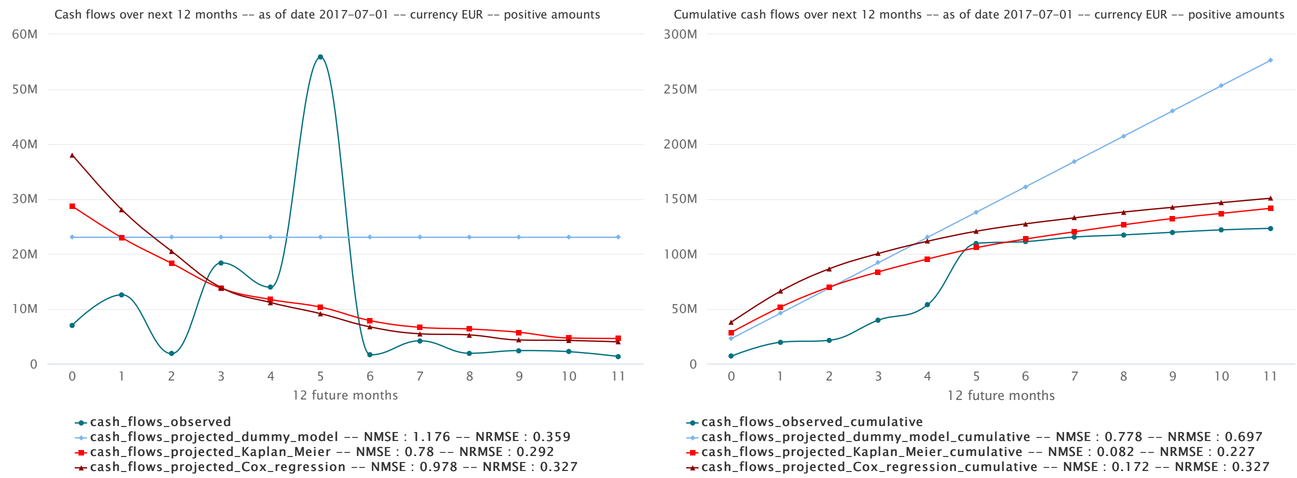


Figure 113: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2017-07-01 – currency EUR – positive amounts

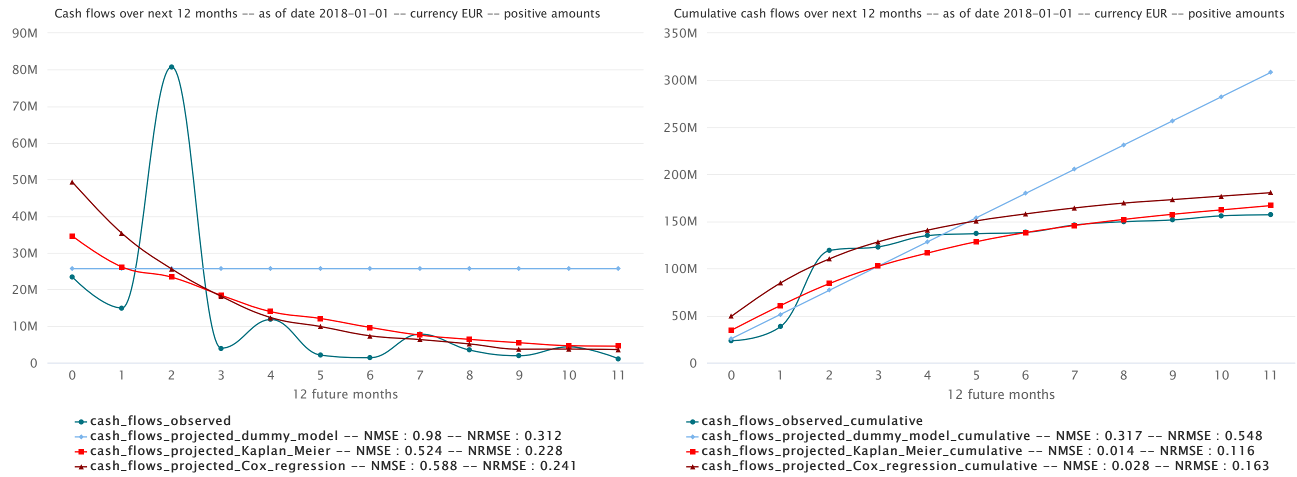


Figure 114: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2018-01-01 – currency EUR – positive amounts

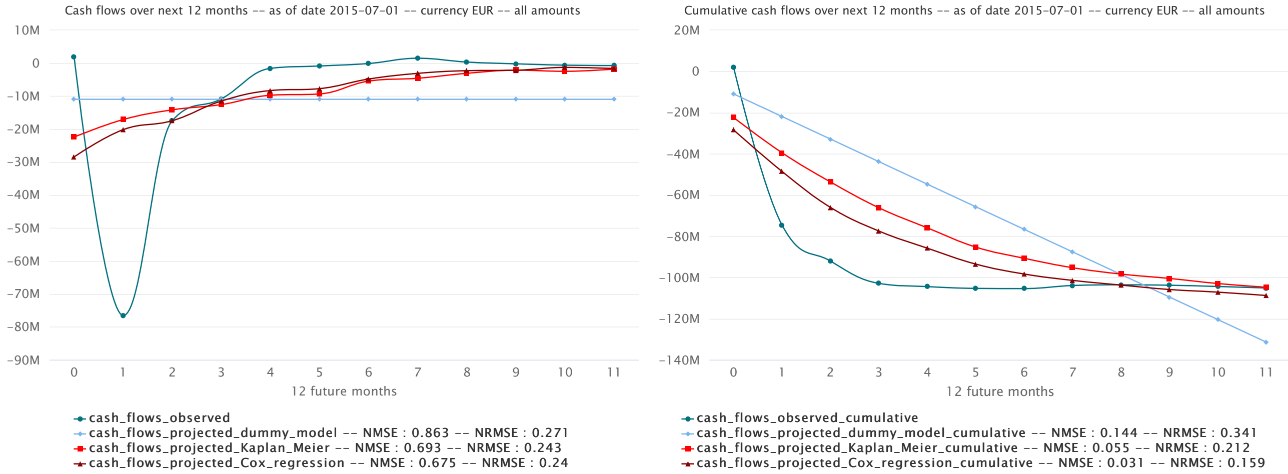


Figure 115: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2015-07-01 – currency EUR – all amounts

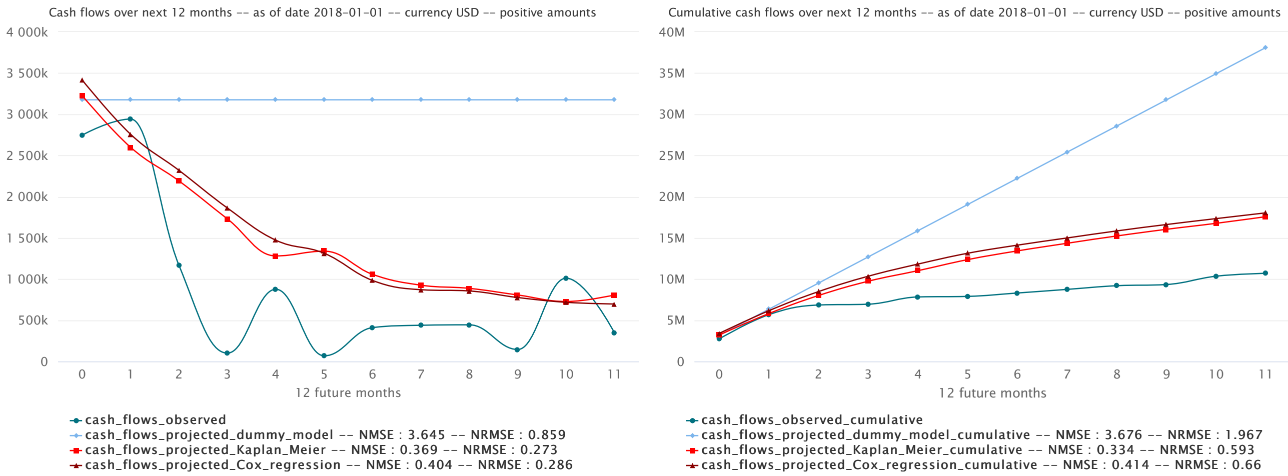
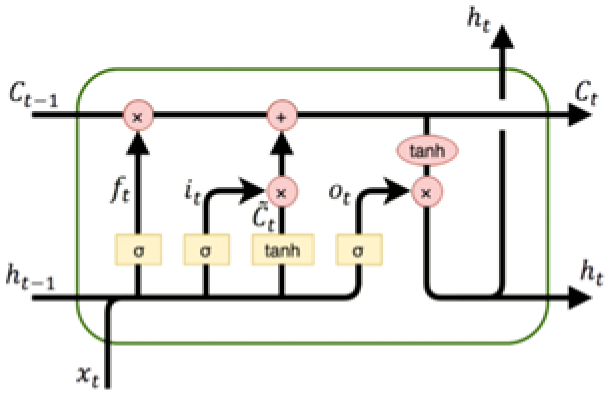


Figure 116: Individual (left) and cumulative (right) cash flows over next 12 months – as of date 2018-01-01 – currency USD – positive amounts

E LSTM step-by-step walk through

LSTM networks are neural networks composed of an input layer, one or more hidden layers and an output layer. The specificity of LSTM comes from the hidden layers since they are composed of memory cells, each one of them storing a part of previous data history and adjusting it as data is processed.

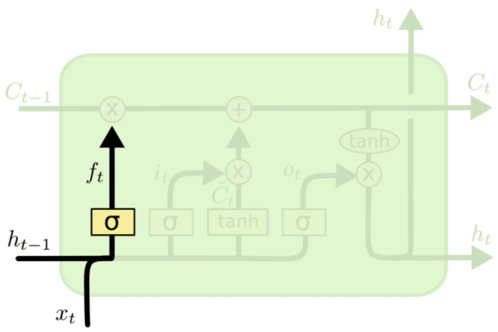


Variables:

- x_t: input vector to the LSTM unit;
- f_t: forget gate's activation vector;
- i_t: input/update gate's activation vector;
- o_t: output gate's activation vector;
- h_t: hidden state vector also known as output vector of the LSTM unit;
- $\tilde{C}_t$: cell input activation vector;
- C_t: cell state vector;
- weight matrices and bias vectors parameters which need to be learned during training.

The first step decides what information is thrown away from the cell state. This decision is made by a sigmoid layer, called the "forget gate layer", that looks at h_{t-1} and x_t, and outputs a vector f_t of numbers between 0 and 1 corresponding to each number in the cell state C_{t-1}.

- 1: all information detained by C_{t-1} is preserved.
- 0: all information detained by C_{t-1} discarded / forgotten.



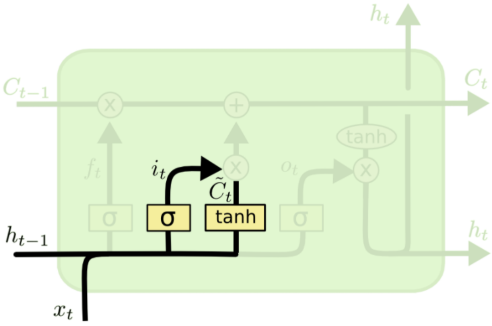
$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \ + \ b_f\right)$$

The next step amounts to determining, what new information is going to be stored in the cell state.

- On one hand, a sigmoid layer called the "input gate layer" decides which values will be updated, and so produce the input gate's activation vector i_t.

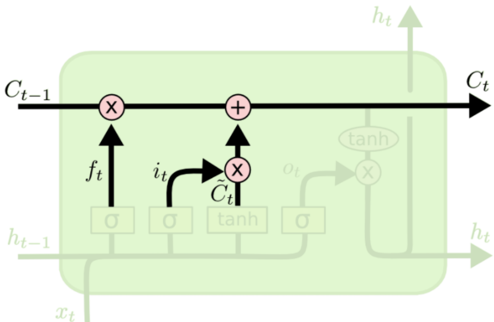
- On the other hand, a layer with an hyperbolic tangent activation function generates a vector of new candidate values $\tilde{C}_t$ potentially to be added to the cell state.

These two vectors are then combined to build up an update to the state:



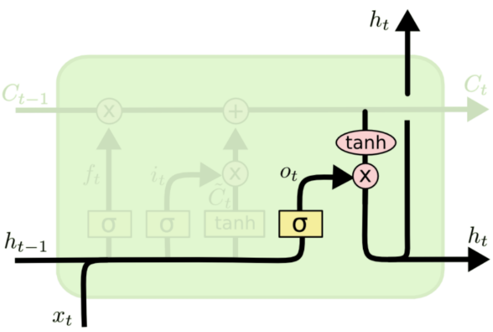
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \ + \ b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \ + \ b_C)$$

The old cell state vector C_{t-1} is now updated into the new one C_t, multiplying C_{t-1} by f_t component wise, and then adding $i_t * \tilde{C}_t$. This is the new candidate vector, scaled by how much previous steps decided to update each of its state value:



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Finally, the next hidden state vector h_t will be based on this brand new cell state, but will be a filtered version. First, a sigmoid layer determines, on the basis of both the input vector to the LSTM x_t and the previous unit hidden state vector h_{t-1}, to what extent the cell state C_t components will be part of h_t. That gives us what is called the output gate's activation vector o_t. Then, we put the cell state through a tanh activation layer (to push its values between 1 and 1) and multiply it component wise by o_t, so that the LSTM unit only keeps in the hidden state the scaled components of C_t that were previously chosen.



$$o_t = \sigma\left(W_o \ [h_{t-1}, x_t] \ + \ b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$

The sequential adjustments are operated with weights and bias terms that are optimized during the training phase. Those parameters optimization aims to minimize the specified loss objective function of the given problem across the sample training.

F Credibility Theory

We develop in this appendix section an overview of the mathematical concepts (with their interpretations) underlying the theory of credibility. Most of it is no more than a summary of relevant (for our current work) parts of the book "A Course in Credibility Theory and its Applications" by Hans Bühlmann and Alois Gisler [3].

F.1 Introduction and Bayesian framework

F.1.1 Introduction

Credibility theory belongs mathematically to the area of Bayesian statistics as a special type of statistical inference.The application of such a theory fits well when one have multiple estimates of an unknown quantity and wish to combine these estimates in such a way to get a more accurate and relevant estimate. It has numerous possible application in many different area.

We will not yet use formalized mathematical notations and framework for now. Instead, let's just focus on the intuition behind the credibility theory.

From an actuarial point of view, the theory of credibility can be applied by actuaries to improve statistical estimates. We will see that this approach can be interpreted through either a frequentist or a Bayesian statistical framework. One classical application entails an estimate X based on a small set of observed data, and an estimate M based on a larger but less relevant set of observed data. X and M can be any scalar quantities (estimated premiums, claims numbers, scores, loss ratios, technical reserves, etc.) vectors (claims development factors, etc.), or even more complex objects. The corresponding credibility estimate is then $\alpha X + (1 - \alpha)M$, where α is a number between 0 and 1 (called the "credibility weight" or "credibility factor") calculated to balance the sampling error of X against the possible lack of relevance (and therefore modeling error) of M.

Let's begin with a typical and introductory application example: the insurance problematic of differentiating premium values for homogeneous risk subgroups constituting a given insurance portfolio.

For the purpose of rating risks and calculating a premium to be charged, an insurance company generally groups its insured risks into classes of "similar risks", on the basis of certain of their characteristics. In motor insurance, those characteristics can be cylinder capacity, car brands and power/weight ratio, as well as individual characteristics such as the driver's age, sex and region (a young man driving a fast car being considered a high risk, while a middle age old woman driving a small car being considered a low risk). In industrial fire insurance, important characteristics might be the construction type and year of the insured building, the kind of business conducted or the fire extinguishing facilities in that building. The division is made balancing the two requirements that the risks in each group are sufficiently similar and the group sufficiently large that a meaningful statistical analysis of the claims experience can be done to calculate the risk premium to apply.

Within the credibility theory framework we do not consider each risk (or risk subgroup) individually but rather as being embedded in a larger group (here the whole portfolio), called the collective. The goal being to determine the "correct" premium for each subgroup of homogeneous insurance contracts, taking into account not only the individual experience with the subgroups, but also the collective experience of the whole portfolio, it is then necessary to devise a way of combining in an intelligent manner the two historical claims experiences. Credibility theory provides a solution to this issue.

We can clearly discern two extreme positions:

- One is to charge everyone the same premium P^{coll} for every member (risk subgroup) of the collective (portfolio), estimated by the overall mean $P^{coll} = \bar{X}$ of the portfolio observed data. This makes sense only if the portfolio is homogeneous, which means that all risk subgroups have identical mean claims. The total premium will then be equal to the expected aggregate claim amount, and thus a balanced situation will be reached. However, if the portfolio is heterogeneous, and the insurance company sets rates at the same level for all risks in this heterogeneous collective, the "good" risks will pay too much and the "bad" ones too little, possibly leading to disastrous consequences, such as adverse selection, in a competitive insurance market (the "good" risks will take their business elsewhere, leaving the insurer with only "bad" risks).
- The other way around is to charge to group i its own historical average claims, $P_i^{ind} = \overline{X_i}$, as an individual risk premium.

Competition forces companies to differentiate their rates and offer the fairest ones possible. That is why, it is relevant for an insurance company to assess the "correct" individual premiums P^{cred}, for being the most competitive rates to apply to each of the considered homogeneous risk subgroups.

With the view to calculate the "best" individual risk premium (i.e. the theoretical expected claims amount) for a particular risk subgroup i, the insurer will likely compromises the two previous extreme positions and resort to both the estimate of the portfolio historic overall claims experience, as well as the more specific estimate for the subgroup in question. Assigning a credibility factor, α_i, to the subgroup experience (and the reciprocal to the overall claims experience) allows the insurer to get a more accurate estimate of the risk premium in the following manner:

$$P_i^{cred} = \alpha_i P_i^{ind} + (1 - \alpha_i) P^{coll}$$

α_i has the following intuitive meaning: it expresses how "credible" the individual i observations are. Higher is the level of confidence in individual observations, higher will be the weight attached to the individual estimator (in our case $P_i^{ind} = \overline{X_i}$) in the weighted average. If the overall collective group were completely homogeneous then it would be reasonable to set $\alpha_i = 0$ for all i. On the contrary, if the collective were completely heterogeneous (with completely homogeneous risk subgroups), then we could consider $\alpha_i = 1$ for all i. Using intermediate values is reasonable to the extent that both individual subgroups' and overall group's historical observations are useful in inferring future individual behavior.

Assuming the variances of P_i^{ind} and P^{coll} are known quantities taking on the values σ_i and τ respectively, we will show that the credibility factors α_i should be equal to: $\alpha_i = \frac{\tau}{\sigma_i + \tau}$ Therefore, the more uncertainty the estimate has, the lower is its credibility.

To sum up, credibility theory is based on the two fundamental concepts of the "individual" within a "collective", and provides a procedure, backed by a rigorous and sound mathematical framework, to exploit optimally all the information stemming from these two sources. It is a mathematical tool to describe heterogeneous collectives and measure the confidence one should have in individual experience (observations) relatively to an encompassing collective's. It aims at individualizing an estimated quantity based on collective data, such as combining individual and collective claims experience to reach a "correct" individual risk premium.

F.1.2 Bayesian inference

Let's denote $\boldsymbol{x} = (x_1, \ldots, x_n)$ realizations of random variables $X_1, \ldots, X_n$. In the case i.i.d., those variables follow the same law $X_i \overset{\text{i.i.d.}}{\sim} X$. Non-parametric estimation doesn't involve assumptions on the law of X, whereas parametric estimation presuppose that $X \sim F_X(.;\theta)$ where θ is the parameter of interest. Frequentist inference assumes that θ is an unknown deterministic real value, whereas Bayesian inference assumes that θ is the value observed of a random variable Θ. Generally, Θ follows a continuous law with density $f_\Theta(.)$, known as the a priori law or the prior distribution.

Using the relationship for conditional probabilities $P(A \mid B) = P(A \cap B)/P(B)$, we define the marginal law, for the discrete case, as:

$$P(X = x) = \int_{\mathbb{R}} P(X = x \mid \Theta = \theta) dF_\Theta(\theta) = \int_{\mathbb{R}} p_{X|\Theta=\theta}(x) f_\Theta(\theta) d\theta$$

and for the continuous case,

$$f_X(x) = \int_{\mathbb{R}} f_{X|\Theta=\theta}(x) dF_\Theta(\theta) = \int_{\mathbb{R}} f_{X|\Theta=\theta}(x) f_\Theta(\theta) d\theta$$

For a given observed sample $\boldsymbol{x} = (x_1, \ldots, x_n)$, we then define different likelihoods.

The conditional likelihood:

$$\mathcal{L}(\boldsymbol{x} \mid \theta) = \begin{cases} \prod_{i=1}^n p_{X|\Theta=\theta}(x_i) & \text{if discrete variable} \\ \prod_{i=1}^n f_{X|\Theta=\theta}(x_i) & \text{if continuous variable} \end{cases}$$

The marginal likelihood:

$$\mathcal{L}(\boldsymbol{x}) = \begin{cases} \prod_{i=1}^n \int_{\mathbb{R}} p_{X|\Theta=\theta}(x_i) f_\Theta(\theta) d\theta & \text{if discrete variable} \\ \prod_{i=1}^n \int_{\mathbb{R}} f_{X|\Theta=\theta}(x_i) f_\Theta(\theta) d\theta & \text{if continuous variable} \end{cases}$$

The posterior (or a posteriori) likelihood, which is the one of interest:

$$\mathcal{L}(\theta \mid \boldsymbol{x}) = \frac{\mathcal{L}(\boldsymbol{x} \mid \theta) f_\Theta(\theta)}{\mathcal{L}(\boldsymbol{x})} \propto \mathcal{L}(\boldsymbol{x} \mid \theta) f_\Theta(\theta).$$

The denominator is indeed a normalization constant independent of θ. Bayesian estimator is generally defined as the expectation of the posterior law.

$$\hat{\theta}^B = E[\Theta \mid \boldsymbol{X} = \boldsymbol{x}] = \int_{\mathbb{R}} \theta f_{\Theta|\boldsymbol{X}=\boldsymbol{x}}(\theta) d\theta = \frac{\int_{\mathbb{R}} \theta \mathcal{L}(\theta, \boldsymbol{x}) f_\Theta(\theta) d\theta}{\mathcal{L}(\boldsymbol{x})}$$

Bayesian estimator happens to be the one that minimizes the marginal mean squared error

$$\hat{\theta}^B = \arg\min_{\hat{\theta}} E\left[(\hat{\theta} - \Theta)^2 \mid \boldsymbol{X}\right]$$

In the special case where the prior distribution p_Θ is assumed to come from a family of distributions called conjugate priors, the corresponding posterior distribution $p_{\Theta|X}$ will be in the same probability distribution family, and the calculation of the bayesian estimator's integrals may be expressed in closed form. In other cases, numerical procedures, such as Monte Carlo method, have to be implemented to approximate those integrals.To summarize, a conjugate prior is an algebraic convenience, giving a closed-form expression for the posterior.

The law of Θ can depend on parameters, called hyperparameters. They can either be estimated by expert judgement or through maximization of the marginal likelihood. One other option is to resort to a parameter-free distribution, like the uniform law or a centered and standardized distribution ($\mathcal{N}(0,1)$ for example).

When the a posteriori distribution is known, one can determine confidence intervals for θ:

$$IC_\alpha(\theta) = [y, z], \text{ such that } P(y \leq \Theta \leq z \mid \boldsymbol{X} = \boldsymbol{x}) = 1 - \alpha.$$

F.1.3 Bayesian premium and heterogeneity structure

Let's continue to use the intuitive thematic of insurance pricing. We must nonetheless keep in mind that the formulas we are going to introduce are way more general and can be applied to any other similar problematic of defining a credible individual estimator on the basis of both individual and collective observed data.

We then place ourselves in the intentionally restricted perimeter of an individual risk (or risk class) which can be regarded as a black box that produces a random (the past periods observed values could have been different) aggregate claim amounts history $\boldsymbol{X}_{i,\leq n} = (X_{i,1}, \ldots, X_{i,n})$ $1 \leq i \leq I$ and $1 \leq j \leq n$., where I is the number of individuals considered in the studied collective and $X_{i,j}$ denotes the claim amount in year j (or some other well-specified time period j). The "black box" associated with an individual could be:

- an individual driver in third-party motor liability insurance portfolio,
- a specified group of insured individuals in a life insurance portfolio,
- all employees of a firm in collective workman's compensation insurance,
- the France property risks segment in an international P&C insurance portfolio,
- USD currency casualty reinsurance treaties in the whole casualty reinsurance business of a reinsurer,
- a ceding company in a reinsurance portfolio.
- an insurance company's life business among other market participant's.
- etc.

The objective contemplated here is to determine the risk premium for the aggregate claims in a future period, for example $X_{i,n+1}$.
For that, we make the following assumptions:

- $X_{i,j} \mid \Theta_i = \theta \overset{\text{i.i.d.}}{\sim} X_\theta$, i.e. the variables of interest are conditionally i.i.d. with distribution $F_X(\cdot;\theta)$;
- $\Theta_i \overset{\text{i.i.d.}}{\sim} \Theta$, i.e. the heterogeneity structure is i.i.d.;
- $(X_{i,1}, \ldots, X_{i,n}, \Theta_i)_{i \in [\![1,I]\!]}$ (i.e. the individuals) are independent.

The heterogeneity structure of the considered collective, embedded in the random variable Θ and its distribution, represents and materializes the dissimilarities of its constituting elements (individuals diverse in character or content). As already explained, the collective (a given risk portfolio, a given line of business, etc.) is indeed composed of different and dissimilar subcomponents that we called individuals, each one having it own specificities.

To go back to our concrete example of individual risk (or risk subgroup) rating problem, every risk i in a particular collective is characterized by its individual risk profile θ_i. These parameters θ_i are realisations of Θ, where Θ covers randomly the set of all potential and possible values of the (unknown) risk profiles in the collective. In the special case of a completely homogeneous collective, Θ is deterministic and consists of just one element. This corresponds to the classical point of view of insurance: every member of the collective has exactly the same risk profile and therefore the same distribution function for its corresponding aggregate claim amount. However, the risk groups or collectives considered in insurance are in truth mostly heterogeneous. In other words, the θ-values of the different risks in the collective are not all the same. They are rather samples taken from a random Θ covering more than one element. But while the risks in the collective are different, they also have something in common: they all belong to the same collective, i.e. they are drawn from the same distribution.

The specific θ-values attached to the different risks in the collective are typically unknown to the insurer. But, on the basis of a priori knowledge and statistical information, the insurer does know something about the structure of the collective. He knows, for example, that most car drivers are "good" risks and seldom make a claim, while a small percentage of drivers make frequent claims. Formally, this information can be summarized by a probability distribution, thus the relevance of modeling heterogeneity through a random variable Θ. Being able to identify individual risk profiles, and thus having a good representation of individual risks, is equivalent to finding an estimator of θ for each individual risk.

The probability distribution of Θ is called the structural function of the collective, the interpretation of which is not unique:

- In the frequentist (or empirical Bayes) interpretation, we consider the θ 's in the collective as being a random sample from some fixed set. The structural function then describes the idealized frequencies of the θ 's over that set.
- In the pure Bayesian interpretation we consider the distribution function of Θ as a description of the personal beliefs, a priori knowledge, and experience of the actuary.

In such a context, we define different premiums.

1. The individual premium: $\mu\left(\theta_i\right)=E\left[X_{i,n+1} \mid \Theta_i=\theta_i\right]$.

2. The collective premium: $\mu_0=E\left[X_{i,n+1}\right]=E[\mu(\Theta)]=\int_{\mathbb{R}} \mu(\theta) dF_{\Theta}(\theta)$.

We can remark that, for $k \neq j$, variables $X_{i,j}$ and $X_{i,k}$ are positively correlated for being conditionally i.i.d., whatever the law of Θ. As a matter of fact, for all i, and $k \neq j$

$$\begin{aligned}
\operatorname{Cov}\left[X_{i,j}, X_{i,k}\right] & =\operatorname{Cov}\left[E\left[X_{i,j} \mid \Theta_i\right], E\left[X_{i,k} \mid \Theta_i\right]\right]+E\left[\operatorname{Cov}\left[X_{i,j}, X_{i,k} \mid \Theta_i\right]\right] \\
& =\operatorname{Cov}[\mu(\Theta), \mu(\Theta)]+0=\operatorname{Var}[\mu(\Theta)]
\end{aligned}$$

But this correlation is smaller than the variance:

$$\text{Var}\left[X_{i,j}\right] = \text{Var}\left[E\left[X_{i,j} \mid \Theta_i\right]\right] + E\left[\text{Var}\left[X_{i,j} \mid \Theta_i\right]\right] = \text{Var}[\mu(\Theta)] + E[\text{Var}[X \mid \Theta]]$$

3. The Bayesian premium:

$$\widehat{\mu(\Theta)}_{i,n+1}^{B} = E\left[\mu(\Theta) \mid \boldsymbol{X}_{i,\leq n} = \boldsymbol{x}_{i,\leq n}\right] = \int_{\mathbb{R}} \mu(\theta) dF_{\Theta \mid \boldsymbol{X}_{i,\leq n} = \boldsymbol{x}_{i,\leq n}}(\theta)$$

with $F_{\Theta \mid \boldsymbol{X}_{i,\leq n} = \boldsymbol{x}_{i,\leq n}}$ the posterior distribution , previously designated $\mathcal{L}(\theta \mid \boldsymbol{x})$.

If we suppose furthermore that Θ follows a continuous law with density f_Θ, we get

$$\mu_0 = \int_{\mathbb{R}} \mu(\theta) f_\Theta(\theta) d\theta \quad \text{and} \quad \widehat{\mu(\Theta)}_{i,n+1}^{B} = \int_{\mathbb{R}} \mu(\theta) f_{\Theta \mid \boldsymbol{X}_{i,\leq n} = \boldsymbol{x}_{i,\leq n}}(\theta) d\theta$$

One can show that the Bayesian premium minimizes the marginal mean squared error

$$\widehat{\mu(\Theta)}^{B} = \underset{\widehat{\mu(\Theta)}}{\arg\min}\, E\left[\left(\widehat{\mu(\Theta)} - \mu(\Theta)\right)^2\right]$$

which is why it is known as the best experience premium.

We can compare the mean squared errors reached by the Bayesian premium and by the collective premium:

$$\begin{aligned} E\left[\left(\widehat{\mu(\Theta)}_{i,n+1}^{B} - \mu(\Theta)\right)^2\right] &= E\left[\text{Var}\left[\mu(\Theta) \mid \boldsymbol{X}_{i,\leq n}\right]\right] \leq E\left[\text{Var}\left[\mu(\Theta) \mid \boldsymbol{X}_{i,\leq n}\right]\right] + \text{Var}\left[E\left[\mu(\Theta) \mid \boldsymbol{X}_{i,\leq n}\right]\right] \\ &\leq E\left[(\mu_0 - \mu(\Theta))^2\right] \end{aligned}$$

The procedure for calculating the Bayesian premium is therefore as follows:

1. list the hypotheses;
2. write the approximate expression of $\mathcal{L}(\theta \mid \boldsymbol{x})$;
3. simplify the terms so as not to keep the terms dependent on theta;
4. identify a known distribution (with its parameters) for $\mathcal{L}(\theta \mid \boldsymbol{x})$;
5. conclude on the value of $\widehat{\mu(\Theta)}_{i,n+1}^{B}$.

As already mentioned, if the posterior distribution $\Theta \mid X$ is in the same probability distribution family as the prior probability distribution of Θ, the prior and posterior are then called conjugate distributions, and the Bayesian premium can be expressed explicitly. Hyperparameters of Θ distribution will have nonetheless to be determined.

F.2 Credibility Estimators

We have seen that the Bayes premium $\widehat{\mu(\Theta)}^{B} = E[\mu(\Theta) \mid \mathbf{X}]$ is the best possible estimator in the class of all estimator functions. In general, however, this estimator cannot be expressed in a closed analytical form and can only be calculated by numerical procedures. Therefore it does not fulfil the requirement of simplicity. Moreover, to calculate $\widehat{\mu(\Theta)}^{B}$, one has to specify the conditional distributions as well as the a priori distribution, which, in practice, can often neither be inferred from data nor guessed by intuition.

The basic idea underlying credibility is to force the required simplicity of the estimator by restricting the class of allowable estimator functions to those which are linear in the observations $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$. In other words, we look for the best estimator in the class of all linear estimator functions. "Best" is to be understood in the Bayesian sense and the optimality criterion is again quadratic loss. Credibility estimators are therefore linear Bayes estimators.

To develop this part, we will make use of the same notations as previously introduced, but we will not keep the individuals identification index i. Thus, $\mathbf{X}$ stands for the observations of the individual (risk, risk subgroup, etc.) and $\mu(\Theta)$ for the individual premium. The components of $\mathbf{X}$ are once again assumed to be, conditional on $\Theta = \theta$, independent and identically distributed.

We will deal with a more general interpretation and we shall define the credibility estimator in a general set-up. We will also see that the credibility estimators can be understood as orthogonal projections in the Hilbert space of square integrable random variables, and we will prove some general characteristics and properties.

F.2.1 The credibility premium in a simple credibility model

In Bayesian credibility, we separate each class (individuals belongs to a given class i.e. $\Theta_i = \theta$) and assign them a probability (Θ distribution). Then we find how likely our experience ($\boldsymbol{X}_{i,\leq n} = \boldsymbol{x}_{i,\leq n}$) is within each class (probability of $\boldsymbol{X}_{i,\leq n} = \boldsymbol{x}_{i,\leq n}$ given $\Theta_i = \theta$). Next, we find how likely our experience was over all classes (Probability of $\boldsymbol{X}_{i,\leq n} = \boldsymbol{x}_{i,\leq n}$). Finally, we can find the probability of our class given the corresponding experience. So going back to each class, we weight each statistic with the probability of the particular class given the experience.

As its name suggests, the purpose of non-parametric credibility is to relax the parametric assumptions made about the prior law and the law of heterogeneity.

We consider the following simple credibility model assumptions:

- The random variables $(X_j)_{j\in[\![1,n]\!]}$ are, conditional on $\Theta = \theta$, independent with the same distribution function F_θ, with conditional moments $\mu(\theta) = E[X_j \mid \Theta = \theta]$ and $\sigma^2(\theta) = \operatorname{Var}[X_j \mid \Theta = \theta]$;
- Θ is a random variable with distribution F_Θ.

In this model we have

$$P^{\text{ind}} = \mu(\Theta) = E[X_{n+1} \mid \Theta] \quad \text{and} \quad P^{\text{coll}} = \mu_0 = \int_\Theta \mu(\theta) dF_\Theta(\theta)$$

Our aim is again to find an estimator for the individual premium $\mu(\Theta)$, but this time, which are linear in the observations. We will denote the best estimator within this class by P^{cred} or $\widehat{\mu(\Theta)}^{cred}$, which we are going to derive now. By definition, $\widehat{\mu(\Theta)}^{cred}$ has to be of the form $\widehat{\mu(\Theta)}^{cred} = \widehat{a}_0 + \sum_{j=1}^n \widehat{a}_j X_j$, with $\widehat{a}_0, \widehat{a}_1, \ldots, \widehat{a}_n$ such that

$$E\left[\left(\mu(\Theta) - \widehat{a}_0 - \sum_{j=1}^n \widehat{a}_j X_j\right)^2\right] = \min_{(a_0,a_1,\ldots,a_n)\in\mathbb{R}^n} E\left[\left(\mu(\Theta) - a_0 - \sum_{j=1}^n a_j X_j\right)^2\right]$$

Since the probability distribution of $X_1, \ldots, X_n$ is invariant under permutations of X_j and $\widehat{\mu(\Theta)}^{cred}$ is uniquely defined, one must have $\widehat{a}_1 = \widehat{a}_2 = \cdots = \widehat{a}_n$, i.e. the estimator $\widehat{\mu(\Theta)}^{cred}$ has the form

$$\widehat{\mu(\Theta)}^{cred} = \widehat{a} + \widehat{b}\bar{X}, \quad \text{where} \quad \bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$$

where $\widehat{a}$ and $\widehat{b}$ are the solution of the minimizing problem

$$E\left[(\mu(\Theta) - \widehat{a} - \widehat{b}\bar{X})^2\right] = \min_{(a,b)\in\mathbb{R}^2} E\left[(\mu(\Theta) - a - b\bar{X})^2\right]$$

Taking partial derivatives with respect to a and b, we get

$$E[\mu(\Theta) - a - b\bar{X}] = 0 \quad \text{and} \quad \operatorname{Cov}(\bar{X}, \mu(\Theta)) - b\operatorname{Var}(\bar{X}) = 0$$

From the dependency structure imposed by the simple model assumptions we have

$$\operatorname{Cov}(\bar{X}, \mu(\Theta)) = \operatorname{Var}(\mu(\Theta)) =: \tau^2$$
$$\operatorname{Var}(\bar{X}) = \frac{E\left[\sigma^2(\Theta)\right]}{n} + \operatorname{Var}(\mu(\Theta)) =: \frac{\sigma^2}{n} + \tau^2$$

from which we obtain

$$b = \frac{\tau^2}{\tau^2 + \sigma^2/n} = \frac{n}{n + \sigma^2/\tau^2} \quad \text{and} \quad a = (1-b)\mu_0$$

We finally have the following expression for the credibility estimator under the considered simple model assumptions:

$$\widehat{\mu(\Theta)}^{cred} = \alpha\bar{X} + (1-\alpha)\mu_0, \quad \text{where} \quad \mu_0 = E[\mu(\Theta)] \quad \text{and} \quad \alpha = \frac{n}{n + \sigma^2/\tau^2}$$

Let us highlight the following points:

- P^{cred} is a weighted mean of P^{coll} and the individual observed average $\bar{X}$.
- The quotient $\kappa = \sigma^2/\tau^2$ is called the credibility coefficient, which can also be written as $\kappa = (\sigma/\mu_0)^2 (\tau/\mu_0)^{-2}$. Note that τ/μ_0 is the coefficient of variation of $\mu(\Theta)$, which is a good measure for the heterogeneity of the portfolio, whereas $\sigma/\mu_0 = \sqrt{E\left[\operatorname{Var}\left[X_j \mid \Theta\right]\right]}/E\left[X_j\right]$ is the expected standard deviation within risk divided by the overall expected value, which is a good measure for the within risk variability.
- The credibility weight α increases as the number of observed years n increases, the heterogeneity of the portfolio (as measured by the coefficient of variation τ/μ_0) increases, the within risk variability (as measured by σ/μ_0) decreases.
- The formula for P^{cred} involves the structural parameters σ^2, τ^2 and μ_0. If there exists a collective composed of similar risks, these parameters can be estimated using the data from this collective (empirical Bayes procedure). In next section we will explicit estimators for these parameters in a more general model. The sizes of the structural parameters could also be intuitively "decided" using the a priori knowledge of an experienced actuary or underwriter (pure Bayesian procedure).
- The class of estimators that are linear in the observations is a subclass of the class of all estimators based on the observations. Hence the Bayes estimator is equal to the credibility estimator, if the former is linear. We refer to such cases as exact credibility.

The credibility premium as a weighted mean of P^{coll} and $\bar{X}$ can also be interpreted as follows:

- $P^{\text{coll}} = \mu_0$ is the best estimator based on the a priori knowledge alone. It has the quadratic loss $E\left[(\mu_0 - \mu(\Theta))^2\right] = \operatorname{Var}(\mu(\Theta)) = \tau^2$.

- $\bar{X}$ is the best possible linear and individually unbiased (i.e. conditionally unbiased given Θ) estimator, based only on the observation vector $\mathbf{X}$. It has the quadratic loss $E\left[(\bar{X} - \mu(\Theta))^2\right] = E\left[\sigma^2(\Theta)/n\right] = \sigma^2/n$.

- $P^{\text{cred}} = \widehat{\mu(\Theta)}^{cred}$ is a weighted mean of these two, where the weights are proportional to the inverse quadratic loss (precision) associated with each of the two components, i.e.

$$\widehat{\mu(\Theta)}^{cred} = \alpha\bar{X} + (1-\alpha)\mu_0 \quad \text{where } \alpha = \frac{n/\sigma^2}{n/\sigma^2 + 1/\tau^2} = \frac{\tau^2}{\tau^2 + \sigma^2/n} = \frac{n}{n + \sigma^2/\tau^2}$$

This is a very intuitive principle. For the estimation of the pure risk premium, two sources of information are available: the a priori knowledge and the individual observations. First, one looks and sees what one can learn from each of these two sources on its own. The a priori knowledge contains information about the collective and the best estimator that we can derive from this information is the a priori expectation μ_0. On the other hand, the observations contain information about the individual risk and the individual risk profile Θ. It is reasonable here to consider linear estimators that are individually, i.e. conditionally, unbiased and to choose from these the one with the greatest precision, i.e. the smallest variance. Finally, we have to weight the estimators derived from each of these information sources. It is intuitively reasonable to weight them according to their precision, respectively the inverse value of the quadratic loss. This simple principle applies also in more general probability frameworks.

With this simple model framework assumptions, we have the following quadratic loss of the different estimators for $\mu(\Theta)$:

Premium	Quadratic loss
a) $P^{coll} = \mu_0$	$\operatorname{Var}(\mu(\Theta))$
b) $P^{cred} = \widehat{\mu(\Theta)}^{cred}$	$(1-\alpha)\operatorname{Var}(\mu(\Theta))$
c) $P^{Bayes} = \widehat{\mu(\Theta)}^{B}$	$E[\operatorname{Var}(\mu(\Theta) \mid \mathbf{X})]$

Regarding the quadratic loss we have that $a) \geq b) \geq c$). The improvement from a) to b) is, as a rule, considerable. The closer that the Bayes premium is to a function which is linear in the observations, the smaller is the improvement from b) to c). The advantage of the credibility premium is its simplicity and the fact that, contrary to P^{Bayes}, we don't have to specify the family of kernel distributions and the prior distribution, a specification which would imply an additional model risk. The somewhat greater loss associated with b) over c) has therefore to be looked at in relative terms and is often acceptable.

F.2.2 The Bühlmann model and the homogeneous credibility estimator

Once again, we will speak of individual risks (or risk sub-groups) contained in a collective portfolio, but what follows can be generalized to any type of individuals constituting a collective. The same remark holds for the quantity of interest (here historical claims amounts).

So far we have derived the credibility estimator based only on the observations of one particular individual risk or risk subgroup. We will now consider a whole portfolio (collective) of similar risks (individuals) numbered $i = 1, 2, \ldots, I$. We denote by $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{in})'$ the observation vector of risk i, and by Θ_i its risk profile.

Bühlmann model assumptions are for $i \in [\![1, I]\!]$ and $j \in [\![1, n]\!]$,

- $\forall k \neq j, X_{i,j} \perp X_{i,k} \mid \Theta_i = \theta$, i.e. the variables of interest are conditionally independents;
- The first two conditional moments are finite $E\left[X_{i,j} \mid \Theta_i = \theta\right] = \mu\left(\theta\right) < +\infty$ and $\text{Var}\left[X_{i,j} \mid \Theta_i = \theta\right] = \sigma^2\left(\theta\right) < +\infty$;
- $\Theta_i \overset{\text{i.i.d.}}{\sim} \Theta$, i.e. the heterogeneity structure is i.i.d., and allows that the risks in the portfolio have different risk profiles;
- The pairs $(\Theta_1, \mathbf{X}_1), \ldots, (\Theta_I, \mathbf{X}_I)$ are i.i.d., i.e. an heterogeneous collective (portfolio) is modelled within which the individuals (risks) are independent and a priori equal (they cannot be recognized as being different).

On the basis of these assumptions we will strive to estimate for each risk i its individual premium $\mu(\Theta_i)$, that is to say, find the corresponding credibility estimators $\widehat{\mu(\Theta_i)}^{cred}$ for $i = 1, 2, \ldots, I$. The credibility estimator $\widehat{\mu(\Theta_i)}^{cred}$ can either be a linear function of the observations of risk i only, or a linear function of all observations in the portfolio. Fist of all, it is then important to specify the quantity that we want to estimate and the statistics that the credibility estimator should be based on.

Generally, the credibility estimator is defined as the best estimator which is a linear function of all observations in the portfolio, i.e. the credibility estimator $\widehat{\mu(\Theta_i)}^{cred}$ of $\mu(\Theta_i)$ is by definition the best estimator in the class

$$\left\{\widehat{\mu\left(\Theta_i\right)}: \quad \widehat{\mu\left(\Theta_i\right)} = a + \sum_{k=1}^{I}\sum_{j=1}^{n} b_{kj}X_{kj}, \quad a, b_{kj} \in \mathbb{R}\right\}$$

By the same invariance and permutation arguments as in the first simple credibility model exposed previously, we find that $\widehat{\mu(\Theta_i)}^{cred}$ must be of the form

$$\widehat{\mu(\Theta_i)}^{cred} = \widehat{a}_0^{(i)} + \sum_{k=1}^{I} \widehat{b}_k^{(i)} \bar{X}_k, \quad \text{where } \bar{X}_k = \frac{1}{n}\sum_{j=1}^{n} X_{kj}$$

To find the coefficients $\widehat{a}_0^{(i)}$ and $\widehat{b}_k^{(i)}$ we have to minimize

$$E\left[\left(\mu\left(\Theta_i\right) - a_0^{(i)} - \sum_{k=1}^{I} b_k^{(i)} \bar{X}_k\right)^2\right]$$

Taking partial derivatives with respect to $b_k^{(i)}$ and $a_0^{(i)}$ we find

$$\text{Cov}\left(\mu\left(\Theta_i\right), \bar{X}_k\right) = b_k^{(i)} \,\text{Var}\left[\bar{X}_k\right]$$

Since the left-hand side is equal to zero for $i \neq k$, it follows that $\widehat{b}_k^{(i)} = 0$ for $k \neq i$. Hence in this model the credibility estimator of $\mu\left(\Theta_i\right)$ depends only on $\bar{X}_i$, the observed mean of risk i, and not on the observations of the other risks in the collective. Therefore, the credibility estimator based on all observations of the portfolio and not only on the corresponding individual ones (i.e. based on $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2', \ldots, \mathbf{X}_I')'$ and not only on $\mathbf{X}_i$) is again:

$$\widehat{\mu\left(\Theta_i\right)} = \alpha \bar{X}_i + (1-\alpha)\mu_0 \text{ with } \alpha = \frac{n}{n + \sigma^2/\tau^2}$$

Given a whole portfolio of risks, we can also consider another type of credibility estimator, and define the homogeneous credibility estimator $\widehat{\mu(\Theta_i)}^{hom}$ of $\mu(\Theta_i)$ as the best estimator in the class of collectively unbiased estimators

$$\left\{ \widehat{\mu\left(\Theta_i\right)} : \quad \widehat{\mu\left(\Theta_i\right)} = \sum_{k=1}^{I} \sum_{j=1}^{n} b_{kj} X_{kj}, \quad E\left[\widehat{\mu\left(\Theta_i\right)}\right] = E\left[\mu\left(\Theta_i\right)\right], \quad b_{kj} \in \mathbb{R} \right\}$$

Contrary to $\widehat{\mu(\Theta_i)}^{cred}$, the homogeneous estimator $\widehat{\mu(\Theta_i)}^{hom}$ does not contain a constant term, but is required to be unbiased over the collective, a condition which is automatically fulfilled for the inhomogeneous credibility estimator.

$$\widehat{\mu(\Theta_i)}^{hom} = \underset{b_{kj} \in \mathbb{R}}{\arg\min}\, E\left[\left(\mu(\Theta_i) - \sum_{k=1}^{I} \sum_{j=1}^{n} b_{kj} X_{kj}\right)^2\right]$$

under the unbiasedness requirement constraint $\sum_{k=1}^{I} \sum_{l=1}^{n} b_{kl} = 1$ resulting from

$$E\left[\widehat{\mu(\Theta_i)}^{hom}\right] = E\left[\mu\left(\Theta_i\right)\right] = E\left[X_{kj}\right] = \mu_0 \quad \text{for all } i, j, k$$

We can obtain these homogeneous credibility estimators from the inhomogeneous ones, simply replacing the overall expected value μ_0 by a suitable linear function of all observations, namely the observed collective mean $\bar{X} = \frac{1}{In} \sum_{i=1}^{I} \sum_{j=1}^{n} X_{ij}$. The homogeneous credibility estimator thus contains a built-in estimator for the overall mean μ_0.

Finally, The (inhomogeneous) credibility estimator $\widehat{\mu(\Theta_i)}^{cred}$ and the homogeneous credibility estimator $\widehat{\mu(\Theta_i)}^{hom}$ in the simple Bühlmann model are given by

$$\begin{aligned}
\widehat{\mu(\Theta_i)}^{cred} &= \alpha \bar{X}_i + (1-\alpha)\mu_0 \\
\widehat{\mu(\Theta_i)}^{hom} &= \alpha \bar{X}_i + (1-\alpha)\bar{X}
\end{aligned}$$

where

$$\alpha = \frac{n}{n + \frac{\sigma^2}{\tau^2}}, \quad \bar{X}_i = \frac{1}{n} \sum_{j=1}^{n} X_{ij}, \quad \bar{X} = \frac{1}{In} \sum_{i=1}^{I} \sum_{j=1}^{n} X_{ij}$$

F.2.3 A general perspective on credibility estimators

Until know, in the setting of our illustrative example, the quantity to be estimated was the individual risk premium in some future period of a particular risk (or risk subgroup) i, which we denoted by $\mu(\Theta) = \mu\left(\Theta_i\right) = E\left[X_{i,n+1} \mid \Theta_i\right]$. The underlying statistics for the credibility estimator was $\mathbf{X}$ containing the observations (claims data available to the insurer) from a whole portfolio of risks.

In the simple Bühlmann model, the quantity to be estimated and the observation vector $\mathbf{X}$ were well defined, and their probability structure well specified by accompanying model assumptions.

In the general set-up, the goal is also to find the credibility estimator of $\mu(\Theta)$ based on some observation vector $\mathbf{X}$. However, we do not define $\mu(\Theta)$ and $\mathbf{X}$ exactly, nor do we specify their probability structure. Hence the only mathematical structure in the general set-up is that we want to estimate some unknown real-valued random variable $\mu(\Theta)$ based on some known random vector $\mathbf{X} = (X_1, \ldots, X_n)'$. All results in this subsections are still valid if we replace $\mu(\Theta)$ by any other square integrable random variable.

Credibility estimators as the best estimators in an a priori given class

Given the observation vector $\mathbf{X}$, let us define two such classes $L(\mathbf{X}, 1)$ and $L_e(\mathbf{X})$ and their corresponding credibility estimators.

We thus consider the two sets of estimators:

$$L(\mathbf{X}, 1) := \left\{ \widehat{\mu(\Theta)} : \quad \widehat{\mu(\Theta)} = a_0 + \sum_{j=1}^{n} a_j X_j, \quad (a_0, a_1, \cdots, a_n) \in \mathbb{R}^{n+1} \right\}$$

and

$$L_e(\mathbf{X}) := \left\{ \widehat{\mu(\Theta)} : \quad \widehat{\mu(\Theta)} = \sum_{j=1}^{n} a_j X_j, \quad (a_1, \cdots, a_n) \in \mathbb{R}^{n}, \quad E\left[\widehat{\mu(\Theta)}\right] = E[\mu(\Theta)] \right\}$$

The credibility estimator of $\mu(\Theta)$ based on $\mathbf{X}$ is the best possible estimator in the class $L(\mathbf{X}, 1)$.

$$P^{cred} = \widehat{\mu(\Theta)}^{cred} = \underset{\widehat{\mu(\Theta)} \in L(\mathbf{X},1)}{\arg\min} \ E\left[(\widehat{\mu(\Theta)} - \mu(\Theta))^2\right].$$

The homogeneous credibility estimator of $\mu(\Theta)$ based on $\mathbf{X}$ is the best possible estimator in the class of collectively unbiased estimators $L_e(\mathbf{X})$

$$P^{cred\_hom} = \widehat{\mu(\Theta)}^{hom} = \underset{\widehat{\mu(\Theta)} \in L_e(\mathbf{X})}{\arg\min} \ E\left[(\widehat{\mu(\Theta)} - \mu(\Theta))^2\right].$$

The homogeneous credibility estimator related to a particular individual (risk, risk subgroup, ect.) is relevant only if the observation vector $\mathbf{X}$ also incorporates collateral data (other similar individuals' observation data comprising a collective). Moreover, as we already saw, forcing the constant term to be zero, together with the collectively unbiasedness condition, automatically implies a built-in estimator of μ_0, which is the main reason for considering the homogeneous credibility estimator.

Credibility estimators as orthogonal projections in the $\mathcal{L}^2$ Hilbert space

The estimators $\widehat{\mu(\Theta)}^{cred}$ and $\widehat{\mu(\Theta)}^{hom}$, defined as a solution of least squares problems, can also be understood as projections in the Hilbert space of all square integrable functions $\mathcal{L}^2$.

Inversion of a large matrix through computational techniques can always solve least squares optimization issues. Nevertheless, as the model structure becomes more complex, this procedure can rapidly grow inconvenient and give up the intuitive insight for the resulting credibility formulae. Hilbert space theory applied to credibility is not so burdensome and allows to take advantage of our intuitive understanding of linear vector spaces' properties.

We define

$$\mathcal{L}^2 := \left\{ X : \quad X = \text{ random variable with } E\left[X^2\right] = \int X^2 dP < \infty \right\}$$

For 3 random variables X, X' and Y elements of $\mathcal{L}^2$, and for M and M' closed subspaces (or closed affine subspaces) of $\mathcal{L}^2$ with $M' \subset M$, we have the following properties:

- $< X, Y >:= E[XY]$ (inner product)
- $\|X\| :=< X, X >^{1/2}= E\left[X^2\right]^{1/2}$ (corresponding norm)
- X and X' are considered identical if $E\left[(X - X')^2\right] = \|X - X'\|^2 = 0$
- $(X \perp Y)$ if and only if $< X, Y >= 0$
- For a closed subspace $M \subset \mathcal{L}^2$, $Y^* \in M$ is the orthogonal projection of Y on M, i.e. $Y^* = \text{Pro}(Y \mid M)$ if and only if $Y - Y^* \perp M$, i.e. $Y - Y^* \perp Z_1 - Z_2$ for all $Z_1, Z_2 \in M$ (Y^* always exists and is unique)
- $Y^* = \text{Pro}(Y \mid M) \iff Y^* \in M$ and $< Y - Y^*, Z - Y^* >=< Y - Y^*, Z >= 0 \quad \forall Z \in M \iff Y^* \in M$ and $\|Y - Y^*\| \leq \|Y - Z\| \quad \forall Z \in M$
- $\text{Pro}(Y \mid M') = \text{Pro}\left(\text{Pro}\left(Y \mid M\right) \mid M'\right)$ (iterativity of projections)
- $|Y - \text{Pro}(Y \mid M')\|^2 = \|Y - \text{Pro}\left(Y \mid M\right)\|^2 + \|\text{Pro}\left(Y \mid M\right) - \text{Pro}(Y \mid M')\|^2$ (theorem of Pythagoras)

Within this framework, we consider the random variables

- $\mu(\Theta) \in \mathcal{L}^2$ the individual premium to be estimated,
- $\mathbf{X}' = (X_1, X_2, \ldots, X_n)$ the individual observation vector with elements $X_j \in \mathcal{L}^2$,

as elements of the Hilbert space $\mathcal{L}^2$ with corresponding moments $\mu_0 = E[\mu(\Theta)]$ and $\boldsymbol{\mu}'_{\mathbf{X}} := (\mu_{X_1}, \ldots, \mu_{X_n}) = (E\left[X_1\right], \ldots, E\left[X_n\right])$, and with the covariance matrix of $\mathbf{X}$, $\Sigma_{\mathbf{X}} := \text{Cov}\left(\mathbf{X}, \mathbf{X}'\right)$.

In addition to already introduced space $L(\mathbf{X}, 1)$ and $L_e(\mathbf{X})$, we define

$$G(\mathbf{X}) := \left\{ Z : \quad Z = g(\mathbf{X}), \quad g = \text{real-valued function and} \quad g(\mathbf{X}) \in \mathcal{L}^2 \right\}$$

$L(\mathbf{X}, 1)$ and $G(\mathbf{X})$ are closed subspaces of $\mathcal{L}^2$, $L_e(\mathbf{X})$ is a closed affine subspace of $\mathcal{L}^2$.
From this mathematical perspective, we can now reformulate the definition of the Bayes as well as the credibility estimators of $\mu(\Theta)$ based on $\mathbf{X}$.

- The (inhomogeneous) credibility estimator is defined as $\widehat{\mu(\Theta)}^{cred} = \text{Pro}(\mu(\Theta) \mid L(\mathbf{X}, 1))$
- The homogeneous credibility estimator is defined as $\widehat{\mu(\Theta)}^{hom} = \text{Pro}\left(\mu(\Theta) \mid L_e(\mathbf{X})\right)$
- The Bayes estimator is defined as $\widehat{\mu(\Theta)}^{B} = \text{Pro}(\mu(\Theta) \mid G(\mathbf{X}))$

We thus interpret both the credibility estimator and the Bayes premium as orthogonal projections of the (unknown) individual premium $\mu(\Theta)$ on appropriately defined subspaces (respectively affine subspaces) of $\mathcal{L}^2$. The following relationships between the credibility estimators and the Bayes premium, are natural outcomes of this orthogonal projection point of view:

- $\widehat{\mu(\Theta)}^{cred}$ is the best linear approximation to $\widehat{\mu(\Theta)}^{B}$
- $\underbrace{E\left[(\widehat{\mu(\Theta)}^{cred} - \mu(\Theta))^2\right]}_{\text{total error}} = \underbrace{E\left[(\widehat{\mu(\Theta)}^{cred} - \widehat{\mu(\Theta)}^{B})^2\right]}_{\text{approximation error}} + \underbrace{E\left[(\widehat{\mu(\Theta)}^{B} - \mu(\Theta))^2\right]}_{\text{Bayes risk}}$
- $\widehat{\mu(\Theta)}^{hom}$ is the best homogeneous linear approximation to $\widehat{\mu(\Theta)}^{cred}$ as well as to $\widehat{\mu(\Theta)}^{B}$

These three results can be easily proved (from the iterativity of projections and the theorem of Pythagoras), given the fact that $L_e(\mathbf{X}) \subset L(\mathbf{X}, 1) \subset G(\mathbf{X})$ are subspaces of $\mathcal{L}^2$.

Orthogonality conditions and normal equations

Once again regarding the credibility estimators as orthogonal projections of $\mu(\Theta)$ on subspaces of $\mathcal{L}^2$, we derive the following orthogonality conditions (called also "normal equations") characterizing those estimators:

- $\widehat{\mu(\Theta)} = \widehat{\mu(\Theta)}^{cred}$
  $\iff \left\langle \mu(\Theta) - \widehat{\mu(\Theta)}, 1 \right\rangle = 0 \quad \text{and} \quad \left\langle \mu(\Theta) - \widehat{\mu(\Theta)}, X_j \right\rangle = 0 \quad \forall j \in [\![1, n]\!]$
  $\iff E[\mu(\Theta) - \widehat{\mu(\Theta)}] = 0 \quad \text{and} \quad \operatorname{Cov}[\mu(\Theta), X_j] = \operatorname{Cov}\left[\widehat{\mu(\Theta)}, X_j\right] \quad \forall j \in [\![1, n]\!]$

- $\widehat{\mu(\Theta)} = \widehat{\mu(\Theta)}^{hom}$
  $\iff \left\langle \mu(\Theta) - \widehat{\mu(\Theta)}, 1 \right\rangle = 0 \quad \text{and} \quad \left\langle \mu(\Theta) - \widehat{\mu(\Theta)}, Z - \widehat{\mu(\Theta)} \right\rangle = 0 \quad \forall Z \in L_e(\mathbf{X})$
  $\iff E\left[\mu(\Theta) - \widehat{\mu(\Theta)}\right] = 0$
  and $\operatorname{Cov}\left(\widehat{\mu(\Theta)} - \mu(\Theta), \widehat{\mu(\Theta)}\right) = \operatorname{Cov}\left(\widehat{\mu(\Theta)} - \mu(\Theta), Z\right) \quad \forall Z \in L_e(\mathbf{X})$

If $\widehat{a}_0$ and $\widehat{\mathbf{a}} = (\widehat{a}_1, \ldots, \widehat{a}_n)'$ are the coefficients of the credibility estimator $\widehat{\mu(\Theta)}^{cred}$, that is $\widehat{\mu(\Theta)}^{cred} = \widehat{a}_0 + \sum_{j=1}^n \widehat{a}_j X_j$ then the corresponding orthogonality conditions can be written as

$$\widehat{a}_0 = \mu_0 - \sum_{j=1}^n \widehat{a}_j \mu_{X_j} \quad \text{and} \quad \sum_{j=1}^n \widehat{a}_j \operatorname{Cov}(X_j, X_k) = \operatorname{Cov}(\mu(\Theta), X_k) \, \forall k \in [\![1, n]\!]$$

The normal equations imply that the credibility estimator depends only on the first two moments of the joint distribution of $\mu(\Theta)$ and $\mathbf{X}$. Consequently, one do not need to know or estimate the whole distribution $\mathcal{F}_\Theta$, but only the first and second order moments.
Furthermore, those equations can be applied in very general set-ups, since we have not assumed any special structure, not even a Bayes structure with a prior and conditional distributions. Instead of $\mu(\Theta)$, we could wish to predict any random variable from $\mathcal{L}^2$, based on X_j 's being any random variables from $\mathcal{L}^2$.

Let $\widehat{\mu(\Theta)}^{cred} = \widehat{a}_0 + \widehat{\mathbf{a}}'\mathbf{X}$. The corresponding normal equations can be written in matrix notation as $\widehat{\mathbf{a}}'\Sigma_{\mathbf{X}} = \mathbf{c}'$ where $\mathbf{c}' = (\mathrm{Cov}\,(\mu(\Theta), \mathbf{X}_1)\,, \ldots, \mathrm{Cov}\,(\mu(\Theta), \mathbf{X}_n))$. If the covariance matrix $\Sigma_{\mathbf{X}}$ is non-singular, then $\widehat{\mathbf{a}}' = \mathbf{c}'\Sigma_{\mathbf{X}}^{-1}$, and this expression substituted in the first condition of the normal equations gives $\widehat{a}_0 = \mu_0 - \mathbf{c}'\Sigma_{\mathbf{X}}^{-1}\boldsymbol{\mu}_{\mathbf{X}}$, from which we finally get

$$\widehat{\mu(\Theta)}^{cred} = \mu_0 + \mathbf{c}'\Sigma_{\mathbf{X}}^{-1}\left(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}\right)$$

The mathematical tools presented in this subsection enable us to find a credibility estimator either by following the computational approach, or instead, by deriving a credibility formulae through intuitive reasoning based on the general principle that the credibility estimator is a weighted average of the best individual, unbiased estimator based on the data, and the best estimator based on a priori knowledge, with weights being proportional to the corresponding precisions (inverse of the quadratic loss). Then we verify that the formulae in question fulfil the orthogonality conditions (or normal equations).

F.3 The Bühlmann-Straub model

F.3.1 Introduction and model assumptions

The Bühlmann—Straub Model developed by Bühlmann and Straub in 1970 has had a multitude of applications in life and nonlife insurance as well as in reinsurance. It is still by far the most used and the most important credibility model for insurance practice.

We place ourselves once again in the situation where, on the basis of certain risk characteristics, a given portfolio's risks have been grouped into various risk classes and we seek to calculate a risk premium for each of these classes. We thus consider I risks (or risk categories). The term risk is used to describe either an individual risk in the physical sense or a risk class. Risk i being characterized by an individual risk profile θ_i, which is itself the realization of a random variable Θ_i.

In order to estimate the risk premiums, we can first estimate the corresponding risk premium rates on the basis of standardized" aggregate claim amounts (claims in relation to volume called also "claims ratios" or "loss ratios").
We denote

- X_{ij} claims ratio (i.e. aggregate claim amount over volume, but it could be claim frequency, average claim size, etc.) of the risk $i \in [\![1, I]\!]$ in year $j \in [\![1, n]\!]$,
- w_{ij} the associated deterministic and known weight.

The weights w_{ij} are in general interpreted as volume or risk exposition measures.
For each risk i, we have to determine the corresponding individual claims ratio
$\mu(\Theta_i) = E\left[X_{ij} \mid \Theta_i\right]$.
The pure premiums are then calculated as $w_{ij} \cdot \mu(\Theta_i)$ where the volume measure is appropriately chosen for the particular line of business under consideration. Volume measures can be:

- number of years at risk in motor insurance,
- total amount of annual (or monthly) wages in the collective health or collective accident insurance,
- sum insured in fire insurance,

- annual turnover in commercial liability insurance,
- annual premium written (or earned) by the ceding company in excess of loss reinsurance.

The model assumptions are the same as Bühlmann model's, that is

- Given Θ_i, $\{X_{ij}\}_{j\in[\![1,n]\!]}$ are conditionally independent with two finite first moments
$$E\left[X_{ij} \mid \Theta_i\right] = \mu\left(\Theta_i\right) \quad \text{and} \quad \operatorname{Var}\left[X_{ij} \mid \Theta_i\right] = \frac{\sigma^2\left(\Theta_i\right)}{w_{ij}}$$
$\Longrightarrow$ The risks are independent, i.e. random variables that belong to different contracts are independent.
- The pairs $\left(\Theta_1, \mathbf{X}_1\right), \left(\Theta_2, \mathbf{X}_2\right), \ldots \left(\Theta_I, \mathbf{X}_I\right)$ are independent, and $\Theta_1, \Theta_2, \ldots \Theta_I$ are independent and identically distributed.
$\Longrightarrow$ A priori, the risks cannot be recognized as being different: they are a priori equal. This assumption is not always valid in practice and one must check if this is reasonable in concrete applications.

The number of observation years (or periods) n may also vary between risks. This can be taken into account by setting $w_{ij} = 0$ for non-observed years.

The assumption about homogeneity over time (the "true" individual claims ratio $\mu(\Theta_i)$ is constant over time), remains but is weaker than in the simple Bühlmann model, in that the variance is allowed to change with varying values of the volume measure. In practice, a constant claims ratio over time is not valid a valid assumption: the X_{ij}'s of interest can depend on inflation or on changing insurance conditions. Sometimes, one can mitigate this problem by an as-if transformation of the observed data.

Let's review the quantities of interest we will manipulate within this model:

Individual: risk class i	**Interpretation**
$\mu\left(\Theta_i\right) := E\left[X_{ij} \mid \Theta_i\right]$	individual risk premium (rate)
$\sigma^2\left(\Theta_i\right) := w_{ij} \operatorname{Var}\left[X_{ij} \mid \Theta_i\right]$	variance within individual risk

Collective: portfolio	**Interpretation**
$\mu_0 := E\left[\mu\left(\Theta_i\right)\right]$	collective premium (rate)
$\sigma^2 := E\left[\sigma^2\left(\Theta_i\right)\right]$	average variance within individual risk
$\tau^2 := \operatorname{Var}\left[\mu\left(\Theta_i\right)\right]$	variance between individual risk premiums

F.3.2 The Credibility Premium in the Bühlmann-Straub Model

- Data available: $\mathcal{D} = \{\mathbf{X}_i : i = 1, 2, \ldots, I\}$, where $\mathbf{X}_i = \left(X_{i1}, X_{i2}, \ldots, X_{in}\right)'$ is the observation vector of the i th risk.
- Goal: to estimate for each risk i its individual claims ratio $\mu(\Theta_i)$, i.e. we seek the credibility estimator based on the data $\mathcal{D}$ from all the risks in the portfolio.

The credibility estimator based on $\mathbf{X}_i$ alone, $\widehat{\mu(\Theta_i)}^{cred} = a_{i0} + \sum_j a_{ij} X_{ij}$, verify, for $k \neq i$ and all l,

$$\mathrm{Cov}\left(\widehat{\mu(\Theta_i)}^{cred}, X_{kl}\right) = \mathrm{Cov}\left(\mu\left(\Theta_i\right), X_{kl}\right) = 0$$

i.e. the normal equations seen in previous section are satisfied and $\mu(\Theta_i)^{cred}$ is therefore the credibility estimator based on all data.

The credibility premium of the risk i depends only on the individual claim experience (observations from the i th risk), and not on the claim experience of the other risks. If we know the a priori expected value μ_0 (collective premium), then the other risks cannot supply any extra information, because they are independent of the risk being rated. In practice, however, μ_0 is usually unknown and the other risks then contain information for the estimation of this a priori expected value, which justify the use of the homogeneous credibility estimator.

Before continuing, let's just adopt the notation $w_{i\bullet} = \sum_j w_{ij}$.

The best linear estimator, which is individually unbiased and which has the smallest conditional variance, is given by the weighted average $X_i = \sum_j \frac{w_{ij}}{w_{i\bullet}} X_{ij}$

$$E\left[X_i \mid \Theta_i\right] = \mu\left(\Theta_i\right) \quad \text{and} \quad \mathrm{Var}\left[X_i \mid \Theta_i\right] = \sum_j \left(\frac{w_{ij}}{w_{i\bullet}}\right)^2 \mathrm{Var}\left[X_{ij} \mid \Theta_i\right] = \frac{\sigma^2\left(\Theta_i\right)}{w_i}$$

X_i is also the homogeneous credibility estimator based on the observation vector $\mathbf{X}_i$, since $X_i = \mathrm{Pro}\left(\mu\left(\Theta_i\right) \mid L_e\left(\mathbf{X}_i\right)\right)$.

We now derive the credibility estimator based on X_i. Because of the normal equations and the fact that $E\left[X_i\right] = \mu_0$, the credibility estimator must be of the form $\widehat{\mu(\Theta_i)}^{cred} = \alpha_i X_i + (1-\alpha_i)\,\mu_0$ and satisfy

$$\mathrm{Cov}\left(\widehat{\mu(\Theta_i)}^{cred}, X_i\right) = \alpha_i \,\mathrm{Cov}\left(X_i, X_i\right) = \mathrm{Cov}(\mu\left(\Theta_i\right), X_i)$$

Since

$$\mathrm{Var}\left[X_i\right] = E\left[\mathrm{Var}\left[X_i \mid \Theta_i\right]\right] + \mathrm{Var}\left[E\left[X_i \mid \Theta_i\right]\right] = \frac{\sigma^2}{w_{i\bullet}} + \tau^2$$

$$\mathrm{Cov}\left(\mu\left(\Theta_i\right), X_i\right) = E\left[\mathrm{Cov}\left(\mu\left(\Theta_i\right), X_i \mid \Theta_i\right)\right] + \mathrm{Cov}\left(\mu\left(\Theta_i\right), E\left[X_i \mid \Theta_i\right]\right) = 0 + \mathrm{Var}\left[\mu\left(\Theta_i\right)\right] = \tau^2$$

it follows that

$$\alpha_i = \frac{\tau^2}{\frac{\sigma^2}{w_{i\bullet}} + \tau^2} = \frac{w_{i\bullet}}{w_{i\bullet} + \frac{\sigma^2}{\tau^2}}$$

We have already seen that the credibility estimator of risk i depends only on the data of risk i. Now we show that the credibility estimator based on X_i is also the credibility estimator based on $\mathbf{X}_i = (X_{i,j})_{j\in[\![1,n]\!]}$, and thus based on all data. This implies that the compressed data vector $(X_i)_{i\in[\![1,I]\!]}$ is a linear sufficient statistic of all data, i.e. the credibility estimator of risk i depends only on all the observations through X_i.

$\forall i \in [\![1, I]\!]$ and $\forall j \in [\![1, n]\!]$

$$
\begin{aligned}
\operatorname{Cov}\left(\widehat{\mu(\Theta_i)}^{cred}, X_{ij}\right) &= \alpha_i \left(\sum_k \frac{w_{ik}}{w_{i\bullet}} \operatorname{Cov}\left(X_{ik}, X_{ij}\right)\right) \\
&= \alpha_i \left(\sum_k \frac{w_{ik}}{w_{i\bullet}} \left(\frac{\sigma^2}{w_{ik}} \delta_{kj} + \tau^2\right)\right) \quad \text{where } \delta_{kj} = \begin{cases} 1 \text{ for } k = j \\ 0 \text{ otherwise.} \end{cases} \\
&= \alpha_i \left(\sum_k \frac{w_{ik}}{w_{i\bullet}} \tau^2 + \frac{\sigma^2}{w_{i\bullet}}\right) \\
&= \tau^2
\end{aligned}
$$

At last, the (inhomogeneous) credibility estimator in the Bühlmann-Straub model is given by

$$
\begin{gathered}
\widehat{\mu(\Theta_i)}^{cred} = \alpha_i X_i + (1 - \alpha_i)\,\mu_0 = \mu_0 + \alpha_i\left(X_i - \mu_0\right) \\
\text{where } X_i = \sum_j \frac{w_{ij}}{w_{i\bullet}} X_{ij}, \quad w_{i\bullet} = \sum_j w_{ij} \quad \text{and} \quad \alpha_i = \frac{w_{i\bullet}}{w_{i\bullet} + \frac{\sigma^2}{\tau^2}} = \frac{w_{i\bullet}}{w_{i\bullet} + \kappa}
\end{gathered}
$$

The quadratic losses of μ_0, X_i and $\widehat{\mu(\Theta_i)}^{cred}$ are respectively

$$
\begin{aligned}
&E\left[\left(\mu_0 - \mu\left(\Theta_i\right)\right)^2\right] = \operatorname{Var}\left(\mu\left(\Theta_i\right)\right) = \tau^2 \\
&E\left[\left(X_i - \mu\left(\Theta_i\right)\right)^2\right] = E\left\{E\left[\left(\sum_j \frac{w_{ij}}{w_{i\bullet}}\left(X_{ij} - \mu\left(\Theta_i\right)\right)\right)^2 \mid \Theta_i\right]\right\} = E\left[\sigma^2\left(\Theta_i\right)/w_{i\bullet}\right] = \sigma^2/w_{i\bullet} \\
&E\left[\left(\widehat{\mu(\Theta_i)}^{cred} - \mu\left(\Theta_i\right)\right)^2\right] = \left(1 - \alpha_i\right)\tau^2 = \alpha_i \frac{\sigma^2}{w_i}
\end{aligned}
$$

F.3.3 Interpretation of the credibility estimator

Some observations are worth being pointed out:

- $\kappa = \sigma^2/\tau^2 = \frac{(\sigma/\mu_0)^2}{(\tau/\mu_0)^2}$ is called the credibility coefficient.
- τ/μ_0 is the coefficient of variation of $\mu(\Theta_i)$, which is a good measure for the heterogeneity (the between risk variability) of the portfolio.
- The quadratic loss of the collective premium μ_0 (best estimator based only on the a priori knowledge) is reduced by the factor $1 - \alpha_i$ when using the credibility estimator instead.
- The quadratic loss of X_i (best linear and individually unbiased estimator based only on the observation vector $\mathbf{X}_i$) is reduced by the factor α_i when using the credibility estimator instead.
- The credibility estimator is a weighted mean of the two estimators where the weight assigned to each summand is proportional to its inverse quadratic loss (precision), i.e.

$$
\widehat{\mu(\Theta_i)}^{cred} = \alpha_i X_i + (1 - \alpha_i)\,\mu_0 \quad \text{where} \quad \alpha_i = \frac{\left(\sigma^2/w_{i\bullet}\right)^{-1}}{\left(\tau^2\right)^{-1} + \left(\sigma^2/w_{i\bullet}\right)^{-1}}
$$

- The general intuitive principle that the credibility estimator is a weighted average of the best individual and unbiased estimator based on the data and the best estimator based on a priori knowledge, with weights being proportional to the corresponding precisions (inverse of the quadratic loss) is fulfilled here.

- σ is the average standard deviation within risk normalized for weight 1, and μ_0 is the expected value averaged over the whole portfolio. Hence σ/μ_0 can be interpreted as some kind of an average within risk coefficient of variation and is a good measure of the within risk variability.

- The greater the weight $w_{i\bullet}$, summed over the years, the greater is α_i (notice also that w_i assumes the role of the number of observation years n in the simple Bühlmann model).

- The smaller the credibility coefficient $\kappa = \sigma^2/\tau^2$, the greater is α_i.

- $E\left[\widehat{\mu(\Theta)}^{cred}\right] = \mu_0$, i.e. on average, over the collective, the risk is correctly rated (unbiased in the collective).

- For the credibility estimator, we need, besides the observed claims ratios X_{ij} and their corresponding weights w_{ij}, the so-called structural parameters μ_0, σ^2 and τ^2. These can be determined based either on a priori knowledge, for example from the opinions of experts (pure Bayesian procedure), or they can be determined from observations of a collective of similar risks (empirical Bayes).

F.3.4 The homogeneous credibility estimator in the Bühlmann-Straub model

Since the homogeneous credibility estimator has no constant term, the overall expected value μ_0 appearing in the inhomogeneous credibility estimator must be replaced by a collectively unbiased estimator $\widehat{\mu}_0$, which is a linear function of the observations. Intuitively, we might first think of using the weighted average of the observed claims ratios

$$\bar{X} := \frac{\sum_{i,j} w_{ij} X_{ij}}{w_{\bullet\bullet}} \quad \text{where} \quad w_{\bullet\bullet} = \sum_{ij} w_{ij}$$

But, from the iterativity of projections property we have

$$\widehat{\mu(\Theta_i)}^{hom} = \text{Pro}\left(\widehat{\mu(\Theta_i)}^{cred} \mid L_e(\mathcal{D})\right) \quad \text{where} \quad \mathcal{D} = (X_{i,j})_{i\in[\![1,I]\!], j\in[\![1,n]\!]}$$

and from the normed linearity property in affine subspaces, we have

$$\widehat{\mu(\Theta_i)}^{hom} = \alpha_i \,\text{Pro}\left(X_i \mid L_e(\mathcal{D})\right) + (1-\alpha_i)\,\text{Pro}\left(\mu_0 \mid L_e(\mathcal{D})\right)$$

$X_i \in L_e(\mathcal{D})$, since $E[X_i] = E[E[X_i \mid \Theta_i]] = E[\mu(\Theta_i)] = \mu_0$, and therefore

$$\widehat{\mu(\Theta_i)}^{hom} = \alpha_i X_i + (1-\alpha_i)\,\widehat{\mu_0} \quad \text{where} \quad \widehat{\mu_0} = \text{Pro}\left(\mu_0 \mid L_e(\mathcal{D})\right)$$

X_i being the best linear, individually unbiased estimator of $\mu(\Theta_i)$ based on the individual observation vector $\mathbf{X}_i$ we could presume that X_i contains all the information in $\mathbf{X}_i$ relating to $\mu(\Theta_i)$, that we could derive from a linear estimator. It seems reasonable then to suppose that $\widehat{\mu}_0$ depends only on the data $(X_i)_{i\in[\![1,I]\!]}$. It can be proved by determining $\text{Pro}\left(\mu_0 \mid L_e(X_1,\ldots,X_I)\right)$ and then showing that $\widehat{\mu_0} = \text{Pro}\left(\mu_0 \mid L_e(\mathcal{D})\right) = \text{Pro}\left(\mu_0 \mid L_e(X_1,\ldots,X_I)\right)$.

Finally, the homogeneous credibility estimator of $\mu(\Theta_i)$ in the Bühlmann-Straub model is

$$\widehat{\mu(\Theta_i)}^{hom} = \alpha_i X_i + (1-\alpha_i)\,\widehat{\mu}_0 = \widehat{\mu}_0 + \alpha_i\left(X_i - \widehat{\mu}_0\right)$$

$$\text{where} \quad \widehat{\mu}_0 = \sum_{i=1}^{I} \frac{\alpha_i}{\alpha_\bullet} X_i, \quad \alpha_i = \frac{w_{i\bullet}}{w_{i\bullet} + \frac{\sigma^2}{\tau^2}} \quad \text{and} \quad \alpha_\bullet = \sum_{i=1}^{I} \alpha_i$$

The corresponding quadratic Loss is

$$\mathbf{E}\left[\left(\widehat{\mu(\Theta_i)}^{\text{hom}} - \mu(\Theta_i)\right)^2\right] = \tau^2 (1-\alpha_i)\left(1 + \frac{1-\alpha_i}{\alpha_\bullet}\right)$$

obtained by iteratively applying the projection operator

$$\widehat{\mu(\Theta_i)}^{hom} = \operatorname{Pro}\left(\operatorname{Pro}\left(\mu(\Theta_i) \mid L(\mathbf{X}, 1)\right) \mid L_e(\mathbf{X})\right)$$

and using the "theorem of Pythagoras"

$$E\left[\left(\widehat{\mu(\Theta_i)}^{hom} - \mu(\Theta_i)\right)^2\right] = E\left[\left(\widehat{\mu(\Theta_i)}^{cred} - \mu(\Theta_i)\right)^2\right] + E\left[\left(\widehat{\mu(\Theta_i)}^{cred} - \widehat{\mu(\Theta_i)}^{hom}\right)^2\right]$$

Let's follow with noteworthy properties of the homogeneous credibility estimator:

- Intuition fails here, as an estimator for μ_0 should not be the observed average $\bar{X} = \sum_{i=1}^{I} \frac{w_{i\bullet}}{w_\bullet} X_i$ but rather the credibility weighted average $\widehat{\mu}_0 = \sum_{i=1}^{I} \frac{\alpha_i}{\alpha_\bullet} X_i$
- It is important to note that it may happen that $\widehat{\mu(\Theta_i)}^{\text{hom}} > X_i$, even though $\bar{X} < X_i$, i.e. the homogeneous credibility estimator is not necessarily between the observed individual and the observed collective mean.
- In contrast to the (inhomogeneous) credibility estimator, we use for the homogeneous estimator the observations from the entire collective (and not only those from the i th individual): they are needed for the estimation of μ_0, automatically built in to the homogeneous estimator's formula.
- We have a balance property for the homogeneous credibility estimator in the Bühlmann-Straub model $\sum_{i,j} w_{ij} \widehat{\mu(\Theta_i)}^{\text{hom}} = \sum_{i,j} w_{ij} X_{ij}$. The equation therefore says that with respect to the past observation period and in total over the whole portfolio, the resulting total credibility premium and the aggregate claim amount are equal. This is true independent of the choice, respectively the estimates, of the structural parameters σ^2 and τ^2. Under stationary conditions, the premium level over the whole portfolio will be fair even if not all model assumptions are strictly satisfied.

F.3.5 Estimation of the structural parameters σ^2 and τ^2

We have seen that the formula for the (inhomogeneous) credibility estimator involves the three structural parameters μ_0, σ^2, τ^2. An estimate for μ_0 is already built in to the formula for the homogeneous credibility estimator, so that only the two structural parameters σ^2 and τ^2 remain to be determined. In practice, these two parameters are also unknown and must be estimated from the data of the collective.

A reasonable unbiased ($E[\widehat{\sigma}^2] = \sigma^2$) estimator for σ^2 is

$$\widehat{\sigma^2} = \frac{1}{I}\sum_{i=1}^{I}\frac{1}{n-1}\sum_{j=1}^{n} w_{ij}\left(X_{ij} - X_i\right)^2$$

An unbiased and consistent estimator for τ^2 is

$$\widehat{\tau}^2 = c \cdot \left\{ \frac{I}{I-1}\sum_{i=1}^{I}\frac{w_{i\bullet}}{w_{\bullet\bullet}}\left(X_i - \bar{X}\right)^2 - \frac{\widehat{\sigma^2}}{w_{\bullet\bullet}} \right\} \quad \text{where} \quad c = \frac{I-1}{I}\left\{\sum_{i=1}^{I}\frac{w_{i\bullet}}{w_{\bullet\bullet}}\left(1 - \frac{w_i}{w_{\bullet\bullet}}\right)\right\}^{-1}$$

$\widehat{\tau}^2$ can possibly be negative. This means that there is no detectable difference between the risks. We will then replace it by $\max\left(\widehat{\tau}^2, 0\right)$, which is no longer unbiased.

F.3.6 Empirical credibility estimator

The empirical credibility estimator is obtained from the homogeneous credibility formula by replacing the structural parameters σ^2 and τ^2 by their estimators

$$\widehat{\mu\left(\Theta_i\right)}^{emp} = \widehat{\alpha}_i X_i + \left(1 - \widehat{\alpha}_i\right)\widehat{\mu}_0, \quad \text{where} \quad \widehat{\alpha}_i = \frac{w_{i\bullet}}{w_{i\bullet} + \widehat{\kappa}}, \quad \widehat{\kappa} = \frac{\widehat{\sigma^2}}{\widehat{\tau^2}}, \quad \widehat{\mu}_0 = \frac{\sum_i \widehat{\alpha}_i X_i}{\sum_i \widehat{\alpha}_i}$$

In the case that $\widehat{\tau}^2 = 0$, we have $\widehat{\alpha}_i = 0$, and $\widehat{\mu}_0$ is thus defined by $\widehat{\mu}_0 = \sum_i \left(w_{i\bullet}/w_{\bullet\bullet}\right) X_i$

If no contract dominates (i.e. if $\sum_i \left(w_{i\bullet}/w_{\bullet\bullet}\right)^2 \to 0$ for $I \to \infty$), then we have

$$\widehat{\mu(\Theta_i)}^{emp} \xrightarrow{P} \widehat{\mu(\Theta_i)}^{cred} \quad \text{and } E\left[\widehat{\mu\left(\Theta_i\right)}^{emp}\right] \longrightarrow \mu_0 \quad \text{for } I \to \infty$$

The premium resulting from the empirical credibility estimator if applied over the past observation period and over the whole portfolio is equal to the observed total claim amount over the same period and over the whole portfolio. And the weighted average of the empirical credibility estimator taken over the whole portfolio is unbiased.

$$E\left[\sum_{i=1}^{I}\frac{w_{i\bullet}}{w_{\bullet\bullet}}\widehat{\mu}(\Theta_i)^{emp}\right] = \mu_0$$

F.4 The hierarchical credibility model

F.4.1 Structure and underlying variables

Hierarchical structures are often encountered both in statistical data analysis and in the calculation of premiums. Insurance portfolio data frequently have a hierarchical structure as is the case for AXA Global Re (cf Figure 78).

Individual risks can be classified according to some criteria and grouped together into subgroups of risks. Those subgroups can be gathered into groups, themselves composing broader groups, which together make up the total of a given line of business. Lines of business comprise a whole insurance portfolio which can be seen in turn as a particular element of a whole industry (at national scale then international and finally world wide).

Hierarchical procedures can also be used in the calculation of premiums, whereby the reasoning follows a hierarchical tree (i.e. a "top down" procedure) in that first the expected aggregate claim amount for the whole line of business is estimated and then successively distributed over the lower levels. Such a hierarchical system has the advantage of leading to a well-founded, properly balanced distribution of the burden of claims, in particular of large claims, within the collective.

We will consider in the following a hierarchical credibility model counting five levels (can easily be generalized to any order). The structure of this model can be visualized bellow.

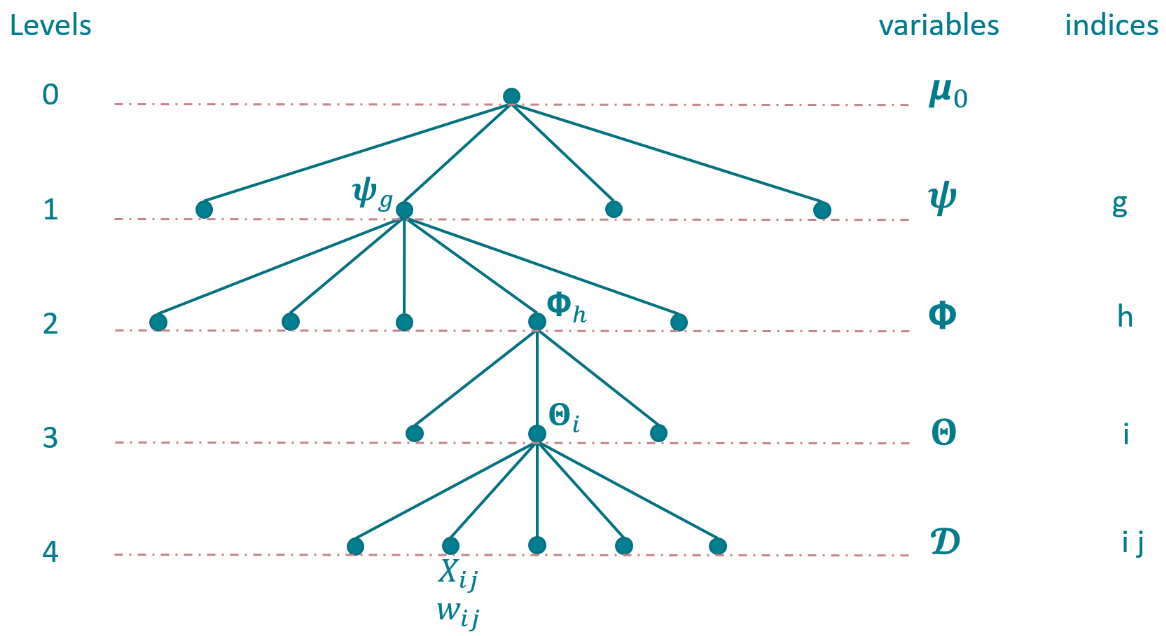


Figure 117: Tree structure of the hierarchical credibility model

Level 0 x Level 1 x Level 2 x Level 3 x Level 4 could be interpreted as line of business x risk groups x risk subgroups x individual risks x data, or as Solvency II business line x AGRe Loss ratio scope (i.e. reserving segment) x currency x individual reinsurance treaties x historical claims data.

We will denote $\Phi\left(\Psi_g\right)$ the set of Φ's that stem from Ψ_g, $\Theta\left(\Phi_h\right)$ the set of Θ's that stem from Φ_h, and $\mathcal{D}\left(\Theta_i\right)$ = the set of observations X_{ij}'s that stem from Θ_i.
More generally, using this notation, $\mathcal{D}\left(\Phi_h\right)$ = represents the set of observations X_{ij}'s issued from Φ_h, etc.

The probability structure in the hierarchical model is based on the following assumptions on the individual variables top down:

- Level 1: the random variables Ψ_g are i.i.d.
- Level 2: given Ψ_g, the random variables $\Phi_h \in \Phi\left(\Psi_g\right)$ are i.i.d. with conditional density depending only on Ψ_g.
- Level 3: given Φ_h, the random variables $\Theta_i \in \Theta\left(\Phi_h\right)$ are i.i.d. with conditional density depending only on Φ_h.
- Level 4: given Θ_i, the observations $X_{ij} \in \mathcal{D}\left(\Theta_i\right)$ are conditionally independent with conditional density depending only on Θ_i and known weights w_{ij}, and with $E\left[X_{ij} \mid \Theta_i\right] = \mu\left(\Theta_i\right)$ and $\operatorname{Var}\left[X_{ij} \mid \Theta_i\right] = \sigma^2\left(\Theta_i\right)/w_{ij}$.

The conditional densities depend only on the variables in the next higher level and not on those from levels higher than that. We say that the random variables Ψ_g, Φ_h, Θ_i, and X_{ij} possess the Markov property.

From Figure 117 we see clearly that hierarchical credibility models are generalizations of the Bühlmann-Straub model to an increased number of levels, the latter also following a tree structure, though only with 3 levels.

F.4.2 Relevant Quantities and Notation

The goal is to estimate correctly the individual quantity $\mu(\Theta_i)$ (can be a premium or a claims development factor) for each of the considered individuals (insurance or reinsurance contracts, or any other groups of risks with arbitrary aggregation). In other words we seek to find the credibility estimator $\widehat{\mu(\Theta_i)}^{cred}$ for $\mu(\Theta_i)$. For that we first define analogous quantities for the higher levels of the hierarchical tree:

$$\begin{array}{rll} \mu_0 & := E[X_{ij}], & \text{(collective premium or claims development factor)} \\ \mu(\Psi_g) & := E[X_{ij} \mid \Psi_g], & \text{where } X_{ij} \in \mathcal{D}(\Psi_g) \\ \mu(\Phi_h) & := E[X_{ij} \mid \Phi_h], & \text{where } X_{ij} \in \mathcal{D}(\Phi_h) \\ \mu(\Theta_i) & := E[X_{ij} \mid \Theta_i], & \text{where } X_{ij} \in \mathcal{D}(\Theta_i) \end{array}$$

From model assumptions (Markov property) and properties of the conditional expectation, it follows immediately that for $\Theta_i \in \Theta(\Phi_h)$, $\mu(\Phi_h) = E[X_{ij} \mid \Phi_h] = E[E[X_{ij} \mid \Theta_i, \Phi_h] \mid \Phi_h] = E[E[X_{ij} \mid \Theta_i] \mid \Phi_h] = E[\mu(\Theta_i) \mid \Phi_h]$.
The same way, we have $\mu_0 = E[\mu(\Psi_g)]$ and $\mu(\Psi_g) = E[\mu(\Phi_h) \mid \Psi_g]$ where $\Phi_h \in \Phi(\Psi_g)$.

The structural parameters of the hierarchical credibility model are the a priori expected value $\mu_0 = E[X_{ij}] = E[\mu(\Psi_g)] = E[\mu(\Phi_h)] = E[\mu(\Theta_i)]$ and the variance components:

$$\begin{array}{ll} \text{at level 1} & \tau_1^2 := \operatorname{Var}[\mu(\Psi_g)] = E\left[(\mu(\Psi_g) - \mu_0)^2\right] \\ \text{at level 2} & \tau_2^2 := E[\operatorname{Var}[\mu(\Phi_h) \mid \Psi_g]] := E[\tau_2^2(\Psi_g)] = E\left[(\mu(\Phi_h) - \mu(\Psi_g))^2\right], \\ \text{at level 3} & \tau_3^2 := E[\operatorname{Var}[\mu(\Theta_i) \mid \Phi_h]] := E[\tau_3^2(\Phi_h)] = E\left[(\mu(\Theta_i) - \mu(\Phi_h))^2\right], \\ \text{at level 4} & \sigma^2 := E[\sigma^2(\Theta_i)] = E\left[w_{ij}(X_{ij} - \mu(\Theta_i))^2\right] \end{array}$$

We further have $\operatorname{Var}[\mu(\Theta_i)] = \tau_1^2 + \tau_2^2 + \tau_3^2$ and $\operatorname{Var}[\mu(\Phi_h)] = \tau_2^2 + \tau_3^2$, the unconditional variances.

F.4.3 Credibility estimator in the hierarchical model

To find the credibility estimators $\widehat{\mu(\Theta_i)}^{cred}$ for the individual quantity $\mu(\Theta_i)$ ($\forall i \in [\![1, I]\!]$), it will be necessary first to find the credibility estimators $\widehat{\mu(\Phi_h)}^{cred}$ ($\forall h \in [\![1, H]\!]$), $\widehat{\mu(\Psi_g)}^{cred}$ ($\forall g \in [\![1, G]\!]$), and (in the homogeneous case) also $\widehat{\mu}_0$.

As we already saw in appendix subsection F.2.3, all credibility estimators can be understood as projections in the Hilbert space of all square integrable functions. In the current hierarchical set-up the credibility estimators of interest will be shown to be linear combinations of μ_0 and all raw data observations X_{ij}.

Using the iterative property of the projection operator, we get for $\Theta_i \in \Theta(\Phi_h)$:

$$\widehat{\mu(\Theta_i)}^{cred} = \operatorname{Pro}(\mu(\Theta_i) \mid L(\mathcal{D}, 1)) = \operatorname{Pro}(\operatorname{Pro}(\mu(\Theta_i) \mid L(\mathcal{D}, \mu(\Phi_h), 1)) \mid L(\mathcal{D}, 1))$$

As demonstrated in Chapter 6 of [3], Pro $(\mu(\Theta_i) \mid L(\mathcal{D}, \mu(\Phi_h), 1))$ must be of the form $\alpha_i^{(3)} B_i^{(3)} + \left(1 - \alpha_i^{(3)}\right) \mu(\Phi_h)$, where $B_i^{(3)}$ corresponds to compressed data from $\mathcal{D}(\Theta_i)$ and $\alpha_i^{(3)}$ is a suitable credibility weight equal to

$$\alpha_i^{(3)} = \frac{\tau_3^2}{\tau_3^2 + E\left[\left(\mu(\Theta_i) - B_i^{(3)}\right)^2\right]}$$

We then obtain

$$\widehat{\mu(\Theta_i)}^{cred} = \alpha_i^{(3)} B_i^{(3)} + \left(1 - \alpha_i^{(3)}\right) \widehat{\mu(\Phi_h)}^{cred}$$

Likewise, for $\Phi_h \in \Phi(\Psi_g)$:

$$\begin{aligned}
\widehat{\mu(\Phi_h)}^{cred} \quad &= \text{Pro}\left(\text{Pro}\left(\mu(\Phi_h) \mid L\left(\mathcal{D}, \mu(\Psi_g), 1\right)\right) \mid L(\mathcal{D}, 1)\right) \\
&= \text{Pro}\left(\alpha_h^{(2)} B_h^{(2)} + \left(1 - \alpha_h^{(2)}\right) \mu(\Psi_g) \mid L(\mathcal{D}, 1)\right) \\
&= \alpha_h^{(2)} B_h^{(2)} + \left(1 - \alpha_h^{(2)}\right) \widehat{\mu(\Psi_g)}^{cred}
\end{aligned}$$

where $B_h^{(2)}$ designates compressed data from $\mathcal{D}(\Phi_h)$, and

$$\alpha_h^{(2)} = \frac{\tau_2^2}{\tau_2^2 + E\left[\left(\mu(\Phi_h) - B_h^{(2)}\right)^2\right]}$$

In the same manner, $\widehat{\mu(\Psi_g)}^{cred} = \alpha_g^{(1)} B_g^{(1)} + \left(1 - \alpha_g^{(1)}\right) \mu_0$, with

$$\alpha_g^{(1)} = \frac{\tau_1^2}{\tau_1^2 + E\left[\left(\mu(\Psi_g) - B_g^{(1)}\right)^2\right]}$$

In order to calculate the credibility quantity for $\mu(\Theta_i)$ we therefore have to work out the credibility quantities for the higher levels as well as the compressed data $B_i^{(3)}, B_h^{(2)}, B_g^{(1)}$ and the corresponding credibility weights $\alpha_i^{(3)}, \alpha_h^{(2)}$ and $\alpha_g^{(1)}$.

At each level and each node, we then need to devise the best way to compress the data stemming from that node. Data compression into $B_i^{(3)}, B_h^{(2)}, B_g^{(1)}$ are once again done trough orthogonal projections on affine subspaces, producing the following "best" linear and individually unbiased estimators:

$$\begin{aligned}
B_i^{(3)} &:= \text{Pro}\left(\mu(\Theta_i) \mid L_e^{ind}(\mathcal{D}(\Theta_i))\right) \\
B_h^{(2)} &:= \text{Pro}\left(\mu(\Phi_h) \mid L_e^{ind}(\mathcal{D}(\Phi_h))\right) \\
B_g^{(1)} &:= \text{Pro}\left(\mu(\Psi_g) \mid L_e^{ind}(\mathcal{D}(\Psi_g))\right)
\end{aligned}$$

where for instance $L_e^{ind}(\mathcal{D}(\Phi_h))$ is defined as the affine subspace of the estimators linear in the data $\mathcal{D}(\Phi_h)$ and whose conditional expected value given Φ_h is equal to $\mu(\Phi_h)$, that is:

$$L_e^{ind}(\mathcal{D}(\Phi_h)) := \left\{ \widehat{\mu(\Phi_h)} : \widehat{\mu(\Phi_h)} = \sum_{\{i,j : X_{ij} \in \mathcal{D}(\Phi_h)\}} a_{ij} X_{ij}, \quad a_{ij} \in \mathbb{R}, \quad E\left[\widehat{\mu(\Phi_h)} \mid \Phi_h\right] = \mu(\Phi_h) \right\}$$

$L_e^{ind}(.)$ differs from $L_e(.)$ introduced in previous sections in the sens that $L_e^{ind}(.)$ is an affine subspace of individually unbiased estimators, whereas $L_e(.)$ is an affine subspace of collectively unbiased estimators.

It is interesting to note that for example $\alpha_h^{(2)}$ can be written

$$\alpha_h^{(2)} = \frac{\left\{E\left[\left(B_h^{(2)} - \mu\left(\Phi_h\right)\right)^2\right]\right\}^{-1}}{\left\{E\left[\left(B_h^{(2)} - \mu\left(\Phi_h\right)\right)^2\right]\right\}^{-1} + \left\{E\left[\left(\mu\left(\Phi_h\right) - \mu\left(\Psi_g\right)\right)^2\right]\right\}^{-1}}$$

which means that $\widehat{\mu(\Phi_h)}^{cred}$ is again a weighted mean with the precisions as weights.

Closed-form solutions for the numerical evaluations of the data compression and of the credibility weights at the various hierarchical levels can be uncovered (demonstration in [3]). The "best", individually unbiased estimators $B_i^{(3)}, B_h^{(2)}, B_g^{(1)}$, and the credibility weights $\alpha_i^{(3)}, \alpha_h^{(2)}$ and $\alpha_g^{(1)}$ for the various levels, can thus be calculated from bottom to top as follows:

Level 3

$$B_i^{(3)} = \sum_j \frac{w_{ij}}{w_{i\bullet}} X_{ij}, \text{ where } w_{i\bullet} = \sum_j w_{ij}$$

$$E\left[\left(\mu(\Theta_i) - B_i^{(3)}\right)^2\right] = \frac{\sigma^2}{w_{i\bullet}} \quad \text{(error term), and} \quad \alpha_i^{(3)} = \frac{w_{i\bullet}}{w_{i\bullet} + \frac{\sigma^2}{\tau_3^2}},$$

Level 2

$$B_h^{(2)} = \sum_{i \in I_h} \frac{\alpha_i^{(3)}}{w_h^{(2)}} B_i^{(3)}, \text{ where } I_h = \left\{i : \Theta_i \in \Theta\left(\Phi_h\right)\right\}, \quad w_h^{(2)} = \sum_{i \in I_h} \alpha_i^{(3)}$$

$$E\left[\left(\mu(\Phi_h) - B_h^{(2)}\right)^2\right] = \frac{\tau_3^2}{w_h^{(2)}} \quad \text{(error term), and} \quad \alpha_h^{(2)} = \frac{w_h^{(2)}}{w_h^{(2)} + \frac{\tau_3^2}{\tau_2^2}}$$

Level 1

$$B_g^{(1)} = \sum_{h \in H_g} \frac{\alpha_h^{(2)}}{w_g^{(1)}} B_h^{(2)}, \text{ where } H_g = \left\{h : \Phi_h \in \Phi\left(\Psi_g\right)\right\}, \quad w_g^{(1)} = \sum_{h \in H_g} \alpha_h^{(2)}$$

$$E\left[\left(\mu(\Psi_g) - B_g^{(1)}\right)^2\right] = \frac{\tau_2^2}{w_g^{(1)}} \quad \text{(error term), and} \quad \alpha_g^{(1)} = \frac{w_g^{(1)}}{w_g^{(1)} + \frac{\tau_2^2}{\tau_1^2}},$$

Level 0

$$\widehat{\mu_0} = \sum_g \frac{\alpha_g^{(1)}}{w^{(0)}} B_g^{(1)}, \text{ where } w^{(0)} = \sum_g \alpha_g^{(1)}$$

We can observe that the data compression estimator of the next higher level is always the weighted average (weighted with the credibility weights) of the current level (for instance $B_h^{(2)}$ is the weighted average of the $B_i^{(3)}$'s, where the weights are the $\alpha_i^{(3)}$'s).

Furthermore, the credibility weights are proportional to the "precisions" with respect to the quantities to be estimated on the next higher level. Indeed, we have

$$E\left[\left(\mu(\Phi_h) - B_i^{(3)}\right)^2\right] = E\left[\left(\mu(\Phi_h) - \mu(\Theta_i) + \mu(\Theta_i) - B_i^{(3)}\right)^2\right] = \tau_3^2 + \frac{\sigma^2}{w_{i\bullet}} = \frac{\tau_3^2}{\alpha_i^{(3)}}$$
$$E\left[\left(\mu(\Psi_g) - B_h^{(2)}\right)^2\right] = E\left[\left(\mu(\Psi_g) - \mu(\Phi_h) + \mu(\Phi_h) - B_h^{(2)}\right)^2\right] = \tau_2^2 + \frac{\tau_3^2}{w_h^{(2)}} = \frac{\tau_2^2}{\alpha_h^{(2)}}$$
$$E\left[\left(\mu_0 - B_g^{(1)}\right)^2\right] = E\left[\left(\mu_0 - \mu(\Psi_g) + \mu(\Psi_g) - B_g^{(1)}\right)^2\right] = \tau_1^2 + \frac{\tau_2^2}{w_g^{(1)}} = \frac{\tau_1^2}{\alpha_g^{(1)}}$$

From this it becomes intuitively clear that for example $B_h^{(2)}$ is the credibility weighted mean of $\left\{B_i^{(3)} : i \in I_h\right\}$.

From the above results, the (inhomogeneous) credibility estimators can be determined top down as follows:

$$\begin{aligned}
&\underline{\textbf{\textit{Level 0}}} && \mu_0 \\
&\underline{\textbf{\textit{Level 1}}} && \widehat{\mu(\Psi_g)}^{cred} = \alpha_g^{(1)} B_g^{(1)} + \left(1 - \alpha_g^{(1)}\right)\mu_0, \\
&\underline{\textbf{\textit{Level 2}}} && \widehat{\mu(\Phi_h)}^{cred} = \alpha_h^{(2)} B_h^{(2)} + \left(1 - \alpha_h^{(2)}\right)\widehat{\mu(\Psi_g)}^{cred}, \quad \Phi_h \in \Phi(\Psi_g) \\
&\underline{\textbf{\textit{Level 3}}} && \widehat{\mu(\Theta_i)}^{cred} = \alpha_i^{(3)} B_i^{(3)} + \left(1 - \alpha_i^{(3)}\right)\widehat{\mu(\Phi_h)}^{cred}, \quad \Theta_i \in \Theta(\Phi_h)
\end{aligned}$$

The credibility estimators depend on the data only through the B-values, which is why the "optimally compressed" data $B_i^{(3)}, B_h^{(2)}, B_g^{(1)}$ are called linear sufficient statistics.

To summarize, estimating credibility estimators in the hierarchical credibility model framework is a two-step procedure, where first the linear sufficient statistics (data compression) are calculated from the tree structure bottom to its top (Level 3 $\rightarrow$ Level 0), after what, the credibility estimators are determined top down (Level 0 $\rightarrow$ Level 3). This two-step procedure is exactly analogous to that for the homogeneous credibility estimator in the Bühlmann-Straub model.

Similarly, the homogeneous credibility estimator can be calculated top down as follows:

$$\begin{aligned}
&\underline{\textbf{\textit{Level 0}}} && \widehat{\mu_0} = \sum_g \frac{\alpha_g^{(1)}}{w^{(0)}} B_g^{(1)}, \text{ where } w^{(0)} = \sum_g \alpha_g^{(1)} \\
&\underline{\textbf{\textit{Level 1}}} && \widehat{\mu(\Psi_g)}^{hom} = \alpha_g^{(1)} B_g^{(1)} + \left(1 - \alpha_g^{(1)}\right)\widehat{\mu_0}, \\
&\underline{\textbf{\textit{Level 2}}} && \widehat{\mu(\Phi_h)}^{hom} = \alpha_h^{(2)} B_h^{(2)} + \left(1 - \alpha_h^{(2)}\right)\widehat{\mu(\Psi_g)}^{hom}, \quad \Phi_h \in \Phi(\Psi_g) \\
&\underline{\textbf{\textit{Level 3}}} && \widehat{\mu(\Theta_i)}^{hom} = \alpha_i^{(3)} B_i^{(3)} + \left(1 - \alpha_i^{(3)}\right)\widehat{\mu(\Phi_h)}^{hom}, \quad \Theta_i \in \Theta(\Phi_h)
\end{aligned}$$

As in the Bühlmann-Straub model, the homogeneous estimator is found by replacing μ_0 with its best linear estimator $\widehat{\mu_0}$.

Finally, the homogeneous credibility estimator in the hierarchical model yields the same balance property as in the Bühlmann-Straub model, that is

$$\sum_{i,j} w_{ij} \widehat{\mu(\Theta_i)}^{hom} = \sum_{i,j} w_{ij} X_{ij}$$

If we adopt for example a premium point of view, the homogeneous credibility estimator $\widehat{\mu(\Theta_i)}^{hom}$ (of the individual premium) defines a tariff which can be seen as a redistribution of the total observed claims.

To close this section on hierarchical credibility model, we invite the reader interested in the estimation of the structural parameters (i.e. components of variance τ_1^2, τ_2^2, τ_3^2 and σ^2) and the analysis of quadratic losses (of the credibility estimator and the homogeneous credibility estimator), to refer to Chapter 6 of [3].

G Credibilized development patterns illustrations

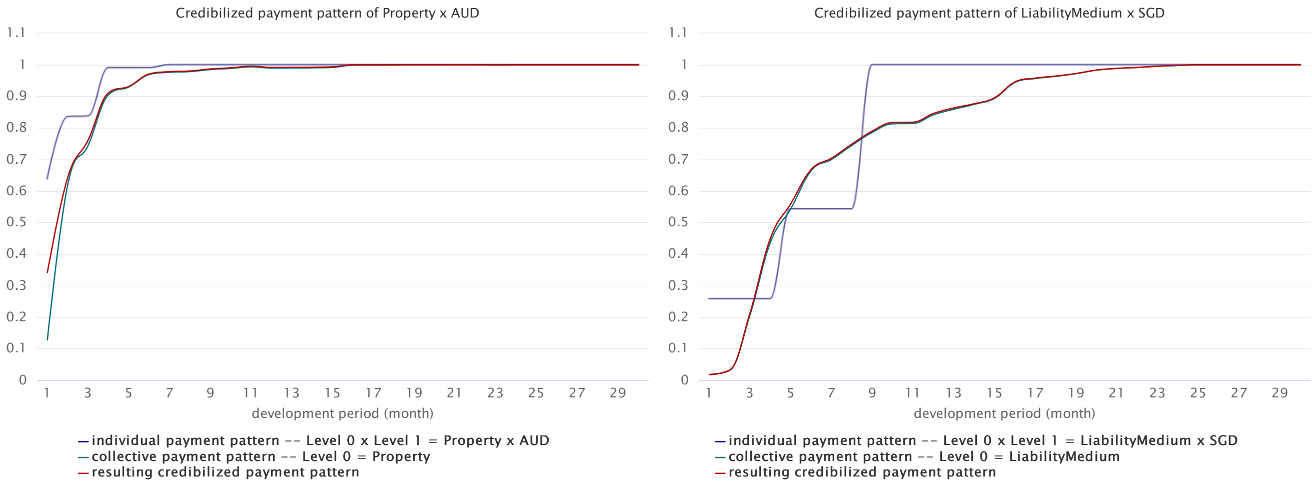


Figure 118: Developments of claims payments issued from individual, collective and credibilized development factors – Level 0 (LR SCOPE) value X Level 1 (CURRENCY) value = Property X AUD (on the left) and LiabilityMedium X SGD (on the right)

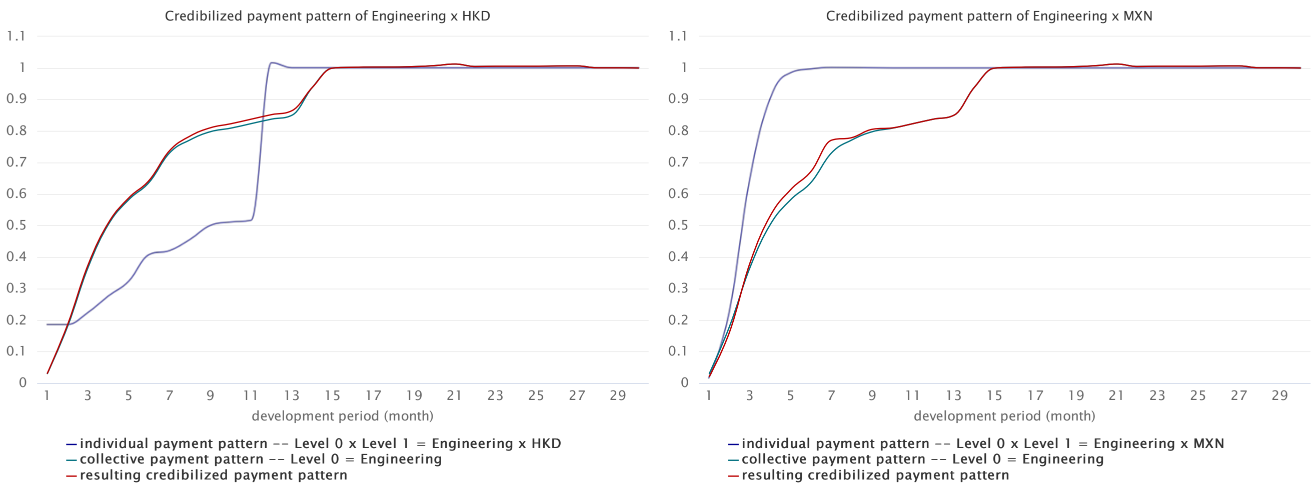


Figure 119: Developments of claims payments issued from individual, collective and credibilized development factors – Level 0 (LR SCOPE) value X Level 1 (CURRENCY) value = Engineering X HKD (on the left) and Engineering X MXN (on the right)

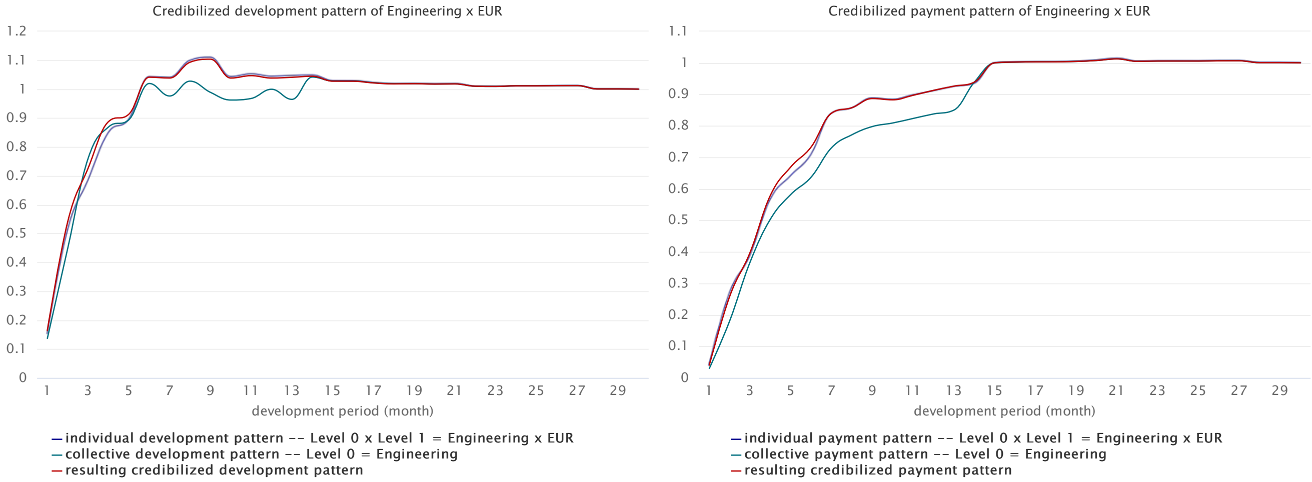


Figure 120: Developments of loss incurred (on the left) and claims payments (on the right) issued from individual, collective and credibilized development factors – Level 0 (LR SCOPE) value X Level 1 (CURRENCY) value = Engineering X USD

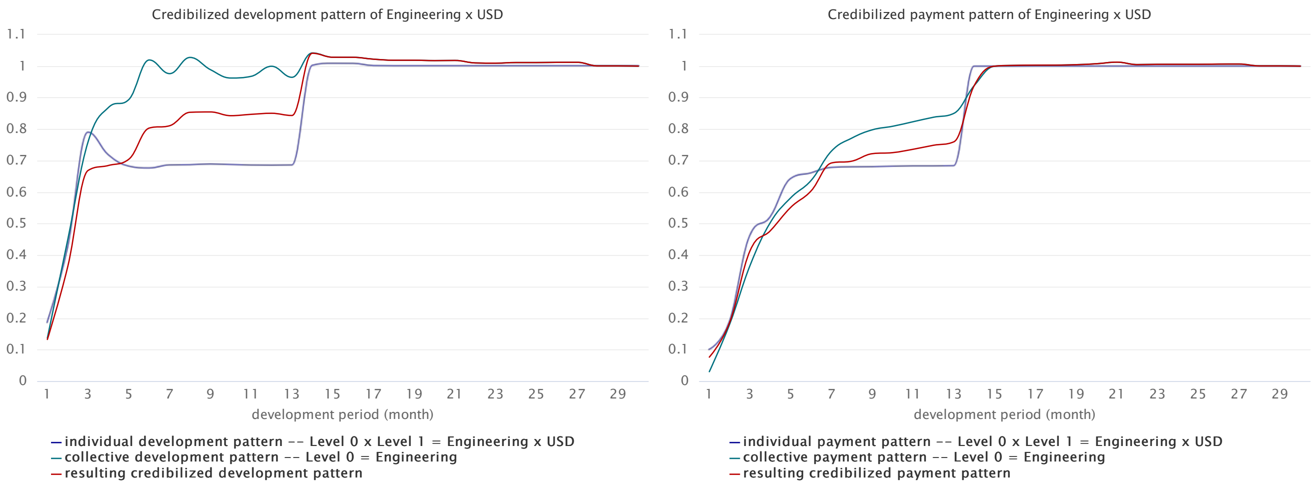


Figure 121: Developments of loss incurred (on the left) and claims payments (on the right) issued from individual, collective and credibilized development factors – Level 0 (LR SCOPE) value X Level 1 (CURRENCY) value = Engineering X USD

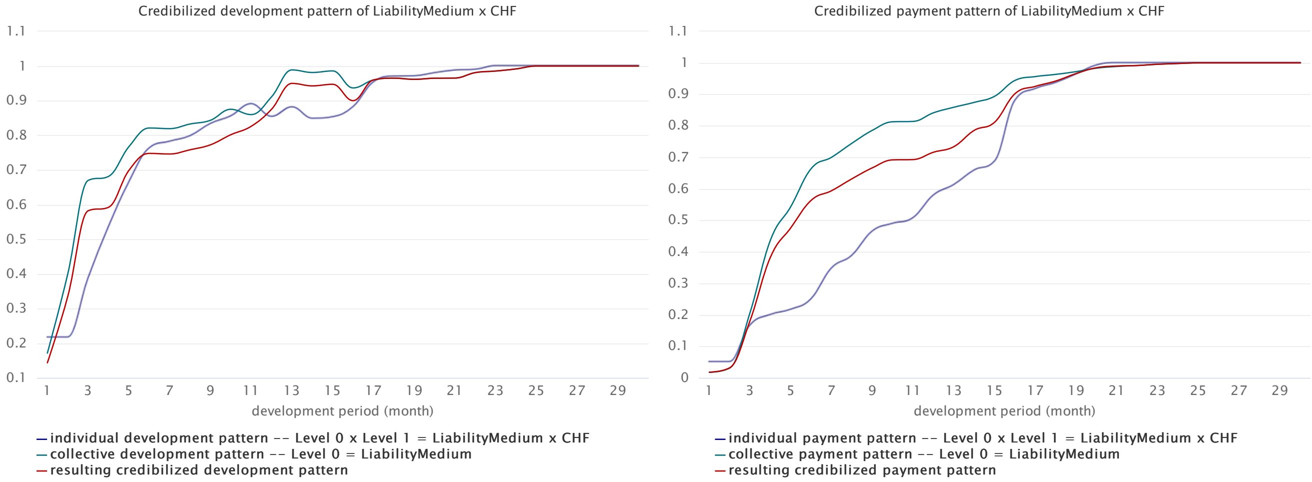


Figure 122: Developments of loss incurred (on the left) and claims payments (on the right) issued from individual, collective and credibilized development factors – Level 0 (LR SCOPE) value X Level 1 (CURRENCY) value = LiabilityMedium X CHF

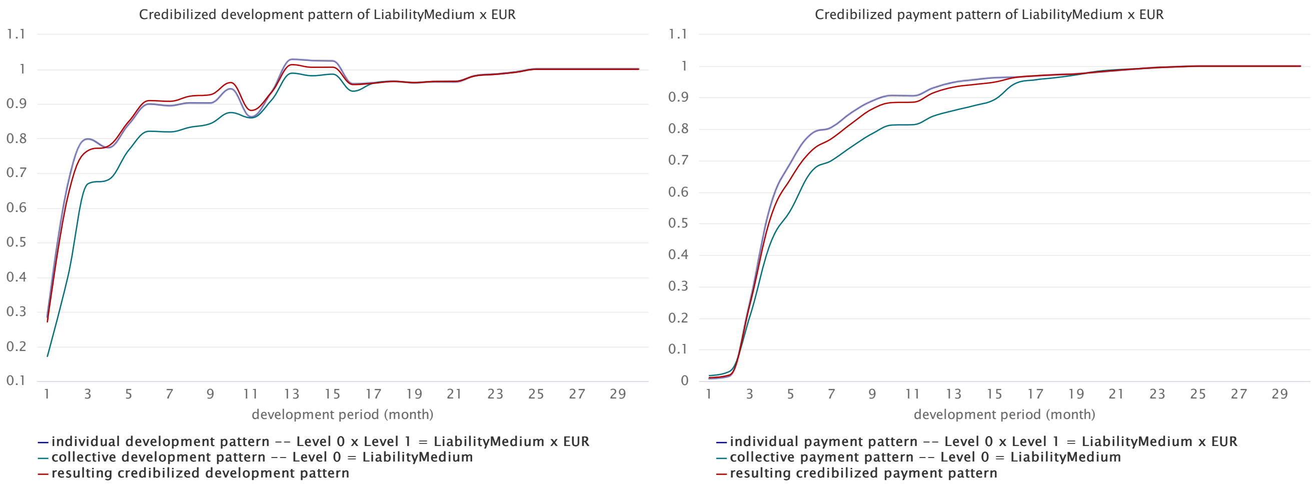


Figure 123: Developments of loss incurred (on the left) and claims payments (on the right) issued from individual, collective and credibilized development factors – Level 0 (LR SCOPE) value X Level 1 (CURRENCY) value = LiabilityMedium X EUR

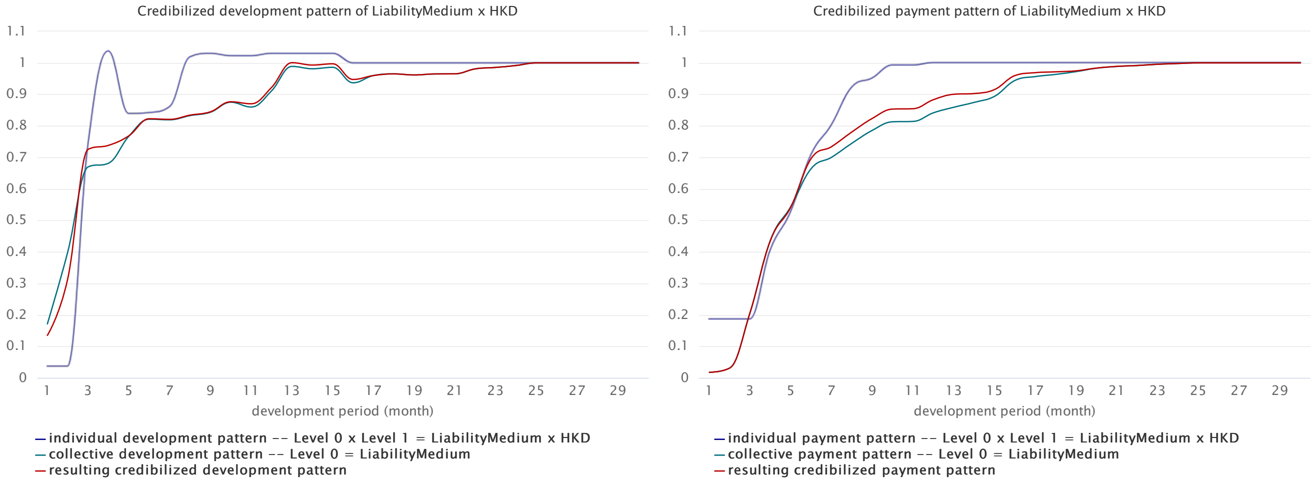


Figure 124: Developments of loss incurred (on the left) and claims payments (on the right) issued from individual, collective and credibilized development factors – Level 0 (LR SCOPE) value X Level 1 (CURRENCY) value = LiabilityMedium X HKD

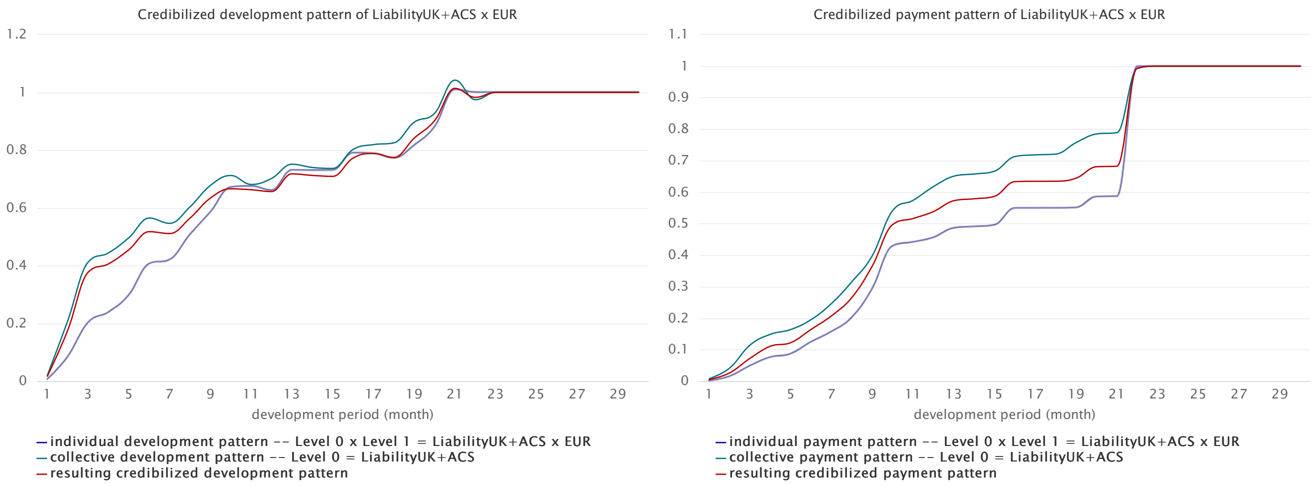


Figure 125: Developments of loss incurred (on the left) and claims payments (on the right) issued from individual, collective and credibilized development factors – Level 0 (LR SCOPE) value X Level 1 (CURRENCY) value = LiabilityUK+ACS X EUR

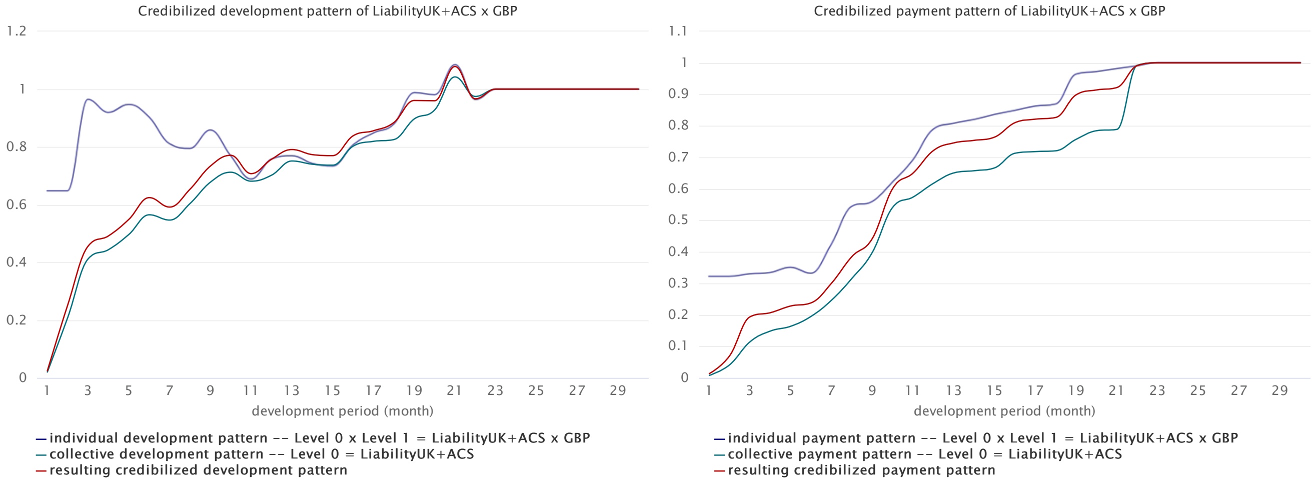


Figure 126: Developments of loss incurred (on the left) and claims payments (on the right) issued from individual, collective and credibilized development factors – Level 0 (LR SCOPE) value X Level 1 (CURRENCY) value = LiabilityUK+ACS X GBP

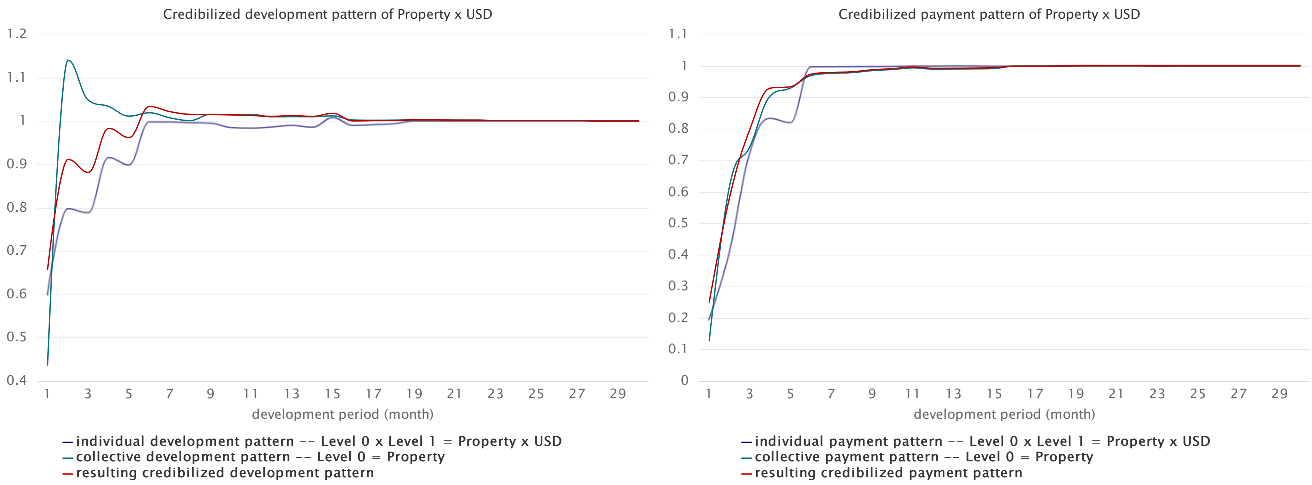
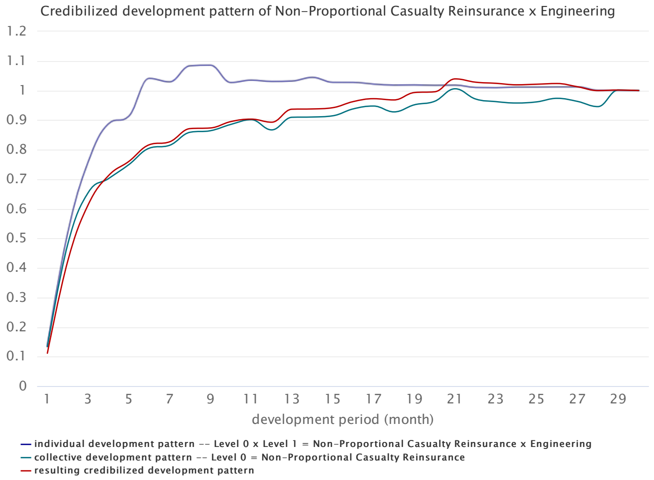


Figure 127: Developments of loss incurred (on the left) and claims payments (on the right) issued from individual, collective and credibilized development factors – Level 0 (LR SCOPE) value X Level 1 (CURRENCY) value = Property X USD



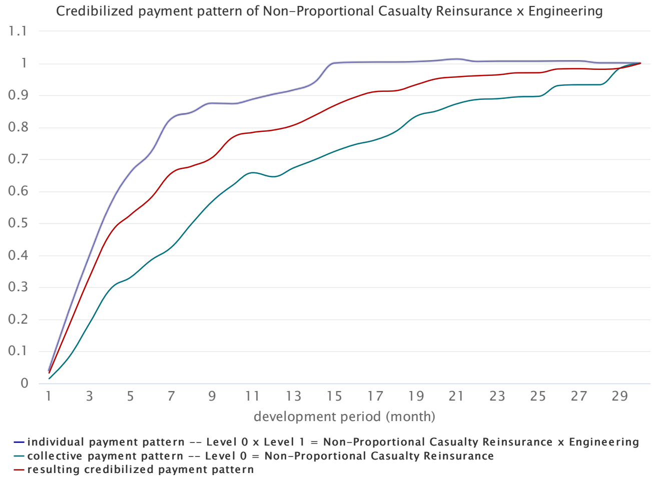
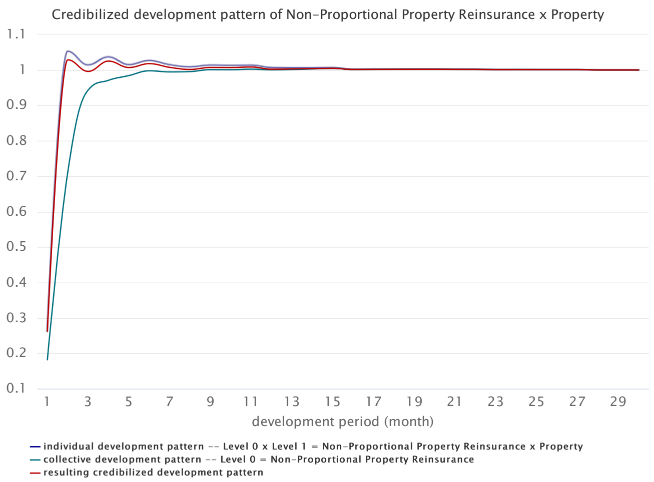


Figure 128: Developments of loss incurred (on the left) and claims payments (on the right) issued from individual, collective and credibilized development factors – Level 0 (SOLVENCY II BUSINESS LINE) value X Level 1 (LR SCOPE) value = Non-Proportional Casualty Reinsurance X Engineering



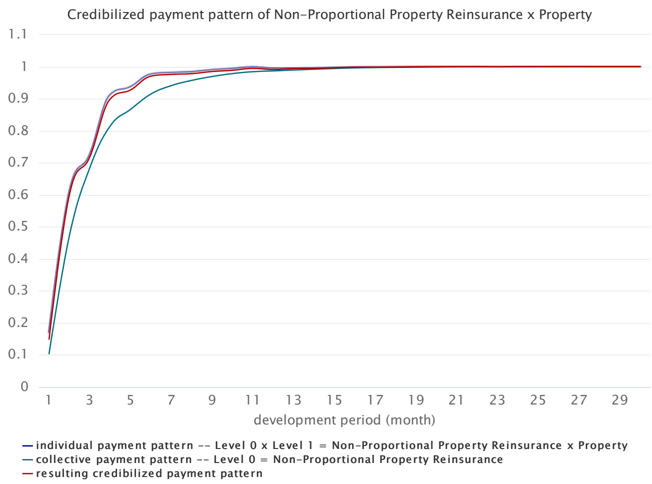


Figure 129: Developments of loss incurred (on the left) and claims payments (on the right) issued from individual, collective and credibilized development factors – Level 0 (SOLVENCY II BUSINESS LINE) value X Level 1 (LR SCOPE) value = Non-Proportional Property Reinsurance X Property

H Long-term cash flows prediction backtest results

	old_allocation__old_projection	old_allocation__proj_1	new_allocation__old_projection	new_allocation__proj_1	new_allocation__proj_2	new_allocation__proj_3	old_allocation__old_projection_cumulative	old_allocation__proj_1_cumulative	new_allocation__old_projection_cumulative	new_allocation__proj_1_cumulative	new_allocation__proj_2_cumulative	new_allocation__proj_3_cumulative
EXERCISE	**NMSE**	**NMSE**	**NMSE**	**NMSE**	**NMSE**	**NMSE**	**NMSE**	**NMSE**	**NMSE**	**NMSE**	**NMSE**	**NMSE**
2014	0.22	0.10	0.22	0.10	0.25	0.26	0.19	0.06	0.19	0.05	0.03	0.03
2015	0.22	0.15	0.22	0.16	0.32	0.34	0.23	0.08	0.23	0.07	0.03	0.04
2016	0.45	0.20	0.45	0.20	0.30	0.32	0.36	0.08	0.36	0.07	0.04	0.04
2017	0.29	0.09	0.29	0.10	0.33	0.34	0.28	0.03	0.28	0.03	0.04	0.04
2018	0.40	0.06	0.40	0.05	0.25	0.26	0.39	0.04	0.39	0.03	0.04	0.04
weighted mean	**0.29**	**0.13**	**0.29**	**0.13**	**0.29**	**0.30**	**0.27**	**0.06**	**0.27**	**0.05**	**0.04**	**0.04**
EXERCISE	**NRMSE**	**NRMSE**	**NRMSE**	**NRMSE**	**NRMSE**	**NRMSE**	**NRMSE**	**NRMSE**	**NRMSE**	**NRMSE**	**NRMSE**	**NRMSE**
2014	0.30	0.20	0.30	0.21	0.33	0.33	0.57	0.31	0.57	0.29	0.23	0.24
2015	0.31	0.26	0.31	0.26	0.37	0.38	0.69	0.40	0.69	0.38	0.26	0.28
2016	0.67	0.45	0.67	0.45	0.55	0.57	0.78	0.37	0.78	0.35	0.26	0.27
2017	0.75	0.42	0.75	0.44	0.81	0.81	0.80	0.26	0.80	0.24	0.31	0.31
2018	1.73	0.66	1.73	0.63	1.37	1.39	1.27	0.42	1.27	0.36	0.41	0.41
weighted mean	**0.59**	**0.34**	**0.59**	**0.35**	**0.56**	**0.57**	**0.74**	**0.35**	**0.74**	**0.32**	**0.27**	**0.28**
EXERCISE	**NMAE**	**NMAE**	**NMAE**	**NMAE**	**NMAE**	**NMAE**	**NMAE**	**NMAE**	**NMAE**	**NMAE**	**NMAE**	**NMAE**
2014	0.49	0.35	0.49	0.37	0.57	0.58	0.43	0.20	0.43	0.19	0.16	0.18
2015	0.52	0.44	0.52	0.45	0.60	0.62	0.47	0.24	0.47	0.23	0.16	0.18
2016	0.60	0.38	0.60	0.37	0.57	0.59	0.61	0.22	0.61	0.20	0.20	0.21
2017	0.50	0.29	0.50	0.31	0.59	0.60	0.54	0.15	0.54	0.14	0.18	0.18
2018	0.60	0.23	0.60	0.20	0.51	0.51	0.64	0.19	0.64	0.16	0.18	0.18
weighted mean	**0.53**	**0.36**	**0.53**	**0.37**	**0.57**	**0.59**	**0.51**	**0.21**	**0.51**	**0.19**	**0.17**	**0.18**
EXERCISE	**MAPE**	**MAPE**	**MAPE**	**MAPE**	**MAPE**	**MAPE**	**MAPE**	**MAPE**	**MAPE**	**MAPE**	**MAPE**	**MAPE**
2014	0.69	0.62	0.69	0.64	0.76	0.74	0.43	0.18	0.43	0.17	0.19	0.20
2015	0.68	0.73	0.68	0.74	0.68	0.69	0.46	0.21	0.46	0.21	0.19	0.20
2016	0.65	0.66	0.65	0.66	0.74	0.76	0.64	0.18	0.64	0.16	0.23	0.24
2017	0.47	0.36	0.47	0.37	0.61	0.62	0.55	0.13	0.55	0.13	0.23	0.23
2018	0.57	0.26	0.57	0.23	0.54	0.54	0.65	0.18	0.65	0.14	0.22	0.22
weighted mean	**0.64**	**0.58**	**0.64**	**0.59**	**0.69**	**0.69**	**0.52**	**0.18**	**0.52**	**0.17**	**0.21**	**0.21**

Table 31: Numerical results of reserves projection into yearly cash flows backtest – reference exercises 2014 to 2018 – all currencies (i.e. all risks) – performance criteria NMSE, NRMSE, NMAE and MAPE

EXERCISE	CURRENCY	NMSE__old__allocation__old__projection	NMSE__old__allocation__proj__1	NMSE__new__allocation__old__projection	NMSE__new__allocation__proj__1	NMSE__new__allocation__proj__2	NMSE__new__allocation__proj__3	NMSE__old__allocation__old__projection__cumulative	NMSE__old__allocation__proj__1__cumulative	NMSE__new__allocation__old__projection__cumulative	NMSE__new__allocation__proj__1__cumulative	NMSE__new__allocation__proj__2__cumulative	NMSE__new__allocation__proj__3__cumulative
2014	**EUR**	0.55	0.12	0.56	0.12	0.25	0.25	0.38	0.05	0.38	0.04	0.02	0.02
2015	**EUR**	0.52	0.18	0.53	0.17	0.22	0.24	0.44	0.09	0.44	0.09	0.02	0.03
2016	**EUR**	0.78	0.27	0.79	0.27	0.30	0.32	0.60	0.10	0.60	0.10	0.04	0.05
2017	**EUR**	0.59	0.21	0.58	0.20	0.51	0.52	0.57	0.04	0.56	0.04	0.07	0.07
2018	**EUR**	1.42	0.15	1.39	0.14	0.52	0.53	1.36	0.04	1.33	0.04	0.10	0.10
weighted mean		**0.68**	**0.18**	**0.68**	**0.18**	**0.32**	**0.33**	**0.56**	**0.07**	**0.56**	**0.07**	**0.04**	**0.04**
2014	**USD**	0.07	0.08	0.07	0.08	0.25	0.25	0.05	0.07	0.04	0.06	0.07	0.07
2015	**USD**	0.08	0.30	0.08	0.31	0.63	0.63	0.01	0.03	0.01	0.03	0.06	0.06
2016	**USD**	0.08	0.11	0.08	0.13	0.31	0.33	0.01	0.02	0.01	0.02	0.04	0.04
2017	**USD**	0.20	0.13	0.23	0.12	0.14	0.15	0.08	0.05	0.10	0.06	0.03	0.03
2018	**USD**	0.03	0.00	0.02	0.00	0.17	0.17	0.01	0.00	0.00	0.00	0.03	0.03
weighted mean		**0.09**	**0.14**	**0.09**	**0.15**	**0.33**	**0.34**	**0.03**	**0.04**	**0.03**	**0.04**	**0.05**	**0.05**
2014	**GBP**	3.56	1.74	3.06	1.20	0.48	0.43	6.38	3.10	5.15	2.00	0.44	0.40
2015	**GBP**	4.96	2.14	4.49	1.82	0.87	0.87	11.09	4.86	9.83	4.07	0.86	0.81
2016	**GBP**	5.55	2.06	5.27	1.78	0.98	0.90	10.20	3.94	9.43	3.34	0.64	0.59
2017	**GBP**	2.79	2.09	2.46	1.54	0.58	0.57	3.76	2.73	3.30	2.04	0.28	0.27
2018	**GBP**	6.56	4.75	5.96	3.63	1.16	1.12	7.11	4.97	6.45	3.85	0.90	0.87
weighted mean		**4.49**	**2.25**	**4.06**	**1.77**	**0.76**	**0.72**	**8.00**	**3.84**	**7.03**	**2.98**	**0.61**	**0.57**
2014	**CHF**	0.32	0.54	0.30	0.12	0.43	0.44	0.08	0.58	0.03	0.10	0.15	0.15
2015	**CHF**	0.40	3.34	0.23	0.53	0.38	0.37	0.31	2.13	0.11	0.40	0.42	0.41
2016	**CHF**	0.70	6.88	0.58	2.21	1.81	1.78	0.46	3.74	0.39	1.07	0.98	1.18
2017	**CHF**	0.85	7.01	0.44	1.74	1.24	1.29	0.80	5.06	0.41	1.37	0.96	0.97
2018	**CHF**	5.16	29.70	3.65	9.20	6.03	5.98	5.02	25.45	3.58	8.30	5.48	5.41
weighted mean		**0.98**	**6.40**	**0.70**	**1.79**	**1.37**	**1.37**	**0.81**	**4.76**	**0.53**	**1.38**	**1.04**	**1.07**

Table 32: Numerical results of reserves projection into yearly cash flows backtest – reference exercises 2014 to 2018 – currencies EUR, USD, GBP and CHF – performance criterion NMSE

EXERCISE	CURRENCY	NRMSE__old_allocation__old_projection	NRMSE__old_allocation__proj_1	NRMSE__new_allocation__old_projection	NRMSE__new_allocation__proj_1	NRMSE__new_allocation__proj_2	NRMSE__new_allocation__proj_3	NRMSE__old_allocation__old_projection_cumulative	NRMSE__old_allocation__proj_1_cumulative	NRMSE__new_allocation__old_projection_cumulative	NRMSE__new_allocation__proj_1_cumulative	NRMSE__new_allocation__proj_2_cumulative	NRMSE__new_allocation__proj_3_cumulative
2014	**EUR**	0.54	0.25	0.54	0.25	0.37	0.36	0.71	0.25	0.72	0.24	0.17	0.18
2015	**EUR**	0.49	0.29	0.49	0.28	0.32	0.33	0.88	0.40	0.89	0.39	0.20	0.23
2016	**EUR**	0.83	0.49	0.83	0.48	0.51	0.53	1.02	0.43	1.02	0.42	0.28	0.30
2017	**EUR**	0.87	0.52	0.87	0.51	0.81	0.82	1.24	0.34	1.24	0.34	0.43	0.43
2018	**EUR**	3.24	1.05	3.21	1.02	1.97	1.99	2.36	0.43	2.33	0.40	0.63	0.63
weighted mean		**0.91**	**0.43**	**0.90**	**0.42**	**0.61**	**0.62**	**1.06**	**0.35**	**1.06**	**0.35**	**0.29**	**0.30**
2014	**USD**	0.14	0.15	0.14	0.15	0.27	0.27	0.36	0.43	0.35	0.42	0.44	0.44
2015	**USD**	0.14	0.28	0.14	0.28	0.40	0.40	0.20	0.37	0.20	0.38	0.55	0.55
2016	**USD**	0.36	0.42	0.36	0.45	0.69	0.71	0.11	0.14	0.13	0.15	0.24	0.22
2017	**USD**	1.27	1.02	1.33	0.99	1.04	1.08	0.26	0.21	0.31	0.23	0.16	0.17
2018	**USD**	0.98	0.30	0.89	0.12	2.35	2.36	0.10	0.03	0.09	0.02	0.26	0.26
weighted mean		**0.44**	**0.38**	**0.44**	**0.37**	**0.71**	**0.72**	**0.23**	**0.28**	**0.24**	**0.29**	**0.37**	**0.37**
2014	**GBP**	1.26	0.88	1.17	0.73	0.46	0.44	1.93	1.35	1.74	1.08	0.51	0.48
2015	**GBP**	1.41	0.93	1.35	0.86	0.59	0.59	2.00	1.33	1.88	1.21	0.56	0.54
2016	**GBP**	1.81	1.11	1.77	1.03	0.76	0.73	2.24	1.40	2.16	1.28	0.56	0.54
2017	**GBP**	2.42	2.10	2.27	1.80	1.11	1.10	1.62	1.38	1.52	1.19	0.44	0.43
2018	**GBP**	17.13	14.57	16.32	12.75	7.21	7.09	3.86	3.23	3.68	2.84	1.38	1.35
weighted mean		**3.17**	**2.49**	**3.02**	**2.18**	**1.33**	**1.30**	**2.16**	**1.55**	**2.02**	**1.35**	**0.61**	**0.59**
2014	**CHF**	0.31	0.40	0.30	0.19	0.36	0.36	0.42	1.12	0.27	0.47	0.57	0.57
2015	**CHF**	0.57	1.63	0.43	0.65	0.55	0.54	0.58	1.52	0.34	0.66	0.68	0.67
2016	**CHF**	0.64	2.01	0.58	1.14	1.03	1.02	0.53	1.52	0.49	0.81	0.78	0.85
2017	**CHF**	1.07	3.09	0.77	1.54	1.30	1.32	1.49	3.77	1.07	1.96	1.64	1.65
2018	**CHF**	7.83	18.77	6.58	10.44	8.46	8.42	4.28	9.64	3.62	5.51	4.47	4.44
weighted mean		**1.31**	**3.27**	**1.09**	**1.72**	**1.49**	**1.49**	**1.03**	**2.55**	**0.79**	**1.31**	**1.19**	**1.20**

Table 33: Numerical results of reserves projection into yearly cash flows backtest – reference exercises 2014 to 2018 – currencies EUR, USD, GBP and CHF – performance criterion NRMSE

EXERCISE	CURRENCY	NMAE__old_allocation__old_projection	NMAE__old_allocation__proj_1	NMAE__new_allocation__old_projection	NMAE__new_allocation__proj_1	NMAE__new_allocation__proj_2	NMAE__new_allocation__proj_3	NMAE__old_allocation__old_projection_cumulative	NMAE__old_allocation__proj_1_cumulative	NMAE__new_allocation__old_projection_cumulative	NMAE__new_allocation__proj_1_cumulative	NMAE__new_allocation__proj_2_cumulative	NMAE__new_allocation__proj_3_cumulative
2014	**EUR**	0.65	0.37	0.65	0.37	0.52	0.52	0.63	0.18	0.63	0.17	0.13	0.13
2015	**EUR**	0.70	0.42	0.70	0.42	0.50	0.51	0.67	0.25	0.67	0.25	0.13	0.16
2016	**EUR**	0.79	0.45	0.79	0.44	0.57	0.60	0.79	0.25	0.79	0.24	0.21	0.22
2017	**EUR**	0.73	0.47	0.73	0.47	0.76	0.76	0.76	0.18	0.76	0.18	0.23	0.24
2018	**EUR**	1.09	0.31	1.07	0.31	0.74	0.74	1.19	0.17	1.18	0.17	0.23	0.24
weighted mean		**0.74**	**0.41**	**0.74**	**0.40**	**0.58**	**0.59**	**0.75**	**0.21**	**0.75**	**0.21**	**0.17**	**0.18**
2014	**USD**	0.33	0.35	0.33	0.35	0.67	0.67	0.19	0.23	0.19	0.22	0.24	0.25
2015	**USD**	0.39	0.73	0.39	0.74	1.01	1.00	0.08	0.14	0.08	0.14	0.19	0.19
2016	**USD**	0.25	0.31	0.26	0.32	0.48	0.52	0.07	0.11	0.09	0.11	0.18	0.15
2017	**USD**	0.41	0.34	0.43	0.34	0.31	0.33	0.28	0.22	0.34	0.25	0.14	0.15
2018	**USD**	0.16	0.05	0.15	0.02	0.41	0.41	0.08	0.02	0.06	0.01	0.14	0.15
weighted mean		**0.33**	**0.40**	**0.33**	**0.41**	**0.64**	**0.65**	**0.14**	**0.16**	**0.15**	**0.16**	**0.19**	**0.19**
2014	**GBP**	2.02	1.35	1.90	1.18	0.66	0.64	2.59	1.85	2.28	1.41	0.66	0.62
2015	**GBP**	2.47	1.60	2.35	1.48	1.04	1.04	3.78	2.52	3.53	2.30	1.00	0.97
2016	**GBP**	2.53	1.56	2.44	1.44	1.01	0.97	3.46	2.17	3.31	1.99	0.81	0.78
2017	**GBP**	1.63	1.41	1.53	1.23	0.49	0.49	2.14	1.84	2.00	1.58	0.51	0.50
2018	**GBP**	2.47	1.99	2.35	1.76	1.00	0.99	2.77	2.35	2.64	2.06	0.89	0.88
weighted mean		**2.22**	**1.53**	**2.11**	**1.37**	**0.83**	**0.82**	**3.01**	**2.13**	**2.79**	**1.84**	**0.78**	**0.75**
2014	**CHF**	0.64	0.72	0.56	0.37	0.71	0.71	0.27	0.76	0.17	0.28	0.35	0.35
2015	**CHF**	0.54	1.22	0.39	0.63	0.64	0.64	0.56	1.56	0.33	0.67	0.65	0.64
2016	**CHF**	0.67	1.80	0.56	1.19	1.31	1.25	0.76	2.20	0.71	1.17	1.10	1.21
2017	**CHF**	0.99	1.94	0.71	1.06	1.02	1.04	0.84	2.31	0.60	1.20	0.98	0.98
2018	**CHF**	2.30	4.34	1.93	2.55	2.08	2.06	2.21	5.22	1.88	2.97	2.41	2.40
weighted mean		**0.84**	**1.60**	**0.68**	**0.92**	**1.00**	**0.99**	**0.72**	**1.93**	**0.55**	**0.96**	**0.87**	**0.89**

Table 34: Numerical results of reserves projection into yearly cash flows backtest – reference exercises 2014 to 2018 – currencies EUR, USD, GBP and CHF – performance criterion NMAE

EXERCISE	CURRENCY	MAPE__old_allocation__old_projection	MAPE__old_allocation__proj_1	MAPE__new_allocation__old_projection	MAPE__new_allocation__proj_1	MAPE__new_allocation__proj_2	MAPE__new_allocation__proj_3	MAPE__old_allocation__old_projection_cumulative	MAPE__old_allocation__proj_1_cumulative	MAPE__new_allocation__old_projection_cumulative	MAPE__new_allocation__proj_1_cumulative	MAPE__new_allocation__proj_2_cumulative	MAPE__new_allocation__proj_3_cumulative
2014	**EUR**	0.80	0.68	0.80	0.68	0.86	0.81	0.66	0.16	0.67	0.15	0.17	0.17
2015	**EUR**	0.92	0.88	0.93	0.88	0.67	0.67	0.69	0.22	0.69	0.21	0.15	0.17
2016	**EUR**	0.92	0.94	0.92	0.93	0.89	0.91	0.84	0.21	0.84	0.20	0.23	0.24
2017	**EUR**	0.70	0.61	0.70	0.60	0.83	0.84	0.77	0.18	0.77	0.17	0.29	0.29
2018	**EUR**	1.02	0.37	1.01	0.36	0.76	0.77	1.23	0.15	1.21	0.15	0.31	0.31
weighted mean		**0.86**	**0.74**	**0.86**	**0.73**	**0.80**	**0.80**	**0.78**	**0.18**	**0.78**	**0.18**	**0.21**	**0.22**
2014	**USD**	2.39	1.78	2.39	1.74	2.42	2.43	0.19	0.21	0.18	0.21	0.25	0.25
2015	**USD**	0.75	1.29	0.76	1.30	1.77	1.74	0.09	0.15	0.09	0.15	0.22	0.22
2016	**USD**	0.29	0.31	0.31	0.33	0.48	0.53	0.07	0.15	0.09	0.16	0.25	0.23
2017	**USD**	0.43	0.35	0.46	0.36	0.29	0.31	0.41	0.32	0.48	0.34	0.14	0.15
2018	**USD**	0.16	0.05	0.15	0.02	0.41	0.41	0.08	0.03	0.08	0.01	0.22	0.22
weighted mean		**1.04**	**0.98**	**1.05**	**0.97**	**1.35**	**1.36**	**0.16**	**0.18**	**0.17**	**0.18**	**0.22**	**0.22**
2014	**GBP**	3.23	2.04	3.05	1.93	1.18	1.16	2.65	1.95	2.25	1.39	0.64	0.60
2015	**GBP**	4.14	2.73	3.89	2.51	1.21	1.20	5.04	3.46	4.58	3.10	1.14	1.12
2016	**GBP**	3.90	2.36	3.79	2.20	1.40	1.35	3.92	2.51	3.69	2.30	0.82	0.78
2017	**GBP**	2.05	1.81	1.91	1.56	0.45	0.46	2.65	2.36	2.47	2.01	0.44	0.44
2018	**GBP**	2.54	2.07	2.42	1.83	0.98	0.96	2.95	2.56	2.81	2.23	0.82	0.81
weighted mean		**3.34**	**2.25**	**3.17**	**2.06**	**1.10**	**1.08**	**3.53**	**2.56**	**3.21**	**2.18**	**0.79**	**0.76**
2014	**CHF**	0.86	0.72	0.67	0.58	0.93	0.93	0.26	0.77	0.18	0.26	0.33	0.33
2015	**CHF**	1.00	1.19	0.75	0.81	0.84	0.82	0.59	1.84	0.36	0.76	0.65	0.63
2016	**CHF**	1.36	3.92	1.21	2.34	2.29	2.27	1.61	4.99	1.50	2.71	2.32	2.42
2017	**CHF**	1.17	1.47	0.84	0.87	1.02	1.06	0.79	2.44	0.57	1.25	0.99	0.99
2018	**CHF**	2.34	3.93	1.95	2.36	1.92	1.90	2.18	5.54	1.86	3.13	2.54	2.52
weighted mean		**1.19**	**1.91**	**0.95**	**1.21**	**1.29**	**1.29**	**0.89**	**2.61**	**0.71**	**1.31**	**1.13**	**1.14**

Table 35: Numerical results of reserves projection into yearly cash flows backtest – reference exercises 2014 to 2018 – currencies EUR, USD, GBP and CHF – performance criterion MAPE

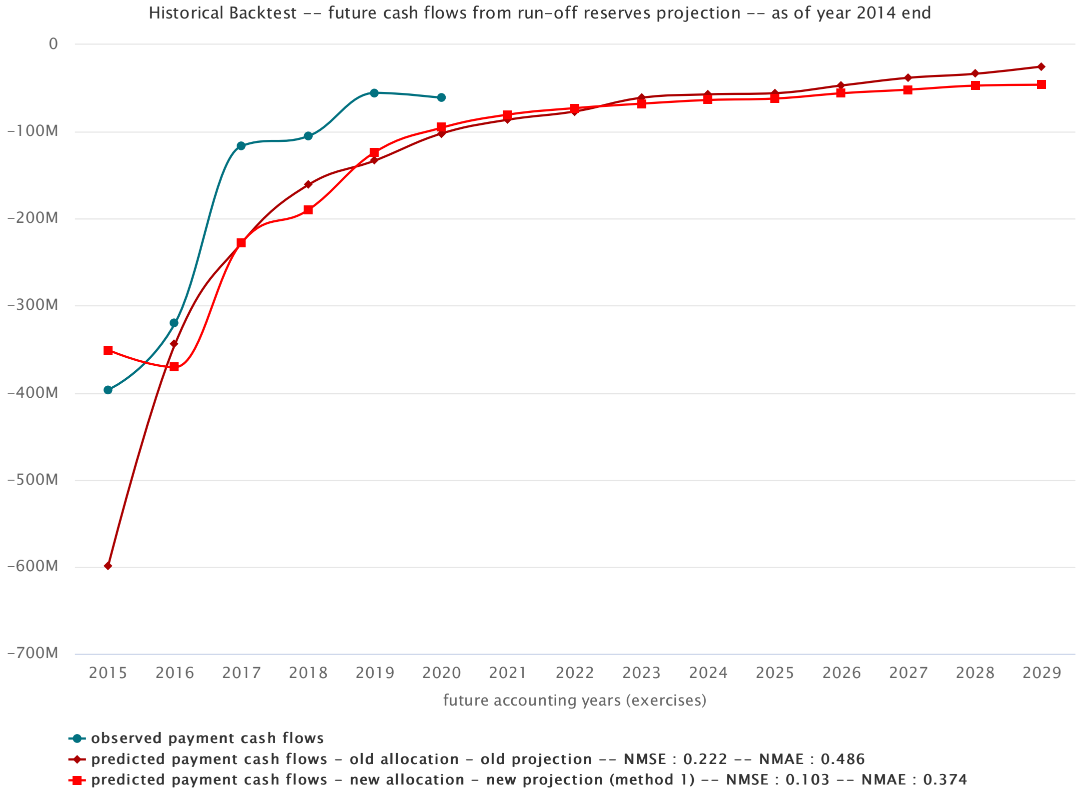


Figure 130: Reserves projections into yearly cash flows vs actually observed claims payment yearly cash flows – as of 2014 – all currencies

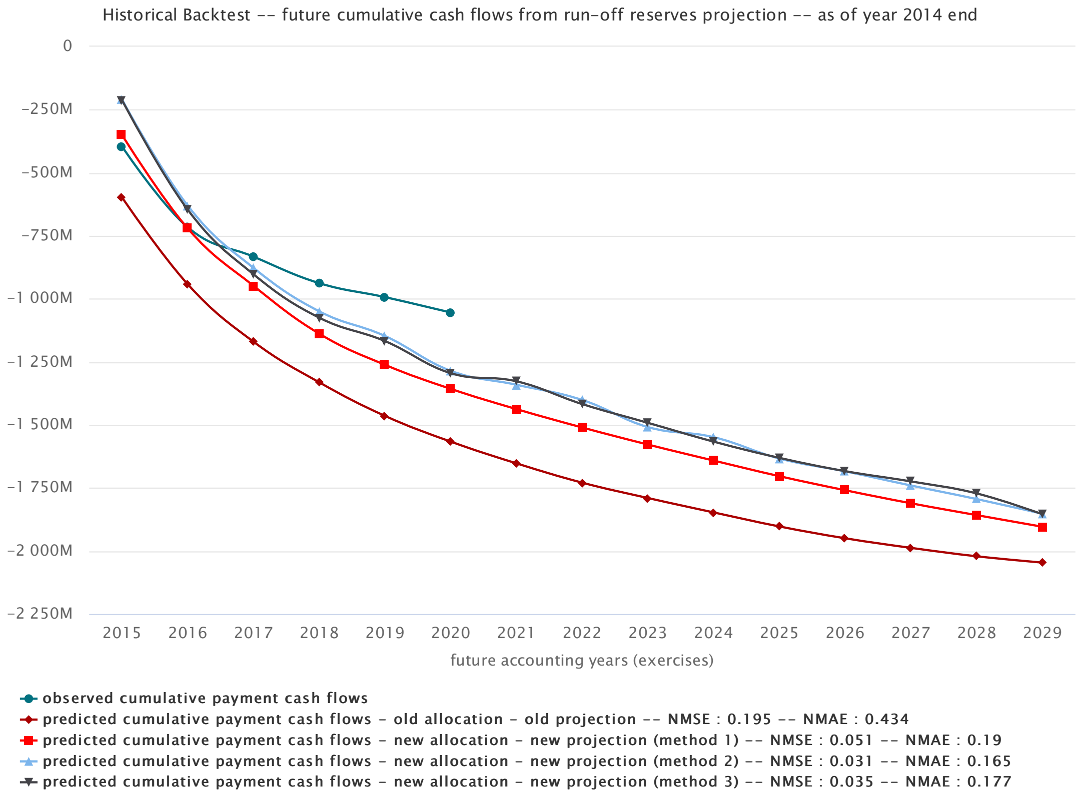


Figure 131: Reserves projections into cumulative yearly cash flows vs actually observed claims payment cumulative cash flows – as of 2014 – all currencies

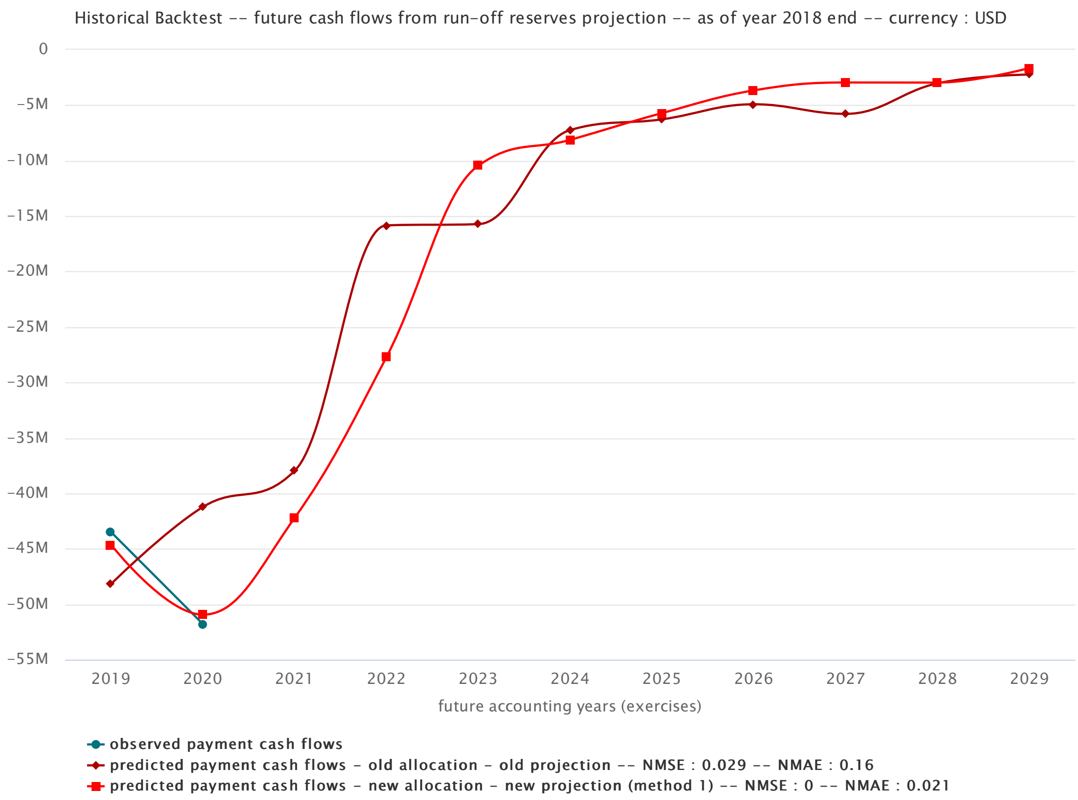


Figure 132: Reserves projections into yearly cash flows vs actually observed claims payment yearly cash flows – as of 2018 – currency USD

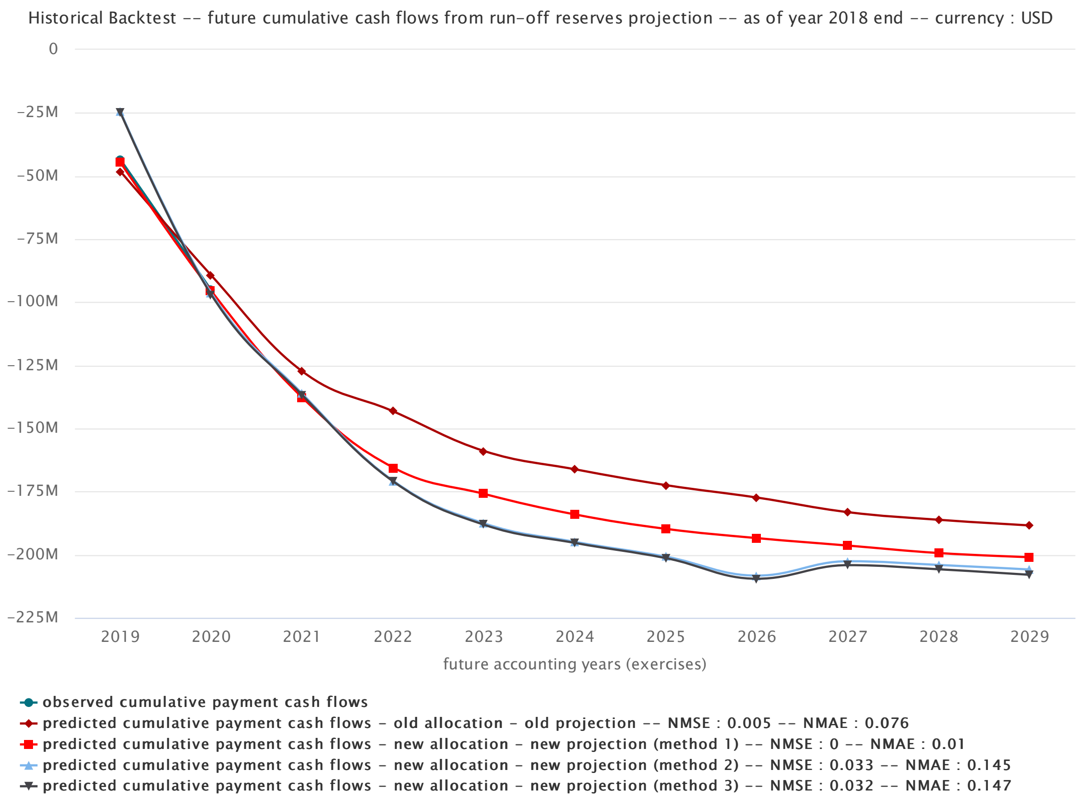


Figure 133: Reserves projections into cumulative yearly cash flows vs actually observed claims payment cumulative cash flows – as of 2018 – currency USD

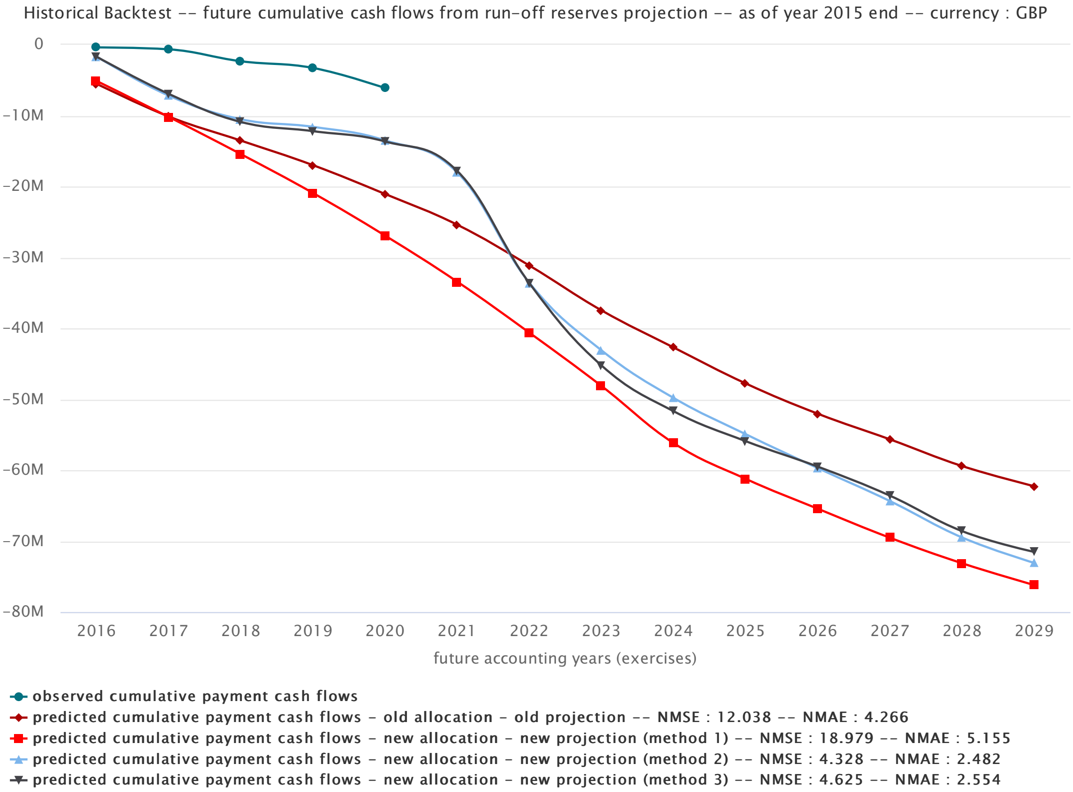


Figure 134: Reserves projections into cumulative yearly cash flows vs actually observed claims payment cumulative cash flows – as of 2015 – currency GBP – polled business