SAPIENZA
UNIVERSITÀ DI ROMA

Claim probability in Credit and Suretyship insurance

Dipartimento di Scienze Statistiche – Scuola di Scienze Statistiche
Dottorato di Ricerca in Scienze Attuariali – XXXIV Ciclo

Candidate
Jacopo Giacomelli
ID number 1839452

Thesis Advisor
Prof. Luca Passalacqua

23 December 2021

Thesis defended on 30 May 2022
in front of a Board of Examiners composed by:

Prof. Emilia Di Lorenzo (chairman, University of Naples Federico II)
Prof. Luca Regis (University of Turin)
Prof. Nicolino Ettore D'Ortona (University of Sannio)

Reviewers:

Prof. Zelda Marino (University of Naples Parthenope)
Prof. Andrea Gheno (Roma Tre University)

Claim probability in Credit and Suretyship insurance
Ph.D. thesis. Sapienza – University of Rome

This thesis has been typeset by LaTeX and the Sapthesis class.

Author's email: jacopo.giacomelli@uniroma1.it

Contents

Introduction

In the last decades, both regulatory and technological evolutions have made a vast and still growing amount of data available. Indeed, the recent Solvency 2 and Basel 3 regulatory frameworks have strongly incentivized insurance and financial industries to increase their investments in creating and maintaining high-quality databases. Furthermore, the continuous technological improvement, both on the database and the measurement instruments side, has opened new and exciting possibilities.

Therefore, recent efforts of the statistics community have been primarily devoted to developing new techniques to face the problems arising when classical data processing techniques are applied to the new "big data" context. This aroused a renewed interest in machine learning and artificial intelligence topics, both on the theoretical and the applications side. Actuarial science is no exception to this general trend, as it is developing an increasing number of big data and AI applications in pricing, provisioning, and risk management.

However, there are several real-life applications where insurance companies can hardly benefit from this new course. A relevant example is the set of all lines of business whose covered claims are catastrophic - or "cat" - events. A cat event is a claim characterized by expected low frequency and high severity, and only a few sparse observations are available. Thus, a "small data" problem must be faced when inferring probability from frequency.

An especially "unfortunate" case is the Credit and Suretyship insurance, a.k.a. C&S (*i.e.*, the Line of Business 9 in the Solvency 2 framework). This LoB deals with insolvency and unfulfilled bond events - a particular type of cat man-made risks. The similarities between LoB 9 and a part of banking activities (*e.g.*, factoring, surety) led the scientific community to avoid the effort of developing dedicated tools. However, there are also relevant differences between LoB 9 insurance companies and the banking/financial sector.
In contrast to C&S insurance companies, banks often have the possibility of implementing hedging strategies in financial markets. Furthermore, national banking supervisory authorities promote centralized databases where historical information about debtors' payment behavior is collected and shared among banks. These elements have reduced the need for improving the precision of real-world probabilities inference of insolvency/unfulfillment events in the banking sector.

The lack of models and instruments dedicated to C&S insurance applications incentivizes the use of methodologies developed for the similar needs of the banking sector. However, a liquid market for C&S insurance liabilities is missing to date, while the banks have access to financial markets, where credit derivatives are available for hedging purposes. Further, C&S insurance companies do not benefit from centralized databases that reduce the "small data" problem through shared information in the banking sector.

This work investigates the problems faced by a C&S insurance company when inferring claim probabilities and proposes a set of dedicated tools to address these problems. The topic is vast and largely unexplored to date, and hence there is no claim to completeness. However, our effort has been devoted to investigating both the estimation of a single claim probability and the dependence structure that relates the future claim events together. These two tasks are related to the two components of the LoB 9 cat risk identified by the Solvency 2 framework:

- the catastrophe *default* risk, which measures the impact of a single claim event originated by the largest risk (influenced by severity but also by marginal claim probability);
- the catastrophe *recession* risk, where a sudden increase of claim frequency across the whole portfolio is considered as the effect of a systemic crisis that can be modeled through the dependence structure of a given portfolio model.

The thesis is organized as follows.

Chapter 1 introduces the features and the context of C&S LoB. The chapter outlines a brief historical background, from the ancient origins of this class of products up to its contemporary development. The main C&S insurance products are described and compared with similar products offered by the banking sector.

Chapter 2 compares credit insurance and suretyship to discuss the appropriateness of modeling these two types of insurance business as a whole. The chapter highlights relevant similarities (*e.g.*, in both cases, claims are absorbing events) and differences (*e.g.*, only credit insurance is affected by stochastic censoring events) to describe their respective claim probabilities in a single framework. To this purpose, a selection of classic credit risk models from the mathematical finance literature is presented. Each model is briefly examined with regards to its applicability to the C&S LoB. The Standard Formula approach to model this LoB is described as well. It is considered a reference for the elements that a proper modeling approach must include to represent the C&S future claims probability.

Chapter 3 describes the classic methods to infer the probability of absorbing events, either in the case of complete information available or not. In particular, the effect of censoring events on frequency estimation is discussed, considering the seminal papers that addressed the problem in medical statistics. The limits of these studies are investigated in the case when their results have to be applied to credit insurance, where the remarkable presence of stochastic censoring events affects the claims frequency estimation.

Chapters 4−6 present the original research results obtained in this thesis.

Chapter 4 investigates some mathematical properties of the CreditRisk$^+$ model, presented in chapter 2 as a sounding candidate to describe future claims occurring in both credit insurance and suretyship in a unitary framework. In particular, the model is generalized to a multi-period framework, and the consistency of its fundative assumptions is investigated when introducing temporal autocorrelation. These results are applied to improve the model's parameters calibration when short time series are available and observed claim events are rare. Hence, the improved calibration framework enables a more punctual description of the dependency structure among risks in a C&S insurance portfolio.

Chapter 5 investigates the possible relevance of creditworthiness variations in claim probability estimation, with specific reference to the bid bonds - a typical suretyship product introduced in chapter 1. It is shown that suretyship insurance companies can completely prevent this effect by applying a proper risk appetite framework. Further, this effect is shown to be negligible in comparison with the effect of a wrong starting price choice in the public tender where the bid bond is issued. Hence, the considered CreditRisk$^+$ model, analyzed in chapter 4, is confirmed as a fitting choice for C&S applications, although not being designed to describe rating migration and other credit dynamics phenomena.

Chapter 6 addresses the problem introduced in chapter 3, about the inapplicability of the classic frequency estimators when dealing with the stochastic censoring events that the insurer experiences in credit insurance. A new estimator is introduced, developed in this work, and specifically designed to overcome these issues.

Finally, the *Concluding Remarks* chapter summarizes the results obtained and highlights the future perspectives of this work, outlining a research agenda that the candidate intends to develop in the following years.

Chapter 1

Credit and Suretyship insurance

Credit and suretyship insurance (also referred to as C&S insurance in the following) is a type of non-life insurance focused on providing protection against the breach of contracts, duties, or other obligations. Business entities purchase credit insurance to insure their accounts receivable against customers that fail to pay what they owe. On the other hand, suretyship insurance (also *surety*) has a broader range of applications, aiming to guarantee the accomplishment of a generic underlying obligation that can be related either to outstanding debt or performance duty or other objectives.

In this chapter, the main features of this line of business are introduced, with a particular focus on the elements relevant to the C&S claim probability estimation - which is the key topic investigated in this work.

The chapter is organized as follows. §1.1 presents a brief historical outline of the evolution of the C&S line of business (also LoB in the following), without claim to completeness, from its origins to the current situation. The following §§1.2 and 1.3 discuss the main products available in credit insurance and suretyship respectively. Finally, §1.4 highlights similarities and differences between C&S products and comparable products that are offered by the banking sector.

1.1 Brief history of Credit and Suretyship insurance

A brief picture of the evolution of credit insurance and surety contracts is outlined hereinafter, without claim to completeness. The early evolution of surety in ancient times and the middle ages mainly follows the dissertation available in [1], where the same topic is discussed in greater detail. On the other hand, the source of information considered about the origins of credit insurance is [2] and references therein. Figure 1.1 summarizes the contents addressed in the next subsections.

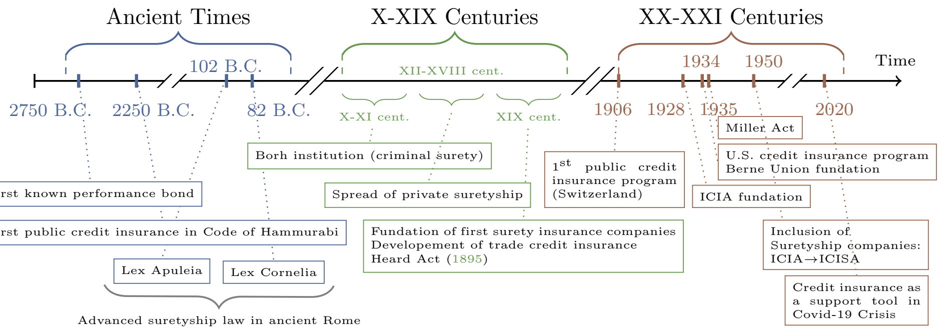

Figure 1.1. Brief summary of the contents reported in §§1.1.1–1.1.3.

1.1.1 Ancient history and the origins of credit and surety contracts

Suretyship is an ancient form of contract. The first known example of such a contract dates back to 2750 B.C., found in a tablet belonging to the library of Saigon I, king of Accad and Sumer [1, 3].

A farmer had been drafted into the military service of the king. A second farmer agreed to cultivate the soldier's farm for the period of his absence, under the contractual obligation to fertilize the land and maintain the property, to be returned to the owner in the same condition he left it. Half of the farm production during the lease was intended for the lessee in return for his services. The owner was unable to check the performance of the contract. Hence a merchant of Accad became the *surety* of the lessee: he was responsible for the lessee's behavior and had to indemnify the landowner in case the lessee would breach his obligations. The three players of a modern surety contract were established: the surety (*i.e.* the merchant), the beneficiary (*i.e.* the owner) and the contractor (*i.e.* the lessee). Also, the underlying risk of performance is the same guaranteed by one of the main modern surety products (the so-called *performance bond*, presented in paragraph 1.3) - *mutatis mutandis*.

A first example of surety law is available in the *Code of Hammurabi* (2250 B.C.) [4]. Section 32 rules the case of an official captured by the enemy and ransomed by a merchant. If the official had the means to pay his ransom, he was obliged to refund the merchant. Otherwise, a public entity (the *temple treasury* of his city or the state) was obliged to refund the merchant. It is worth noticing two features of this case: the involvement of a public entity and the presence of underlying credit risk. A public entity that provides protection against credit risk in a private transaction prefigures to some extent the current role of Export Credit Agency in long-term trade credit insurance (discussed in §1.2.2). Furthermore, the Code of Hammurabi testifies the introduction of invoices in private commercial relations[1], which are a central element in modern credit insurance contracts.

Further evolution of surety law took place in ancient Rome, as testified by Gaius in his *Commentaries* (150 A.D.). The *adpromissor* (*i.e.* the surety) was classified

[1] See sections 104-105.

in three distinct types: *sponsor*, *fidepromissor* and *fideiussor*, depending on the type of the underlying contract they could act on and the requirement of being Roman citizens or not. In the Roman world, surety was disciplined by a complex and modern legal framework. For example, *Lex Apuleia* (102 B.C.) disciplined the right of recovery among co-adpromissors[2]. In comparison, *Lex Cornelia* (81 B.C.) imposed a cap to the exposure at risk that a single adpromissor could be liable for (namely, 20.000 sesterces). Roman Law also introduced the right of the surety to attempt a recovery action against the obligor, which is a foundational element of current surety law.

1.1.2 From middle ages to the 19th century

During the middle ages, a relevant application of surety can be found in Anglo-Saxon England. In this context, surety was mainly intended as a means of enforcing criminal laws. Men were commonly required to have a *borh* (*i.e.* a surety), who was responsible for the criminal acts of his principal. This form of surety was originated from the early structure of Anglo-Saxon society, divided into clans, that needed to disciple the case of injuries provoked to a clan member by someone belonging to another clan. Two significant improvements were implemented in years 960 and 1150, respectively. The *Law of Edgar* (960) introduced the mandatory application of the institution of the *borh* to every man in the kingdom. In order to address the problem of criminals escaping from justice after committing an offense - and their respective borhs doing the same as well - the peculiar system of *frankpledge* was introduced in 1150: men were clustered by groups of ten individuals, who were surety to each other. Hence, if one of them escaped after a crime, the other nine would bring him to justice.

The application of borh was progressively extended from criminal law to commercial practices, coming back the original purpose of financial and/or contractual guarantee that surety had in ancient history. A law of Ethelred (early XI century) imposed the borh as a necessary condition to "buy or exchange". Private surety agreements continued to be stipulated for various purposes also in the centuries following the middle ages. Attestations of this are available even in literature: the plot of Shakespeare's "*Merchant of Venice*" is an example of the wide use of surety in Elisabethean England[3].

The weakness of a private suretyship agreement lies in the fact that the guarantee's creditworthiness is as solid as the standing of the person who acts as a surety. Indeed, this was well understood since the Roman law[4]. However, the need for a more reliable surety institution arose only during the 19th century under the spur of the first industrial revolution, leading to establish the first suretyship insurance companies.

[2] Namely, a sponsor or a fidepromissor who had paid more than his share, was entitled to recover the excess from his co-sponsors or co-fidepromissors.

[3] W.D. Morgan remarks [1]: *we must remember that this play was written, as were all of Shakespeare's plays, not as a literary production but primarily for presentation to Elizabethan audiences. Shakespeare, the actor, and the dramatist would have chosen as the center of his plot only a subject which would have been familiar to London's play-goers.*

[4] As discussed in paragraph 1.1.1, a cap to the single adpromissor liability was introduced by the *Lex Cornelia*.

The first proposal for establishing a surety company dates back to 1837: the American W.L. Haskins published a pamphlet titled "*Considerations on the Project and Institution of a Guarantee Company, on a New Plan, with some general views on Credit, Confidence, and Currency*" where the main features of a new type of company were outlined. The proposed company, hypothetically named *New York Guarantee Co.*, would have had the purpose to guarantee the payment of notes and other written obligations or contracts, whether of individuals, corporations, or private associations. This business was supposed to be sustainable by setting a high capitalization level since its fundation[5]. Although New York Guarantee Co. never became a reality, Haskins deserves credit for introducing the main features of a modern corporate suretyship company. A few years later, an article[6] published by professor De Morgan of London University outlined another type of surety company, proposing to apply the *principle of averages* to fidelity bonds. Namely, he proposed that a large enough group of professionals holding positions of trust[7] should organize themselves in a company. The company aimed to indemnify the losses arising from the fraudulent activities of one of its members. In return, each member had to pay a yearly premium collected in a fund to pay the company's expenses and insured losses. According to De Morgan, 500 individuals were enough to stabilize the company if the new risky professionals joining the club were selected based on their reputation.

Unlike New York Guarantee Co., the De Morgan project led to establishing a surety company, named *the Guarantee Society of London.* Fidelity insurance, which nowadays contributes to a residual share of the suretyship market, was the surety type that enabled the constitution of the first surety insurance companies in the 19th century. Indeed, in 1849 there were already five surety companies active in England. In those years, surety companies began to consider selling protection against credit and performance risk, as done by a private surety. Indeed, the 1852 prospectus of the Contract Guarantee Co. presented the intention to ensure credit risk generated by contracts[8]. Furthermore, the 1853 prospectus of the Achilles Co. expressed a similar intention referred to the performance risk[9]. Unfortunately, as in the case of New York Guarantee Co., none of these ideas was implemented. However, there are records[10] suggesting that by 1875 a suretyship market to protect public contracts

[5]According to the Haskins' project, the company's stakeholders should have paid a capital worth $ 10 millions.

[6]*Dublin Review*, 1840. Further references are available in [1].

[7]At the time when De Morgan wrote his article, such professionals were mainly clerks, secretaries, and bankers.

[8]The Contract Guarantee prospectus reads: "*The object for which this company is incorporated, is to supersede the necessity of individual security under commercial or trading contracts by providing that of an associated body.*"

[9]The Achilles prospectus reads "*The success of those companies which have been established to provide a substitute for personal guarantees for fidelity is well known, but no company at present exists securing performance of contracts.* [...] *One of the objects of this society, therefore, will be to take the place of the surety in those instances so that any contractor of known respectability and ascertained credit may be able immediately to offer to his principal an undoubted and unquestionable security for the due performance of his contract.*" Hence, this record not only implies that the idea of developing a performance bond issued by a company was arising in those days, but also proves that such a product was not available in the insurance market at that time.

[10]In [1], W.D. Morgan refers to some notes found in *The Insurance Encyclopedia* by Walford.

was available in England.

The mandatory application of surety to public works was introduced in the U.S.A. only in the last years of the century: an 1894 law, commonly known as *Heard Act* [5], required each contractor involved in public works to buy a surety policy in order to guarantee the payments due to providers and subcontractors.

In the 19th century also trade credit insurance began to develop. After the Napoleonic wars (1803-1815), the first trade credit insurance practices were introduced to guarantee safety and payments in trades [2,6]. Insurance companies offering this type of protection were established in large ports like Venice, Livorno, Naples, and Genoa. However, the essential elements of the contemporary credit insurance techniques were developed for the first time by the British Commercial Insurance Company in 1820 [2, 7].

1.1.3 From the 20th century to the contemporary framework

In the words of E.H. Cushman [8], "*the Heard Act failed to protect the United States*". Unfortunately, such a bold statement is fully justified by the lawsuits that the Heard Act generated against the United States. The act proved to be ineffective in preventing the occasional insufficiency of bonds to cover subcontractors' claims or various jurisdictional problems.

Although Congress attempted to fix those problems through the 1905 amendment, a completely reliable surety law of public contracts was available only in 1935: the so-called *Miller Act* [9], which is still effective nowadays [10]. This law requires that each contractor who enters a public works contract buys two different surety products: a *performance bond*, that guarantees to U.S. government the fulfillment of the obligation underlying the contract, and a *payment bond*, that protects subcontractors and providers against the counterparty risk generated by the contractor. Since the U.S. government is not accountable in case of contractor's default, nowadays, the standing of the surety company issuing the *payment bond* is critical to stimulate the competition among potential subcontractors. Nowadays, both U.S. and European laws require surety contracts as a means of protecting public works execution. However, there are some relevant differences between the two frameworks, furtherly discussed in section 1.3 by the comparison between Italian and U.S. *performance bond* regulations, presented in paragraphs 1.3.1 and 1.3.2 respectively.

In the early 20th century, while U.S. congress was defining the current American surety law, an increasing interest in managing commercial credit risk arose on both sides of the ocean. In fact, since the second half of the 19th century, the first mercantile credit agencies were established in the U.S.A., prefiguring the constitution of the modern rating agencies [16]. On the other hand, European countries were developing the contemporary trade credit insurance sector [2]. The first European country to introduce a credit insurance program was Switzerland in 1906 to reduce unemployment and stimulate trade. Many other European countries followed in the next three decades[11]. The worldwide economic depression after the 1929 crisis

However, to the best of our knowledge, it is not clear when and how such a market began to develop.

[11] United Kingdom (1919); Belgium (1921); Denmark (1922); Netherlands (1923); Finland (1925);

stimulated the introduction of credit insurance programs in Japan (1930) and the U.S.A. (1934). In the latter case, the first form of credit insurance practice is due to Export-Import Bank, which offered different guarantees similar to trade credit insurance.

As the number of countries active in the credit insurance sector was increasing, the international nature of this business (*i.e.* credit insurance is strongly related to exportations to foreign countries) stimulated the natural constitution of associations among credit insurers from different countries, involving both private and public entities related to the trade credit insurance industry to some extent.

The first international conference on trade credit insurance was held in London in 1926. The conference led to the foundation of the International Credit Insurance Association (ICIA), established in Paris in April 1928[12]. Since 1950 ICIA began to accept also surety insurers among its members. The name was changed to International Credit Insurance & Surety Association (ICISA) in 2001 to reflect that more than half of its members are involved in the underwriting of surety business. Nowadays, ICISA is still active, and its current members account for over 95% of the world's private credit insurance business [17]. According to Sec. 2 of the ICISA statutes [18], the purpose of this association is "*to study questions relating to Credit Insurance and Surety, to provide opportunities for Members' employees to acquire knowledge of the theory and practice of credit insurance and surety underwriting, to represent the Members' interests and to initiate means whereby the common action of the Members can be facilitated in order to develop their mutual relations in the interest of their national and the international economy, in the interest of their insured and in their own interest.*"

Another relevant association related to the C&S insurance sector is *Berne Union*, founded in 1934 and still operating. Unlike ICISA, the Berne Union - also known as the International Union of Credit and Investment Insurers - has remained focused on credit insurance across the decades. Indeed, Berne Union members are Government-backed Export Credit Agencies[13] and private credit and political risk insurers. Berne Union mission is "*to actively facilitate cross-border trade by supporting international acceptance of sound principles in export credit and foreign investment*". To this purpose, the association "*provides a forum for professional exchange, sharing of expertise and networking and coordinates collaborative projects with stakeholders from across the trade-finance industry*" [19].

Nowadays, Credit & Surety LoB has about 450 companies operating in Europe[14], that generate 1.5% of European gross written premium (GWP) volume underwritten by non-life insurance sector [21]. As shown in figure 1.2, this LoB can be considered

Germany (1926); Austria and Italy (1927); France and Spain (1928); Norway (1929); Czechoslovakia and Latvia (1931); Poland (1933); Sweden and Ireland (1935).

[12]From the ICISA website: "*The founding Members were Cobac of Belgium (now Euler Hermes), Crédito y Caución of Spain, Eidgenössische of Switzerland (now Winterthur), Hermes of Germany (now Euler Hermes), NCM of Netherlands (now Atradius), SFAC of France (now Euler Hermes), SIAC of Italy (now Euler Hermes) and Trade Indemnity of the UK (now Euler Hermes).*"

[13]Export Credit Agencies are presented in paragraph 1.2.2 together with long-term credit insurance practice.

[14]Section 2312 of EIOPA's second set of advices [20] counts 456 European companies active in C&S LoB in early 2018.

residual among the various non-life insurance types. Nevertheless, its impact on the global economy is relevant. Credit insurance guarantees almost USD three trillions of trade receivables [17] which is about 13% of world cross border trade for goods and services [19]. On the other hand, the mandatory recourse to the surety in public works - both in Europe and in U.S.A.[15] - makes this LoB have a remarkable impact on contemporary society.

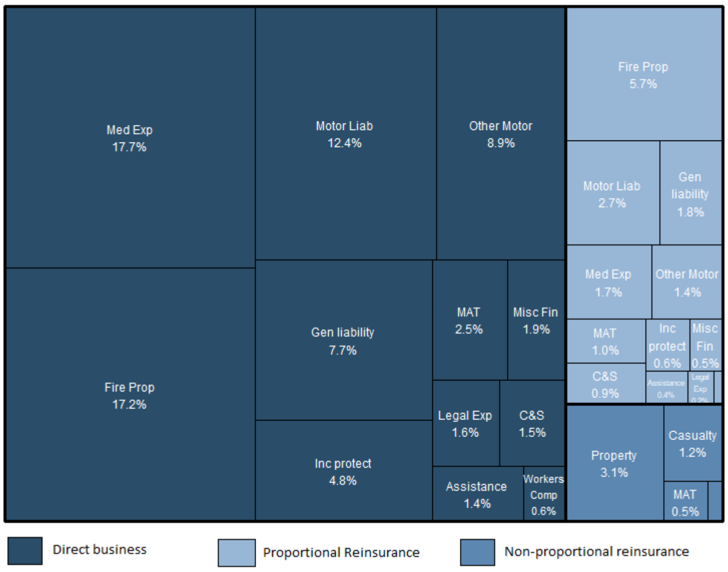

Figure 1.2. European market, year-end 2018: the total non-life insurance market split into lines of business by GWP volume. Split between direct business, proportional reinsurance and non-proportional reinsurance displayed. This picture is contained in EIOPA's *European Insurance Overview 2019* [21] (fig. 20, p. 15).

To date, the most recent example of C&S social impact is represented by the newly arisen interest of European countries in credit insurance as a means of boost the economy to recover from the worldwide 2020 Crysis, originated by the COVID-19 pandemic [34, 35].

1.2 Elements of Credit insurance

Credit insurance - also known as trade credit insurance - protects manufacturers, traders, and providers of services. The risky covered event is that a buyer of the insured subject does not pay its commercial debt or pays very late with respect to the due date. If a claim occurs, the insurer pays a pre-defined percentage of the outstanding debt, typically ranging between 75% and 95%. The percentage of coverage and other forms of risk-sharing that are discussed later in this chapter limit the moral hazard of the insured, who could be otherwise incentivized to an adverse selection of its buyers. The insurer benefits from the eventual recovery of the debt in proportion to its share of the losses. Commonly the insured and the insurer

[15] According to the Surety & Fidelity Association of America [22], surety bonds have protected more than USD 9 trillion in U.S. contracts through the period 1998-2018.

cooperate in the recovery process as a part of the accessory services offered together with the policy. The claim is not necessarily associated with the buyer's bankruptcy, although the trigger event is the violation of a financial obligation between the two parties.

Invoices are commonly provided to the insurer as a certification of the insured credit and its maturity. A copy of the contract between the insured and the buyer may be provided as well, especially in the case of complex transactions, where the two parts have formerly agreed on an amortization schedule or multiple future trades.

Credit insurance is characterized by a high degree of customization and a low degree of information available to the insurer.

The policies are Taylor-made on the insured's needs, ranging from the coverage of a single transaction to insuring the whole portfolio of buyers, including the subjects who are not buyers at the beginning of the coverage. The special clauses of policies are highly customizable as well, including both elements of risk mitigation for the insurer and possibilities of profit-sharing for the insured.

Furthermore, the complexity of this insurance business is increased because the insurer has only partial and delayed information on the covered buyers and the trades that occur between them and the insured during the policy period. Invoices are usually provided to the insurer only on established dates (e.g., quarterly or yearly) unless a claim occurs. Hence the actual exposition of the seller against a covered buyer is typically unknown to the insurer before the buyer generates a claim event or the policy expires. The following paragraph 1.2.1 describes the most common features of a credit insurance policy that is available on the regulated insurance market.

In many countries, a public credit insurance service is also available through the so-called Export Credit Agencies (ECA in the following). These public entities are not in competition with the private insurance sector. On the contrary, ECAs are intended to be a means to offer protection to enterprises where private companies are not able to. Typically ECAs deal with long-term coverages (i.e., beyond 24 months in credit insurance market) and against buyers located in *non-marketable* countries. Paragraph 1.2.2 contains the main features of the typical insurance policies issued by an ECA, compared to the ones offered by a standard credit insurance company.

1.2.1 Short-term Credit Insurance

In the following, the main features of a short-term credit insurance policy are outlined without claiming to be exhaustive in describing a class of insurance products that presents an extremely high degree of customization.

1.2.1.1 Underlying buyers management

A credit insurance policy may be formally underwritten as a master agreement, even without any underlying risk at the beginning of the policy period. Insured sellers request coverage to the insurer for each buyer they want to be protected from. The insurer may agree on each request partially or entirely, granting the insured a part of the whole credit limit needed. It is also possible for the insurer to reject any request for coverage received from the insured. While the protection is active, invoices that are issued to the covered buyer are guaranteed against the event that the buyer does not pay.

However, the insurer usually reserves the right to modify the credit limit anytime during the policy period. In particular, the credit limit on each covered buyer can be reduced or zeroed without any form of insured's consent. Any variation of the credit limit affects only invoices that have been issued afterward. Hence, zeroing the exposure generated by a buyer who has a negative outlook does not protect the insurer from existing invoices when the management action takes place.

The insurer has to take management actions on each existing credit limit on an ongoing basis during the policy period. Furthermore, the insurer's decision is requested each time the insured needs protection against a new buyer. Information needed to handle the portfolio of credit limits underlying a policy may come from three sources: previous experience of the insurer (e.g., in case the buyer has already been covered previously under another policy); information available publicly; data acquisition from info-provider. The latter source, which usually includes an external opinion/rating on the buyer's creditworthiness, has a cost that is added to the policy premium due to the insurer.

Furthermore, it is mandatory for the insured to share any adverse information regarding any covered buyer with the insurer, even if this may reduce the formerly granted credit limit. In fact, credit insurance is intended to be a partnership between the insurer and the insured in handling the risks to which the insured is exposed. Shared information and shared risk should lead to improving the average quality of the buyer's portfolio, creating value to the insured in the medium-long term. Reductions and cancellations of credit limits must be regarded as a part of this process of improvement.

The insurer's right to handle the insured risks may be limited or even nullified, depending on which special clauses have been underwritten in a given policy. The policy may establish a *grace period* between the insurer's declaration of credit limit reductions or withdrawals and their effective date. During the grace period insured can keep issuing new invoices to the buyer under the former coverage conditions. *Non-cancellable limits* are possible as well: in that case, the protection granted by the insurer against a given buyer cannot be modified during the whole policy period.

When a policy provides for grace periods or non-cancellable limits, the insurer reserves the right to receive information about the underlying buyer, although having lost - in full or in part - the right to modify the associated credit limit. Instead, in the case of a *discretionary limit*, the buyer's identity is unknown, and the credit limit is decided by the insured. This clause allows the insured to choose the covered buyers without any insurer review, provided that insurer guidelines - agreed in policy - are respected.

1.2.1.2 The claim event

When a credit limit is active for a given buyer, any invoice issued by the insured to the buyer is covered until the credit limit is reached. The risk generated by outstanding debt in excess is fully retained by the insured.

Each invoice provides for a *credit term*, at the expiry of which the invoice has to be paid. The credit term must not exceed the *maximum credit term*, which is stated in the policy; otherwise, the invoice is not guaranteed by the insurer. On the other hand, there is no inferior limit to the duration of an eligible credit term.

Once the due date is reached, a claim does not occur immediately if the debt is not paid off. The insured has the right to extend the due date until a maximum *extension period* is reached. Anyhow, the claim may be submitted, and the loss is assessed after a given *waiting period* has passed.

1.2.1.3 Risk mitigation and risk-sharing elements

After claim submission, the indemnification is quantified depending on the amount of the overdue account and the specific policy condition. In this paragraph, the most common risk-sharing conditions are introduced. Let us consider a set of losses

$$\left\{L_i^{(j_i)}\right\} \quad i = 1, \ldots, N$$

occurred to the insured during a policy period. In the notation above, $i-$th loss $L_i^{(j_i)}$ has been generated by the j_i-th buyer. Although the insurer shall likely nullify the credit limit on the $j-$th buyer, just after having received the first claim, it is possible to observe further claims on the same buyer due to invoices already issued but not expired when the first claim is communicated to the insurer. I.e., it is possible that $j = j_i = j_{i'}$ where $i \neq i'$.

In a sense, the credit limit can be regarded as a first loss-sharing tool. Indeed, if the total amount of the invoices at risk simultaneously on the same buyer exceeds the credit limit, the excess part is not covered by the policy in case the buyer becomes insolvent. Excluding this case, the first condition applied to reduce the amount of each overdue account is the *deductible*, which is commonly used in all non-life insurance businesses. The deductible amount D is the part of the loss absorbed by the insured before any indemnification under the policy. Hence each loss become

$$L_i^{(j_i),D} := \max\left\{L_i^{(j_i)} - D; 0\right\}$$

The presence of this condition is not mandatory in the policy despite being widely used. An alternative condition is AFL (the acronym for "aggregate first loss"), which has the same effect on the aggregate loss instead of acting on the single overdue account. Without loss of generality, let us suppose that the index "i" sorts the claims according to their chronological order. Hence the variable

$$\Sigma L_i := \sum_{i' \leq i} L_i^{(j_i)}$$

represents the aggregate loss generated to the insured by $i-$th claim and the claims that occurred before. AFL clause allows the insurer to consider the claims when a minimum aggregate loss threshold A has been reached. The part of the $i-$th loss that can be considered for indemnification under the policy is

$$L_i^{(j_i),A} := \min\left\{L_i^{(j_i)}; \max\left\{\Sigma L_i - A; 0\right\}\right\}$$

AFL clause is not mandatory, as in the case of the deductible clause. On the contrary, the *percentage of coverage* C_j must be included amongst the policy conditions (see e.g. [41] for Italy). C_j is commonly defined as a function of the country where the buyer's registered office is located. Sometimes C_j can also depend on other features of the buyer. The aggregate loss is reduced by the percentage of coverage as follows

$$\Sigma\tilde{L} \coloneqq \sum_{i=1}^{N} C_{j_i}\tilde{L}_i^{(j_i)}$$

where $\tilde{L}_i^{(j_i)} \in \{L_i^{(j_i)}; L_i^{(j_i),D}; L_i^{(j_i),A}\}$ depending on the possible presence of deductible clause or AFL clause in the policy. Finally, the aggregate loss is usually bounded above by the so-called *policy limit* M. M can be expressed both as an amount or as a function of the policy premium. The capacity M is reinstated at every policy renewal. The final indemnification L (before any recovery from the insolvent buyers) is

$$L \coloneqq \min\left\{\Sigma\tilde{L}; M\right\}$$

Other risk-sharing clauses may have an effect on the final quantification of the premium due to the insurer at the end of a policy annuity. The *profit-share* clause may reduce the premium that the insured has to pay for an annuity, depending on the value of L at the annuity end. Conversely, the *malus* clause may increase the premium due to the insurer in case L is greater than a given threshold.

1.2.1.4 Credit insurance product types

The types of policy available on the credit insurance market are different in the fraction of the insured's covered risks, given that every credit insurance policy can be customized with its own special clauses, regardless of the type to which it belongs.

The *whole turnover* policy offers total coverage of the insured's credit sales. This is the most suitable product for the insurer in terms of moral hazard mitigation. In fact, the insured cannot perform any adverse selection of buyers to ask for a credit limit since all of them must be reviewed by the insurer. The diagram depicted in figure 1.3 outlines the whole turnover protection mechanism.

The *single risk* policy is designed to satisfy the opposite need of an insured. In fact, it covers only the sales to one specific buyer. This product has two versions: the underlying risk can be either the wholesales to the given buyer over the policy period or just a specific deal between the insured and the buyer.

The *key debtor* policy is an intermediate product between the whole turnover and the single risk policies. It covers only a selection of the insured's buyer, up to a maximum number.

The *top-up* cover is complimentary protection that can be added to an already existing credit insurance policy. It provides an extra credit limit in excess of the one established by another insurer on the same buyer. It can also be issued by the same insurer that has accorded the first credit limit to receive an additional separated premium from the insured for the extension of coverage.

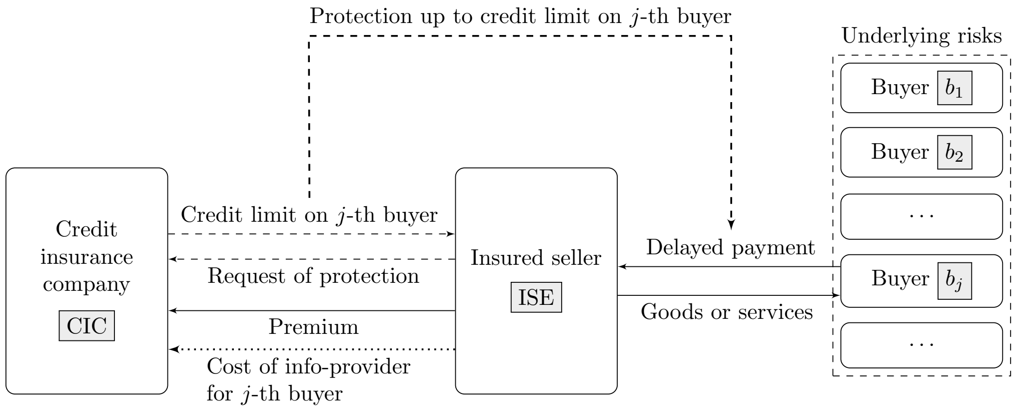

Figure 1.3. Schematics of the most typical credit insurance product: the *whole turnover* policy. The symbol b_j ($j = 1, 2, \dots$) and acronyms CIC and ISE will be used in chapter 6.

1.2.2 Long term Credit Insurance and the Export Credit Agencies

Broadly speaking, credit insurances are classified as short-term (usually one or two years), medium-term (two to five years), and long-term (over five years). This distinction originates from the agreements and understandings between countries within different frameworks. In fact, in almost all OECD countries, medium and long-term export credits are officially supported.

The support is either offered directly by the government or indirectly through a government agency or backing a private insurer through a reinsurance agreement. Institutions dealing with export credits are called Export Credit Agencies (ECAs). In case of official support, an ECA can be a government department or a commercial institution administering an account for or on behalf of the government, separate from the institution's commercial business.

In April 1978, a large number of ECAs had arranged a "Gentlemens Agreement"[16] [42] among its Participants to provide a framework for the orderly use of officially supported export credits with repayment terms of two years or more. It places limitations on the terms and conditions of export credits that benefit from official support. Such limitations include - amongst others - minimum premium rates, the minimum cash payment to be made at or before the starting point of the credit, and maximum repayment terms.

ECAs are not to be considered competitors to the private credit insurance sector. Indeed their role is actually complementary to the private sector in order to support their respective countries' export where insurance coverages are not available on the market. Although this is generally true, in rare cases, public and private credit insurances can overlap.

Apart from the typical duration of the risks and the restrictions imposed by the aforementioned international agreements, the description of a short-term credit in-

[16]This definition stems from the fact that the agreement is not an OECD Act, although it receives administrative support of the OECD Secretariat.

surance policy, available in paragraph 1.2.1, applies to the policies issued by an ECA as well.

1.3 Elements of Surety insurance

Surety policies provide a guarantee of performance and principles of various objectives and duties. They are commonly required to secure the obligations of the principal debtor (generally known as the *principal*) against the beneficiary. Unlike in credit insurance, in this case, the insurer directly relates to the source of risk. In fact, the principal usually underwrites a surety policy because this is a requirement to engage the beneficiary in business.

Risks underlying surety policies can be very diverse from each other, ranging from performance risk in an engineering contract to moral hazard/operational risks in claiming a VAT credit to be refunded. Surety bonds fall under two categories [43]:

- *contract bonds*, intended to guarantee the performance of contractual obligations, mainly in the areas of public works and private construction projects;
- *commercial bonds*, intended to secure the performance of legal or regulatory obligations.

Without claim to completness, examples of products belonging to each category [3, 43, 44] are listed in table 1.1 .

Contract bonds	
bid bond	guarantees that a contractor has submitted a bid in good faith and intends to enter the contract in case of award
performance bond	offers protection from the case that a contractor fails to fulfill the terms of the contract
advance payment bond	guarantees that the contractor will be able to repay the procuring entity any funds received in advance
payment bond	protects the credit of workers, subcontractors and suppliers against the contractor
maintenance bond	guarantees against defective workmanship or materials
Commercial bonds	
costums bond	assures customs authorities that an importer will pay the import duties required
tax bond	ensures the proper declaration and timely payment of taxes
licence/permit bond	guarantees the obligor's compliance with laws
court/fidelity bond	guarantees the performance of fiduciaries' duties and their compliance with court orders

Table 1.1. A brief description of the main suretyship products.

Let us focus on two typical products in surety insurance among the ones listed above: the *bid bond* and the *performance bond*. The life cycle of these two products - depicted in figures 1.4 and 1.5 - can be summarized as follows.

A procuring entity requires a generic "performance", such as the construction of an infrastructure or the supply of specific goods or services. Hence, the aforementioned entity uses a bidding process in order to select the best contractor for the assignment. Each contractor interested in submitting the bid has to underwrite a *bid bond* that guarantees the procuring entity against the case that the awarded contractor is not able to take charge of the required performance. Indeed, the contractor could go bankrupt during the bidding process, or some requirement necessary to fulfill the obligation could not be met (*e.g.* legal authorizations needed to perform the underlying task). In this case, the bidding process has to be reopened, and the insurer indemnifies the procuring entity.

In case the winning contractor satisfies all the other requirements, a *performance bond* is still needed to close the bidding process. It is worth noticing that the insurer who has issued the bid bond may refuse to underwrite the performance bond. However, if the contractor cannot find another insurer available to underwrite the required performance bond, the tender is reopened, and thus the bid bond issuer has to indemnify the procuring entity. This mechanism implies that insurers who issue a bid bond are also sharing the subsequent performance risk with the procuring entity to some extent.

The phase following the conclusion of the bidding process is the *execution phase* when the obligation has to be fulfilled by the winner of the tender. The performance bond guarantees the beneficiary against the risk that the principal is not able to satisfy the timing or any other requirement of the obligation. In case the performance does not meet all the requirements declared in the bidding process, the insurer indemnifies the procuring entity.

Depending on the considered regulatory framework, the procuring entity has the right to increase the duration or to vary other features of the obligation during the execution phase. This is usually the case when the procuring entity is a public institution. When the risk underlying the performance bond is modified, the insurer may require the payment of a premium supplement but must accept to guarantee the beneficiary against the risk extension.

In case of claim, the *subrogation*[17] takes place. Namely, the subrogation can be thought of as the set of rules that defines the role of the insurer when the execution phase is interrupted by a violation of the underlying obligation. Different regulatory frameworks decline this concept in different ways, as shown below by the comparison between the Italian and the American bidding laws, described in the next paragraphs 1.3.1 and 1.3.2.

[17]From late Latin "*subrogare*" (*i.e.* to choose as substitute), "*subrogation*" is a juridic expression specifically referred to the surety context. It means that the "surety" (*i.e.* the insurer) has the right (and/or the oblige) to act as the beneficiary and/or the principal when a claim occurs.

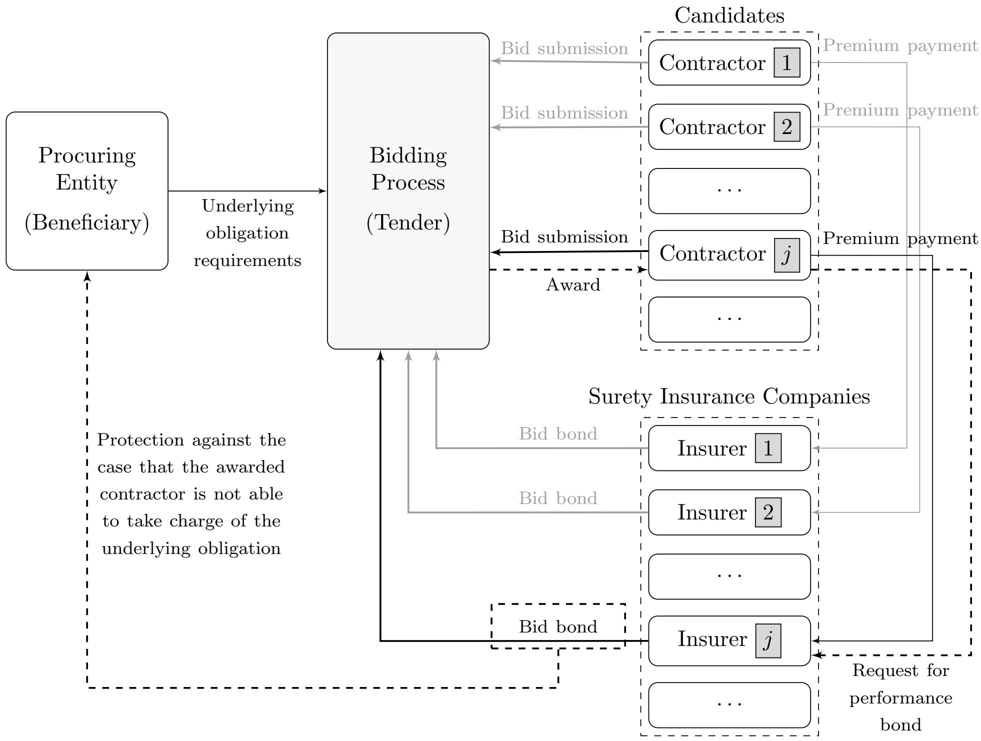

Figure 1.4. Schematics of the most typical surety insurance products: the *bid bond* and the *performance bond*. The bidding process is represented, when the beneficiary receives protection by the bid bond and the following performance bond is not issued yet.

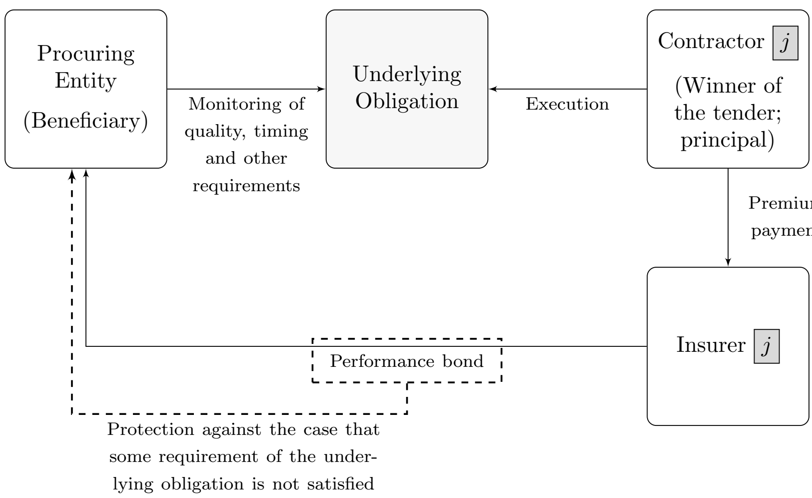

Figure 1.5. Schematics of the most typical surety insurance products: the *bid bond* and the *performance bond*. The execution phase is represented, when the beneficiary receives protection by the performance bond.

1.3.1 Bid and performance bonds: the Italian case

According to the Italian bidding law [45], the exposure generated by a bid bond is typically fixed to a $\alpha_{\text{bid}} = 2\%$ (also α_B in chapter 5) fraction of the underlying obligation notional value NV_0, established by the procuring entity at the beginning $t = 0$ of the bidding process. However, depending on the risk profile of the obligation, the procuring entity may choose a different α_{bid} value in the interval $[1\%, 4\%]$.

The performance bond exposure likewise is represented as a fraction $\alpha_{\text{performance}}$ (also α_P in chapter 5) of the obligation notional values NV_T, re-estabilished at the time $t = T$ when the tender ends. Since each candidate contractor offers to take charge of the obligation at a cost that is lower than the auction base NV_0, it holds $NV_T < NV_0$ by construction. However, the Italian law forbids to choose $\alpha_{\text{performance}}$ and provides a mechanism to protect the beneficiary from the risk arising when $NV_T \ll NV_0$ (*i.e.* quality requirements of the obligation are likely to not be met). In fact it holds [46]

$$\alpha_{\text{performance}} = 10\% + \max\{0; 90\% - \tfrac{NV_T}{NV_0}\} + \max\{0; 80\% - \tfrac{NV_T}{NV_0}\}$$

Namely, the base value of $\alpha_{\text{performance}}$ is equal to 10%, but it is increased by 1% per each percentage point exceeding a 10% bidding discount and by 2% per each percentage point exceeding a 20% bidding discount.

Hence, in the case of $NV_T \simeq NV_0$, $\alpha_{\text{bid}} = 2\%$ and $\alpha_{\text{performance}} = 10\%$, the exposure at risk guaranteed by a given performance bond is approximately five times the exposure at risk covered by the corresponding bid bond. However, depending on the choice of α_{bid} and the value of NV_T/NV_0, the notional value of the performance bond can easily reach an amount ranging between 10 and 20 times the corresponding bid bond exposure. This fact incentivizes the insurer to perform an assessment of the contractor as accurate as possible when underwriting the bid bond to avoid the choice between the bid bond payment and the issue of a performance bond that causes the exposure generated by the contractor to be too big with respect to the contractor's worthiness.

The performance bond exposure decreases over time as the completion percentage of execution increases. The effective exposure can decrease up to 20% of the initial exposure in $t = T$ [47]. However, the insurer is usually not aware of the execution status since there is no obligation for the beneficiary nor for the principal to keep the insurer update unless a claim is notified.

In case of a claim, the insurer indemnifies the beneficiary according to the rules described above, and then the subrogation takes place: the insurer acquires the right to recover from the principal the amount indemnified to the beneficiary [48].

1.3.2 Bid and performance bonds: the U.S.A. case

Public works promoted either by the U.S. federal government or by a specific country in the U.S.A. are subject to a law that shares many features with the Italian bidding law described above in paragraph 1.3.1. Indeed, processes outlined in figures 1.4 and 1.5, for the tender and the subsequent execution phase, respectively, still apply in this case. However, it is worth noticing two relevant differences with respect to the Italian case, regarding the mandatory presence of another type of surety product - the payment bond, described in the following - and a different implementation of the subrogation concept.

Each contract exceeding $ 100.000 and awarded by the U.S. federal government for the construction, alteration, or repair of a public building is disciplined by the Miller Act [10]. When a specific state requires public work in the U.S.A., similar regulatory frameworks are applied - known as "little Miller Acts" to highlight their analogy with the federal law they are inspired by.

As anticipated in paragraph 1.1.3, the Miller Act aims to protect both the government and the subcontractors during the execution phase. Indeed, the risk that the contractor does not pay subcontractors and suppliers of goods would reduce competition and raise construction costs, given the sovereign immunity that prevents a mechanic's lien. Hence, transferring the credit risk generated by the contractor is an indirect defense of the public interest. To this purpose, a *payment bond* is required together with the performance bond[18].

A payment bond can be thought of as a simplified credit insurance policy: the underlying risk is still the violation of a financial obligation, but the surety contract cannot be customized by adding implicit options and risk/profit sharing elements. Furthermore, the surety insurer has not the possibility to handle the credit limit dynamically during the coverage period, nor can the insured vary the exposure at risk by issuing new invoices to the risky debtor (*i.e.* the contractor in this case).

The American law interpretation of the performance bond is less structured and more effective than the one discussed above in paragraph 1.3.1 in the Italian case. In fact, in this case, the surety insurer is generally requested to guarantee the execution of the underlying obligation. Depending on the specific case, this may imply replacing the defaulted contractor with another executioner or the payment of the remaining part of the public work to be completed after the principal default. However, while Italian law requires paying a specific fraction of the contract value, American law requires the contract to be fulfilled somehow. Hence a sort of "double" subrogation takes place: the insurer replaces the contractor who is guaranteeing the work completion and then replaces the beneficiary in the recovery attempt against the defaulted contractor. The concept of double subrogation in American surety law is explicit in the doctrine. For example, in Hall v. Windsor Sav. Bank [49] the court observed that "*whenever the surety of a fiduciary is compelled to answer for the latter's breach of trust, he succeeds to the rights of both the fiduciary and the cestui*"[19].

[18] In the Italian case, there is no sovereign immunity, and hence the presence of a payment bond is not mandatory, although possible. Furthermore, other forms of protection are available as well. In case a given public work is managed as project finance, the banking system provides the liquidity needed to pay subcontractors and suppliers, and the banks involved in the operation can be insured against the counterparty risk generated by the contractor.

[19] *Cestui* or *Cestui que* is a shortened version of "cestui a que use le feoffment fuit fait", literally, "The person for whose benefit the feoffment was made", *i.e.* the beneficiary. The word "fiduciary" (also "trustee") is referred to the principal.

1.4 *Similar but not quite the same*: financial products related to Credit and Suretyship insurance

Subjects belonging to both SME and large corporate segments generate a vast demand for protection from credit events and from loss events belonging to the broader *breach-of-contract* class.

Insurance companies, banks, and other financial institutions have been stimulated to develop and offer a wide range of products to satisfy this need. Without claim to completeness, it is worth recalling some of the most similar products to the ones investigated in this work, to highlight their similarities and differences in comparison with the Credit&Suretyship insurance products. The following sections briefly introduce two examples of banking products that are comparable with credit insurance and suretyship, respectively. §1.4.1 summarizes the basic features of factoring operations that provide support to a seller in handling its unexpired commercial credit, implying natural comparability with credit insurance. §1.4.2 introduces banking guarantees that are a possible replacement to a surety bond, although being a financial (*i.e.*, not insurance) product.

No other financial product is explicitly considered in the following, although there are notable products designed to transfer the credit risk originated by a corporate entity or a basket of generic debtors (SMEs or even retail borrowers). E.g., Credit Default Swap (a.k.a. CDS) belongs to the former category, while Collateralized Debt Obligations or Asset-Backed Securities (a.k.a. CDO and ABS, respectively) are examples of the latter.

1.4.1 Factoring

Factoring [50] is a type of financial transaction requested by subjects who have a commercial credit against a risky counterparty. Credit insurance and factoring have remarkable similarities, although the latter being a banking product. In fact, in both cases, accounts receivable (*i.e.*, unexpired invoices) underlie a contract between the creditor and the bank or the insurer. Further, either in factoring or in credit insurance, the creditor may receive the amount due from the bank or the insurer instead of the debtor, under certain conditions.

The first relevant difference between factoring and credit insurance is that the *factor* (*i.e.*, the bank) pays the creditor in advance, buying the credit and receiving the outstanding amount from the debtor at the due date. Although the invoice is transferred from the creditor to the bank, the credit risk is not necessarily transferred as well. Indeed, the factoring transaction may be *pro-soluto* (a.k.a. without recourse) or *pro-solvendo* (a.k.a. with recourse). The former factoring type implies that the factor must bear the loss in case of the debtor's insolvency, while the latter type gives the factor the right to recover the credit from the invoice seller in case the debtor does not pay. Thus, the *pro-soluto* factoring offers actual protection against commercial credit risk, while the *pro-solvendo* factoring is just a means of financing the credit by anticipating the outstanding amount. Unlike in credit insurance, the factoring may be requested by the debtor instead of the creditor. Such a transaction is known as *reverse factoring* and allows the debtor to maintain the duration of its outstanding commercial debts, although offering an immediate payment to its creditors.

Although *pro-soluto factoring* is a form of protection against commercial credit risk, it is different from credit insurance in one relevant aspect at least. In fact, each unexpired invoice has to be purchased by the bank and, thus, the creditor must present it in advance, and the bank can accept it or not. As discussed in §1.2, the credit insurer grants a credit limit without knowing the actual exposure to the risk generated in a given time instant by the business relationship existing between the insured seller and its buyer. Hence, while the bank is required to pay in advance the amount due but enjoys complete information to decide its risk appetite level, the credit insurer is required to pay the amount due only if a claim is notified but is forced to decide its risk appetite level without knowing the amount of exposure to risk by each time instant.

1.4.2 Banking Guarantees

When a beneficiary requests protection against the breach-of-contract risk generated by a principal obligation, an insurance company can act only as a surety. In contrast, a bank can act either as a surety or as a guarantor.
In the first case, the protection offered by the bank is basically the same as the suretyship insurance products presented in §1.3. Minor differences can be found depending on the regulatory framework applied in the country where the considered principal obligation takes place. Practically speaking, the main difference is that a bank is likely to require full collateralization of the transferred risk in advance, and a suretyship insurance company usually exercises the right of subrogation against the contractor only after the claim [51].

When a bank acts as a guarantor, it always requires the full collateralization of the underlying notional value in advance. The reason for this mandatory requirement lies in the second difference between suretyship and banking guarantees. Indeed, a guarantee is a distinct promise to pay and is not dependent on the principal obligation. The guarantor must pay upon the first written demand of the beneficiary. In that perspective, the guarantee is not an insurance product. Although it provides protection against the breach of the principal obligation, the obligation does not underlie the guarantee contract that gives a unilateral right to the beneficiary [52]. As a practical consequence, a suretyship policy allows a rearrangement of the terms of the obligation, with the intermediation of the surety, before the claim is settled. In public works, it is common practice to extend the principal obligation expiry date when needed, provided that the public procuring entity considers the contractor able to fulfill its obligation, although in a time longer than expected. In such a case, the surety must accept the time extension of the underlying performance risk, while the principal has to pay an additional premium. This flexibility allows to progress and - hopefully - to complete the required public works without interruption. On the other hand, the exercise of a guarantee generates two (ideally) instantaneous cash flows, from the bank to the beneficiary and from the principal to the bank, the latter being possible through the collateralization of the notional amount. In such a case, the related - not underlying - obligation is nullified, and the contractor must be replaced.

Chapter 2

Modeling the future claim event in credit and suretyship insurance

In chapter 1, the main elements of C&S insurance have been outlined in paragraphs 1.2 (credit insurance) and 1.3 (suretyship). Now it is worth comparing these two types of insurance business in the perspective of better understanding to what extent the claim events they generate are similar and therefore can be described by using the same modeling approach.

The following chapter shows that the relevant features shared between these two insurance lines are enough to use a single model to describe the joint probability distribution of claim events generated from credit and surety risks. However, credit insurance is affected by an information asymmetry stronger than the one experienced by a surety insurer and so requires special care in estimating marginal claim probabilities. The latter problem is addressed in chapter 6, as a part of the original research developed in this work.

In the next section 2.1 differences and similarities between credit and surety are discussed concerning the modelization of their claim events probabilities.

In section 2.2, a model is selected among the "classic" credit risk models available in the literature to cope with the features of both these insurance types and so to be used as a possible joint probability distribution of their claim events. This choice allows for parametric estimation of the dependence structure among risks belonging to a mixed C&S portfolio of insurance liabilities. Details about the chosen model - that is, CreditRisk$^+$ - and its calibration are presented in chapter 4. In the same chapter, an original technique is presented to increase the precision of CreditRisk$^+$ calibration when performing it in the C&S context - completing the original research results contained in this work.

Finally, section 2.3 presents the modeling choices made by EIOPA when considering this LoB in the Solvency 2 Standard Formula framework.

2.1 Comparison between credit and surety

Credit insurance and suretyship have many features in common that justify the choice of supranational authorities to represent them as a whole and to issue a unique regulation relevant to both of them [36, 37, 41]. Conversely, national insurance laws may consider them as distinct lines of business, depending on the specific country[1].

Their similarities are also relevant from an actuarial perspective to choose the most suitable modeling approach to represent the various components of their underwriting risk. Among their features, the ones of interest in modeling claim events probability are listed and discussed hereinafter.

i. *Claims as absorbing events.* This can be considered a common feature shared between credit and suretyship to a fair approximation. When the beneficiary notifies a claim to the surety, subrogation is activated, implying the indemnification or the contractor replacement and the subsequent attempt of the insurer to recover the loss. This sequence can occur once at most in the policy life cycle, so it is an absorbing event. On the other hand, an insolvency event in credit insurance may be not related to the buyer's bankruptcy but just to a temporary condition of financial distress. Hence it is possible that a buyer generates a credit insurance claim, improves his or her creditworthiness again, and then generates a second claim after some time. However, this double claim event can be considered negligible since the insurer is likely to exercise the right to nullify the credit limit as the first claim event takes place, avoiding the second event. Furthermore, the insured seller likely decides to interrupt the business relation with the insolvent buyer after the first event due to his or her loss implied by the risk-sharing elements of the credit insurance policy.

ii. *Heterogeneity of loss events.* The suretyship insurance is characterized by a wide range of loss event types. In contrast, credit insurance aims to protect the insured only against the risk of losses arising from the insolvency of a debtor. As discussed in paragraph 1.3, a suretyship policy may guarantee the beneficiary against credit risk as well (*e.g. payment bond* introduced in paragraph 1.3). However, surety insurance products also protect against the risk that almost any other kind of obligation is violated, including, *inter alia*, performance requirements, and self-declarations accuracy. Suretyship also guarantees the commitment not to do something (*e.g.* the obligation not to renovate a historic building while renting it).

iii. *Scarcity of information about (the majority of) sources of risk.* In principle, scientific literature developed for the banking sector offers a wide range of techniques dedicated to estimating the probability of insolvency events. However, these techniques require the availability of specific pieces of information, such as the price of stocks or bonds listed in markets or, at least, an updated financial statement. This information is usually available for subjects

[1] According to Italian insurance law, credit and surety are classified as distinct lines of insurance business (n. 14 and 15 respectively). However, they are regulated by standard administrative orders and laws [41].

belonging to the corporate segment, while the typical source of risk for credit insurers is the SME segment. It is true that also banks have to manage the credit risk generated by the SME segment, but with two relevant advantages. First, the debtor relates directly with the bank, which has the right to request all the information needed to evaluate the debtor's creditworthiness. On the other hand, credit insurers evaluate the debtors only based on the information available to the insured seller - usually another SME - and (when applicable) to info providers.

Furthermore, banks provide information on the payment behavior of customers to each other nationwide. At the same time, a centralized database of claims does not exist in the credit insurance sector, nor credit insurance companies are allowed to access centralized data collected by their domestic banking system on their debtors. In surety, the heterogeneity of insured loss events makes estimating the claim probability for a specific underlying obligation even more difficult. There are no data or models to quantify an obligor's idiosyncratic risk not to fulfill his or her obligation for the majority of obligation types (namely, all the types not directly related to credit risk). As might be expected, technical evaluations are possible in the case of extensive public works, where projects and other features of the underlying obligation are available to the insurer to some extent. However, these evaluations are expert-judgment based, and it is unlikely that they allow quantifying the probability of the corresponding surety claim.

iv. *Information asymmetry with the insured: duration of risk.* This kind of information asymmetry concerns surety insurance specifically. A credit insurer knows precisely how long he or she may receive a claim notification after having nullified the credit limit granted on a specific buyer. In fact, the loss event cannot occur after the maximum credit term passes and must be notified within the maximum extension period. On the other hand, the maturity of a surety policy can be extended even after its expiration date. Indeed, suppose the beneficiary is a public entity, and the underlying obligation is related to public works. In that case, the insured is obliged to accept a premium supplement against providing an extended guarantee. In principle, this can occur multiple times before the actual expiration date is reached, implying a stochastic duration for a relevant fraction of surety insurance products.

v. *Information asymmetry with the insured: exposure at risk.* This kind of information asymmetry concerns credit insurance specifically. The presence of an actual exposure at risk generated by a given buyer depends on the simultaneous existence of two elements: a nonzero credit limit and (at least) one outstanding invoice issued by the insured seller to the underlying buyer. The second condition only depends on the commercial relation ongoing between buyer and seller. Hence, the credit insurer cannot say which risks generate an actual exposure at a specific time - not only in the present but, in most cases, also in the past. On the other hand, the severity of a surety claim may be *a priori* unknown - since the insurer is not (entirely) aware of the progress status of the principal obligation - but exposure at risk is certainly present until the principal obligation is not redeemed.

vi. *Insensitivity to creditworthiness dynamics.* This is a common feature shared between credit and surety. Although credit insurance protects against insolvency events and most of the suretyship loss events are somewhat related to the case of an obligor default, the variation of the underlying risk (debtor/obligor) creditworthiness during the life cycle of the policy does not represent a loss event in itself. Even if the source of risk belongs to the large corporate segment and hence observable and updated information is available, such as a rating issued by an ECAI rating agency, C&S products are triggered only by their specific underlying loss events. Namely, a defaulted debtor that can pay the last invoice issued by the insured seller or a defaulted obligor promptly replaced in fulfilling his or her obligation towards the beneficiary does not generate any loss event. Although there is no direct effect of creditworthiness migration onto claim probability of any C&S product, there is an interesting case of possible *indirect* effect in suretyship insurance. When a surety guarantees a bidder in a public tender process if the granted bidder wins the tender has to obtain a performance bond, commonly issued by the same insurer who issued the bid bond, as a requirement to enter the contract. If the bidder's creditworthiness worsens remarkably during the tender process, the insurer may refuse to issue the performance bond, and the same may apply to the other C&S insurers. This scenario implies an increased claim probability for the bid bond. However, as discussed later in chapter 5, this effect is not worth being considered for practical purposes.

vii. *High volatility of claim probability and loss ratio.* Both credit and surety insurance policies protect low-frequency absorbing events. However, frequency and the degree of dependence among these loss events may grow depending on systemic factors, such as the macroeconomic situation of a considered country. Hence, this line of business results especially volatile over the years if compared to other lines of business of the non-life insurance sector. For example, let us compare the recent claim history of C&S insurance against two among the main[2] non-life LoBs in the Italian market. As shown in table 2.1, the historical default frequency observed in non-performing loans (NPLs) over the period 2013-2018 is approximately five times as volatile as the claim frequency recorded for GLI and FDP during the same period. In relative terms (*i.e.* considering CV instead of RSME), NPLs default frequency still results three to five times as volatile as the claim frequencies of the two considered LoBs. Although being recorded by the banking sector [54], NPLs are a good benchmark to study the trend of credit insurance claim frequency, given the lack of publicly available information about C&S number of claims per year. Similar results are obtained comparing loss ratios (see figure 2.1): in this case, historical data are disclosed also about Credit and Surety LoBs[3]. It is worth

[2]Taken together, GLI and FDP LoBs are worth 25.8% of premiums earned by Italian non-life insurance sector in 2019, while C&S LoB generated 1.4% of Italian non-life insurance premiums in the same year [82]. This premium distribution is comparable to the one observed in the European non-life insurance market in 2018 (see figure 1.2).

[3]In Italian insurance law, Credit insurance and Suretyship are two distinct LoBs (number 14

noticing not only that credit and surety loss ratios exhibit a strong reaction to systemic crises (increasing their volatilities), but also that the reaction is not the same. Indeed, figure 2.1 shows that suretyship seems to be affected especially by the 2003 crisis, while credit insurance results more impacted by 2009 and 2012 crises. This is likely to be related to "feature ii." above. The strong correlation between C&S performance and the economic cycle has been also recognized and considered by CEIOPS in Solvency 2 Standard Formula design[4].

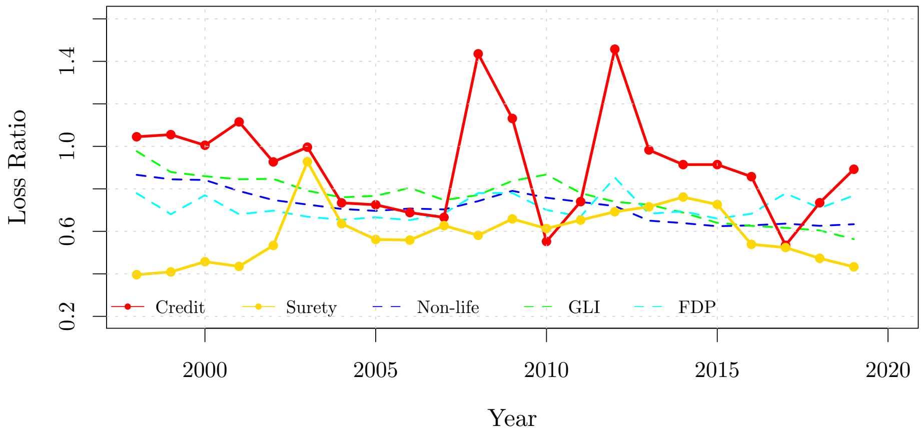

Figure 2.1. Italian market (1998-2019): comparison among historical loss ratio time series of the entire non-life Italian insurance market, and the ones of four lines of non-life insurance business. Data source: ANIA [82].

Features listed above are summarized in table 2.2. These elements allow to draw some conclusions about the representation of a C&S future claim event and the estimation of its probability.

and 15 respectively), while they are considered as a whole in European Solvency 2 framework (*i.e.* LoB n. 9).

[4]Paragraph 3.1143 of *Solvency 2 Calibration Paper* [83] reads as follows: "*In light of the credit crisis, due attention was given to concerns regarding pro-cyclicality of financial systems and their regulatory regimes. One particular insurance field on which this concern has focused is credit insurance and suretyship (C&S). For instance, the EFC report to the Council of the European Union states that credit insurance is, in terms of its risks, substantially similar to the banking business and faces the same pro-cyclical challenges.*[...]."

General liability insurance			
Year	number of risk units	number of claims	claim frequency
2013	17,788,748	364,957	0.021
2014	17,875,067	333,750	0.019
2015	17,739,068	319,358	0.018
2016	19,222,243	310,237	0.016
2017	18,494,292	312,445	0.017
2018	19,695,904	325,578	0.017
Fire and other damage to property insurance			
Year	number of risk units	number of claims	claim frequency
2013	31,556,833	1,186,587	0.038
2014	31,284,433	1,234,661	0.039
2015	30,816,617	1,139,994	0.037
2016	32,340,234	1,127,954	0.035
2017	33,916,290	1,178,279	0.035
2018	36,320,597	1,259,482	0.035
Non-performing loans			
Year	number of borrowers	number of defaults	default frequency
2013	1,314,117	59,389	0.045
2014	1,234,151	53,417	0.043
2015	1,183,963	44,417	0.038
2016	1,156,716	34,995	0.030
2017	1,146,146	27,555	0.024
2018	1,144,550	24,829	0.022
	GLI	**FDP**	**NPL**
Mean	0.018	0.036	0.034
RMSE	0.002	0.002	0.010
CV	9.19%	5.39%	29.34%

Table 2.1. Italian market (2013-2018): comparison among default frequency of Non-performing loans (NPL) and historical claim frequency of two non-life insurance lines of business (*general liability insurance* - GLI and *Fire and other damage to property insurance* - FDP). Mean, root square mean error and coefficient of variation are compared among the three considered time series. Data sources: IVASS [53] for GLI and FDP time series; Bank of Italy [54] for NPL time series.

		Credit	Surety
i.	Claims as absorbing events	✓	✓
ii.	Heterogeneity of loss events	✗	✓
iii.	Scarcity of information about (the majority of) sources of risk	✓	✓
iv.	Information asymmetry with the insured: duration of risk	✗	✓
v.	Information asymmetry with the insured: exposure at risk	✓	✗
vi.	Insensitivity to credit worthiness dynamics	✓	✓
vii.	High volatility of claim probability and loss ratio	✓	✓

Table 2.2. Comparison between credit and surety: recap of similarities and differences presented in paragraph 2.1.

Loss events as Bernoulli variables. *i.* and *vi.* imply that a risk source (*i.e.* an obligation or a commercial relation underlying a C&S policy) can generate losses once at most during the whole coverage period. Hence, the occurrence of the claim over a given time horizon is well represented by a Bernoulli r.v. by construction. Therefore the marginal probability associated with such an event can be estimated as the Bernoulli distribution parameter. Namely, this means that it is necessary to estimate the time series of historical claim frequencies as first and proceed to further model the probability of future events. The parametric estimation of absorbing events probability is discussed in chapter 3. As anticipated in *vi.*, the Bernoulli representation of the underlying risk fits bid bonds claim events as well. In fact, the effects of creditworthiness migration, if any, are *indirect* - as better discussed in chapter 5 - and a claim is still an absorbing event even if related to migration phenomena.

Clustering risk sources into risk-homogenous groups. *ii.* implies that C&S insurance offers protection against claims which are different and therefore have to be clustered into distinct groups, internally homogenous, to avoid the average among the possibly different probability levels associated with each group. Indeed, the mix of distinct populations would lead to confusing their dynamics, causing a bad modelization. Furthermore, *iii.* suggests that the claim probabilities associated with the majority of the debtors have to be evaluated on a clustering basis only. In fact, the categorical variables needed to characterize each debtor in terms of economic sector, geographical origin, and legal form are the only available information in many cases and are perfectly suitable to clusters definition purposes.

Estimating historical claim frequencies under a lack of information. *iv.* and *v.* highlight that C&S insurers have to face a significant lack of information implied by the features of this LoB discussed above. However, it is worth remarking that information asymmetries - though present both in credit and suretyship - are different in the two cases and thus have distinct consequences in terms of claim probability estimation.
In the suretyship case, the beneficiary information advantage (presented in *iv.*) does not cause any problem to the surety when estimating historical claim frequencies since it affects not expired risks only. Hence, the most relevant practical consequence arises when estimating the premium provision component of best estimate liabilities. In fact, Solvency II requires the quantification of all the future cash-in and cash-out installments until run-off, generated by the insurance liabilities underwritten up to the evaluation date[5]. Therefore, surety insurers are required to estimate an effective expiry for each underwritten risk that the beneficiary reserves the unilateral right to extend. This can be easily achieved by the application of historical expiry distributions per policy type. Furthermore, the estimation of claim probability per unit of time is not impacted.
On the other hand, the partial information available to credit insurers about commercial relations between their insured sellers and buyers affects the insurer's ability to measure the historical frequency of potential loss events, due to past default events which may be not observed depending on the insured seller behavior. In fact, given a risk homogeneous cluster of buyers observed over the same time horizon, the insurer will receive a number of claim notifications that is affected by the probability of having an invoice issued by the insured seller and still outstanding when the buyer defaults. Therefore, in the credit insurance case, the commercial relation affects the measure of historical default rates. Such a bias is not easily removable since the information needed (*i.e.* historical insured invoices issued to each risky buyer) is not available in most cases, even after the policy expires. However, the factorization of default probability and claim observability may be relevant if both evolve in time under independent dynamics. In this case, modeling them as a whole may cause a misjudgment of recently observed trends and thus bad forecasting. This problem is addressed in chapter 6.
Furthermore, claim rate estimation in C&S insurance may be made more difficult by a reduced sample size. Indeed, table 2.1 shows that when considering the whole Italian banking system - where information is shared among the banks - the average number of risky borrowers is an order of magnitude lower than the typical number of risk units in other common non-life insurance LoBs (FDP and GLI are compared

[5]Solvency 2 "Commission Delegated Regulation" 2015/35 [37], Chapter III, Section 3, Subsection 3, Article 28 reads: "*The cash flow projection used in the calculation of the best estimate shall include all of the following cash flows, to the extent that these cash flows relate to existing insurance and reinsurance contracts: (a) benefit payments to policy holders and beneficiaries; (b) payments that the insurance or reinsurance undertaking will incur in providing contractual benefits that are paid in kind; (c) payments of expenses as referred to in point (1) of Article 78 of Directive 2009/138/EC; (d) premium payments and any additional cash flows that result from those premiums;* [...] *(g) payments for salvage and subrogation to the extent that they do not qualify as separate assets or liabilities in accordance with international accounting standards, as endorsed by the Commission in accordance with Regulation (EC) No 1606/2002;* [...]".

NPL in the table). Since C&S insurers do not share information about their risks and claims, the number of risk units observed by a specific insurance company can be smaller by two orders of magnitude or more[6]. The necessity to cluster risk sources into risk-homogenous groups reduces the sample size per group further. Hence, samples used to estimate claim frequencies may be scarcely populated. Combined with the generally low claim rate observed in C&S LoB, this fact leads to a relevant estimation error associated with claim frequency measurement.

Modeling multivariate dynamics of claim probabilities. As stated in *iii.*, credit insurance is mainly devoted to offering protection against business entities belonging to the SME segment. Hence, it is reasonable to suppose that their insolvency events are mutually independent in general. Indeed, even in the case of bankruptcy, a small business entity is most likely not to provoke any contagion phenomenon at all. However, as observed in *vii.*, credit insurance is sensible to crises, as well as to macroeconomic evolutions in general. This implies factor models to be a possible "natural" representation of the multivariate dynamics of claim probabilities associated with a generic set of risk homogeneous clusters considered in this context. Models belonging to this class describe the collective dynamics of random variables set (allowed to be conditionally independent) subjected to a dependence structure defined through an additional set of latent variables (commonly known as "factors") that represent the environment affecting the observable variables.
As discussed in *iii.* and *vii.*, the same applies to suretyship, which is influenced by macroeconomic phenomena as well. However, suretyship claims can be reasonably supposed to be conditionally independent since the violation of an obligation underlying a surety policy is not likely to affect the probability of observing further violations of other insured obligations. Furthermore, as discussed in paragraph 1.3 and recalled in *ii.*, suretyship insurance offers protection against a variety of loss event types. Hence, the flexibility provided by a factor model results to be especially fitting in this case, where different loss event types could also exhibit a different dependency on each latent factor.
In general, calibrating a dependence structure is a demanding task. In this case, the lack of information experienced by C&S insurers, due to both the insured behavior and the reduced number of observed risk units per cluster, makes it even harder. The problem of calibrating a factor model for C&S insurance applications is addressed in chapter 4, which presents a technique to increase the precision of the dependence structure among claim probabilities in this context.

2.2 Model selection for a C&S basket of risks representation

As discussed in paragraphs 1.2 and 1.3, a C&S product protects a subject (not necessarily the insured one, as seen in par. 1.3) from the failure of another subject in fulfilling a given obligation. The underlying obligation can have either a financial nature, as in credit insurance products and payment bonds, or a nature that is related to the creditworthiness of the risky subject to some extent (*e.g.* in

[6] *E.g.*: The number of C&S risk units recorded in [84] is approximately equal to 10^5.

performance bond contracts, the contractors who are not able to fulfill their duties are likely to be in a situation of financial distress). Even in the case when the loss event is unrelated to a credit event at all (*e.g.* court bonds not related to financial obligations), it maintains a certain similarity with a default event, at least in a modeling perspective, since it can always be regarded as "absorbing" - as discussed in paragraph 2.1.

Thus, it is worth recalling the main "classical" modeling approaches available in the literature for representing credit risk in a financial context to evaluate their degree of suitability for C&S insurance applications. Some amongst the most relevant credit risk models are briefly reviewed hereinafter, without claim to completeness[7], neither in the list of the models presented below nor in the description provided about each model. The considered models have been initially developed for the banking sector between the 1970s and 1990s but are still an industry standard nowadays.

According to a classification that is widely accepted in the literature, credit risk models are usually categorized as *structural* or *reduced form* models. Structural models describe a firm's bankruptcy event through a microeconomic model of the firm's capital structure. In contrast, reduced-form models directly represent the default probability using a variable or a process not explicitly related to the firm's balance sheet. Examples belonging to both categories are discussed in the following paragraphs.

However, before discussing each considered model and its feasibility for C&S applications, it is worth emphasizing that the distinction between structural and reduced-form models introduced above is weak, although commonly used. Depending on how parameterization is done, the same underlying assumptions can lead either to a structural or a reduced-form model. Loosely speaking, when the underlying microeconomic model of the firm's capital structure is directly used in calibrating the corresponding credit risk model, the latter is considered to be "structural". On the other hand, if the functional form implied by a set of microeconomic assumptions can be calibrated by exogenous default observations, without needing the direct observation of the financial variables involved in the assumptions, the same credit risk model becomes a "reduced-form" model.

Among the models presented in the following, only the Merton model, Moody's KMV model, and credit scoring models are classified as structural. Other models based on the same structural assumptions introduced by the Merton model (*i.e.* ROA being normally distributed and default event caused by a critical decrease in assets value), such as CreditMetrics and Vasicek model, are considered to be reduced-form because of the possibility to calibrate them considering a latent assets dynamics.

Some models are considered neither reduced-form nor structural. This third set includes all the models where the probability of default is explicitly dependent on some non-latent variables, but the considered variables are not microeconomic. The first and most important example of this model type is CreditPortfolioView, presented in the following.

[7] An exhaustive discussion on this topic is available, *e.g.*, in the books of Schönbucher [116] and McNeil [122].

2.2.1 Structural models: Merton model and Moody's KMV model

The so-called "*Merton model*", introduced in 1974 by Robert C. Merton [55], is the first example of a structural model ever developed. Merton model has been furtherly developed by KMV, a financial analysis society founded by Kealhofer, McQuown, and Vasicek. The acquisition of KMV by Moody's in 2002 [56] leaded to the "*Moody's KMV model*" [57] - only partially disclosed to date.

Model hypotheses and structure

The Merton model uses a log-normal process to describe the dynamics of a given firm's assets value V_t

$$dV_t = \mu V_t dt + \sigma_V V_t dW_t$$

where W_t is a Wiener process. The firm's equity is evaluated as the difference between V_t and the debt level B. The default event is supposed to occur when $V_t \leq B$, implying a null or negative equity value.

Hence, the following expression holds for the firm's probability of default (PD) in t

$$\text{Prob}(V_t \leq B|\mathcal{F}_0) = \Phi\left(\frac{\log\frac{B}{V_0} + (r - \frac{\sigma_V^2}{2})t}{\sigma_V\sqrt{t}}\right)$$

where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution and r the risk-free interest rate. Since $\Phi(x) = 1 - \Phi(-x) =: \overline{\Phi}(-x)$ we have that the 1-year PD is $\text{Prob}(V_1 \leq B|\mathcal{F}_0) = \overline{\Phi}(A)$, where

$$A := \frac{\log V_0 - \log B - (r - \frac{\sigma_V^2}{2})}{\sigma_V}$$

According to Moody's KMV framework, PD is also referred to as *Expected Default Frequency* (EDF), so that the current version of the model has been renamed "Moody's Analytics EDF". An easily understandable variable, known as *distance-to-default* (DD) is introduced instead of A in the passage from the original Merton's formulation of the model to the contemporary Moody's KMV framework.

$$\frac{\log V_0 - \log B - (r - \frac{\sigma_V^2}{2})}{\sigma_V} \approx \frac{\log V_0 - \log B}{\sigma_V} \approx \frac{V_0 - B}{\sigma_V V_0} =: DD$$

Moreover, Moody's KMV model improves the original Merton model by weakening its distributional hypothesis. Indeed $\overline{\Phi}(\cdot)$ is replaced by an *empirical DD-to-EDF map* $M(\cdot)$ (not disclosed to date) which is calibrated on historical data. This choice implies that the log-normal distribution hypothesis is abandoned - increasing the flexibility of the model and, according to Moody's, its predictivity [57].

Although the Merton model aims to provide only an explicit description of the PD associated with each considered firm, it can be easily extended to evaluate the distribution of future losses generated by a basket of risks. Indeed, since the PDs are inferred from the dynamics of stock prices, the model can be easily coupled with any structure of dependence consistent with the assumption about the lognormality of the stock prices' marginal dynamics.

2.2.2 Structural models without dynamics: credit scoring models

Merton model, presented in §2.2.1, can be thought as based on two assumptions:

- the (deterministic) relation existing between a microeconomic variable describing the state of a firm and its creditworthiness;
- the (stochastic) marginal dynamics of the considered microeconomic variable.

This setting is not shared among all the structural models. A comprehensive class of structural models - known as "cross-sectional" credit scoring models - represents a commonly accepted approach to model the default probability PD of a firm as a function of the information available from the firm's financial statements, without introducing any additional assumptions about dynamics [58].

Model hypotheses and structure

In a nutshell, a typical credit scoring model is based on the hypothesis that the default probability PD of a firm F, estimated in t over a given time horizon $(t, t+\Delta t]$, can be expressed as a generalized linear function of some financial ratios and/or other numerical values taken from the firm's financial statements.

$$\mathrm{PD}(F, t, \Delta t) = \mathbf{E}\left[\mathbb{1}_{\{\tau_F \in (t, t+\Delta t]\}} | \mathcal{F}_t\right] = f\left(\beta_0 + \sum_{i=1}^{N} \beta_i x_{iF}(t)\right)$$

where τ_F is the time to default of F, $\boldsymbol{\beta} := (\beta_0, \beta_1, \ldots, \beta_N) \in \mathbb{R}^{N+1}$ is the array of the model parameters to be calibrated and $x_{iF}(t) \in \mathbb{R}$ is the value of the i-th considered variable, measured in t from the F's financial statements and/or other selected information sources.

The function $f : \mathbb{R} \to [0, 1]$ is chosen according to tractability criteria. Unlike in Merton model, in this case there are no assumptions about the dynamics of credit worthiness that imply a form of $f(\cdot)$. Two typical[8] choices [58–60] are the standard logistic function

$$f(x) \equiv \frac{e^x}{e^x + 1} = \frac{1}{2} + \frac{1}{2}\tanh\left(\frac{x}{2}\right)$$

and the standard normal CDF

$$f(x) \equiv \frac{1}{2}\left[1 + \frac{1}{2}\mathrm{erf}(x)\right],$$

where $\mathrm{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x \exp\left(t^{-2}\right) dt$ is the error function.

In this context, it is commonly assumed that default events over $(t, t + \Delta t]$ are distributed as i.i.d. Bernoulli random variables, conditionally to the state $\mathbf{x}_F(t)$ of each firm. Under this assumption, the two $f(\cdot)$ forms listed above lead to the *logit* and *probit* models[9], respectively. A reason for the popularity of these models is the

[8] A comparative analysis on the presence of each cross-sectional model in the literature can be found in [60], §3.3: the "logit" and "probit" models emerge as the most commonly studied in terms of number of papers.

[9] These models are named as the corresponding inverse functions $f^{-1}(\cdot)$. In fact, *logit* is the inverse of the logistic function, while *probit* is the quantile function associated with the standard normal distribution.

possibility to calibrate them by ML estimation of their parameters from historical data. Indeed, the likelihood function $\mathcal{L}$ can be easily written in closed form due to the independence among defaults.

$$\mathcal{L}\left(\boldsymbol{\beta}|\mathcal{F}_{t'}\right) = \prod_F f(\boldsymbol{\beta}\cdot\mathbf{x}_F(t))^{D_F}\left[1 - f(\boldsymbol{\beta}\cdot\mathbf{x}_F(t))\right]^{1-D_F}$$

where $t' \geq t + \Delta t$, $\mathbf{x}_F(t) \coloneqq (1, x_{1F}(t), \dots)$ and $D_F \coloneqq \mathbb{1}_{\{\tau_F \in (t, t+\Delta t]\}}$.

Furthermore, in both cases, the optimal parameters choice is easily achievable due to the computable form of the first derivative: in logit model, $f'(\cdot)$ can be represented as a closed-form expression of $f(\cdot)$, since the logistic function is a solution to the differential equation $f'(x) = f(x)[1 - f(x)]$, while in probit model $f'(x)$ is the standard normal PDF. Generally speaking, the choice between the two models is not relevant to practical purposes in most cases since their outcomes are very similar [59, 61].

In banking practice, this technique is commonly embedded in a wider framework [61–63] (*i.e.* an internal rating system), applied to assess the credit risk profile of risky debtors in homogenous[10] portfolios. Without claim to completeness, some of the elements that are usually introduced to apply a *logit/probit* model (or other comparable approaches) to practical purposes are listed below:

- univariate selection of the variables to be included among predictors $\mathbf{x}_F(t)$, according to a measure of their diagnostic ability;
- multivariate validation of the selected predictors and dimensionality reduction of $\mathbf{x}_F(t)$ by the application of PCA or other factor analysis technique;
- partition of the PD domain $[0, 1]$ in a finite set of indexed subintervals (*i.e.* rating classes, also known as *grades*), each of them being associated to a symbol (*e.g.* AA, A, BBB, etc.) and to a qualitative description of the corresponding risk level (the so-called *master scale*);
- allowance for expert-judgment-based override of the rating, leading to a joint usage of quantitative and qualitative results to produce a final evaluation of the firm creditworthiness;
- backtest and measure[11] of model performance.

It is relevant to note that the complete specification of a rating model still need the estimation of historical default rates and forward-looking default probabilities in the same portfolio/cluster to which the model is being applied [62, 63, 65, 66].

ML calibration on historical data is based on Δt-long observations (being Δt typically equal to 1 year) of defaulted/survived enterprises collected in a past period

[10] In this context, a cluster of debtors is *homogenous* if their PDs are supposed to be related to the same predictors $\mathbf{x}(t)$ by the same parameters set $(\alpha, \beta_1, \dots, \beta_N)$. Typical examples of homogeneous clusters consist of enterprises that belong to the same segment, economic sector, and geographical area (*e.g.* European financial large corporate; US agricultural small/medium enterprises).

[11] A typical measure of predictivity is the Accuracy Ratio, that can be expressed as a function of the AUC (area under the curve) of the Receiver Operating Characteristics [64].

that spans several years (*i.e.* a whole economic cycle or more). The resulting default rate associated with the calibration sample is a long-run average, and the "natural" map *grade*↔PD is built accordingly.

However, a financial entity may be interested in defining a different master scale. From a short-term perspective, the forward-looking PDs expected for the next year could be significantly far from the long-run average (either above or below). Depending on the considered application, the short term level, known as *Point-In-Time* (PIT) PD, often results to be more appropriate[12] and thus the master scale has to be adjusted to reflect an average PIT PD level across the grades, instead of the natural long-term PD level.

In literature, the PIT PD concept is usually contrasted with the *Through-The-Cycle* (TTC) PD. It is worth noticing that "TTC" is a slightly ambiguous expression. Indeed, some authors identify the TTC PD with the long-run average PD level through a whole economic cycle - as its name suggests (see *e.g.* [66]). However, according to the Basel Committee, TTC PD concept is associated to a prudential long term PD level[13], instead of the average default rate observed. Considering this second possible meaning, also the application of the TTC PD level to the master scale requires an adjustment.

Several techniques are available in the literature to adjust the PD associated with each grade in a master scale. The seminal work of Falkenstein *et al* [61] suggests[14] to scale each PD in the master scale by the coefficient such that the average PD in the calibration sample is equal to the target PD level (*i.e.* PIT/TTC/other). Hence this approach implies that the shape of the PD(*grade*) profile must not be affected by the adjustment. A discussion where this choice is compared with other non-uniform adjustment techniques is available in [65].

2.2.3 Econometric models: McKinsey's CreditPortfolioView

Before introducing some classic reduced-form models in the following sections, it is worth recalling the well-known CreditPortfolioView model, first proposed by Wilson in September 1997 [68, 69]. In a sense, this model constitutes a third possible approach to credit risk modeling between structural and reduced-form models. In fact, the so-called *econometric* models, started by CreditPortfolioView, assume that PDs depend on exogenous variables - like in the structural models' case. However, these variables are not specific to each considered risky debtor. Indeed, macroeconomic

[12] According to "The Internal Ratings-Based Approach" (BIS, 2001) [62], Section E, paragraph 54, p. 12, banks tend to consider the PIT PD level more often than the long-run average PD. Moreover, IFRS 9 standard (see [67], paragraph B5.5.52) requires the estimation of a PIT PD, as discussed in [66].

[13] From "The Internal Ratings-Based Approach" (BIS, 2001) [62], Section E, paragraph 53, p. 12: [...] *In a "point-in-time" process, an internal rating reflects an assessment of the borrower's current condition and/or most likely future condition throughout the chosen time horizon. As such, the internal rating changes as the borrower's condition changes throughout the credit/business cycle. In contrast, a "through-the-cycle" process requires assessment of the borrower's riskiness based on a worst-case, "bottom-of-the-cycle scenario" (*i.e., *its condition under stress). In this case, a borrower's rating would tend to stay the same throughout the credit/business cycle.*

[14] See *RiskCalc for private companies: Moody's default model* [61], Section VII: Mapping To Default Rates And Moody's Ratings.

indexes $\{X_k\}$ $(k = 1, \dots, K)$ are considered instead, each of them being modeled by an autoregressive process. Applying an autoregressive process is possible due to the availability of historical time series. In case of structural models, the only variable with a populated time series is the stock price, which leads to the KMV model described in §2.2.1. Other idiosyncratic variables available in the financial statement of a debtor are observed yearly. Hence, modeling them through autoregressive processes is not feasible due to the insufficient number of observations. In such a case, dynamics is eliminated from the model, leading to the scoring models introduced in §2.2.2. Macroeconomic variables can be considered instead, enabling the application of autoregressive processes to describe the regressors' dynamics in a framework similar to the scoring models.

Model hypotheses and structure

The model assumes a logistic dependence of the 1-year PD on $\{X_k\}$. Namely, the same functional form introduced in §2.2.2 holds

$$\mathrm{PD}_i(t, \Delta t) = \frac{1}{1 + \exp\left(-\beta_{i0} - \sum_{i=1}^{N} \beta_{ij} X_j(t) + \varepsilon_{it}\right)}$$

Three main differences are worth being highlighted when comparing CreditPortfolioView with a credit scoring model. First, PD_i depends on a i-th economic sector instead of a given specific firm F. Second, an explicit innovation term

$$\underline{\varepsilon}_t \sim \mathcal{N}\left(0, \Sigma_\varepsilon\right)$$

is introduced, where $\underline{\varepsilon}_t$ is the stacked vector of errors ε_{it} associated to each macroeconomic variable and Σ_ε is the covariance matrix $\underline{\varepsilon}_t$. Third, the financial statement's variables $x_{jF}(t)$ are replaced by the macroeconomic variables $X_j(t)$.

As anticipated above, each $X_j(t)$ is modeled through an autoregressive process to allow forecasting

$$X_j(t) = \gamma_0 + \gamma_1 X_j(t-1) + \gamma_2 X_j(t-2) + \vartheta_{jt},$$

where

$$\underline{\vartheta}_t \sim \mathcal{N}\left(0, \Sigma_\vartheta\right)$$

is the stacked vector of errors associated to each macroeconomic variable $X_j(t)$ and Σ_ϑ is the associated covariance matrix.

This frameworks allows to forecast each PD_i by forecasting the set $\{X_j(t)\}$, $(j = 1, 2, \dots)$. The idea underlying the model is easily generalized by considering different autoregressive models and a diverse link function between $\{X_j(t)\}$ and PD_i.

Further, in [68] Wilson proposed to extend the model including also rating migration dynamics. Indeed, the model provides scenarios for the $PD_i(t' > t)|\mathcal{F}_t$ dynamics associated to a given i-th creditworthiness class (*e.g.*, investment grade, high yield, or a specific rating class such as BB). Hence, the ratio $r_i(t') := PD_i(t')/\phi PD_i$ can be evaluated, where ϕPD_i is the average TTC PD associated with the i-th cluster, according to the notation used in [68]. For each scenario and future time

t', $r_i(t')$ identifies the corresponding phase in the macroeconomic cycle. Then the model chooses the historical rating migration matrix observed in the past for the same r_i value - or, at least, an approximately comparable level - and applies it to the considered scenario. This approach can be maintained also in a multi-period framework, where a pattern $r_i(t')$ $(t'-t=1,2,\dots)$ is simulated per scenario and the corresponding 1-year transition matrixes $M(r_i)$ are composed by a product operation:

$$M = \prod_{t'} M\left(r_i(t')\right).$$

It is worth anticipating the main difference between this model and the reduced-form factor models introduced in the remainder of this chapter. CreditPortfolioView describes the risk factors and their relation with the default probabilities explicitly. This feature enables the indirect forecasting discussed above but requires that the dependence of PD_i on $\{X_j(t)\}$ is verified. On the other hand, latent factors models introduced in the following, such as CreditRisk$^+$, do not provide a macroeconomic identification of the latent variables considered. Thus, a not specified set of market factors is always applicable to describe the dependence structure among the risky subjects, without any further assumption, but forecasting PD_i trends based on the market factors dynamics is not feasible.

2.2.4 Reduced-form models with rating migrations: JLT and CreditMetrics models

In the following, two classical reduced-form models are presented. Both the models were developed in the late '90s to describe the creditworthiness dynamics of a risky subject.

Jarrow-Lando-Turnbull model (JLT), published in 1997 [70], is the improved version of the former Jarrow-Turnbull model [71], which is the first reduced-form model developed in credit literature. The model aims to evaluate risky cash flows in a complete, frictionless market and derivative products whose underlying risk is the creditworthiness of a given counterparty.

In the same year, CreditMetrics was disclosed by JP Morgan & Co [73]. CreditMetrics is designed to assess the risk profile of a given portfolio of risky bonds. Although classified as a reduced-form model due to its exogenous PDs, it is worth highlighting that its dependence structure is based on a generalization of the structural Merton model.

Model hypotheses and structure - Jarrow, Lando and Turnbull model

JLT assumes that a risky cash flow C_T receivable in T can be represented in t as

$$C_T|\mathcal{F}_t = \frac{B(t)}{B(T)} N\left[R + \mathbb{1}_{\{\tau>T\}}(1-R)\right],$$

where N is the notional amount due in T, R is a deterministic recovery rate, τ is the stochastic time when the debtor defaults, and $B(t) := \exp[\int_0^t r_s ds]$ is the money market account. No assumption is introduced on the risk-free spot rate dynamics r_t. However, r_t and the creditworthiness of the debtor are supposed to be independent under the risk-neutral measure of probability $\tilde{P}$. Hence, the price in t of the ZCB associated with C_T is

$$c_{t,T} = \tilde{\mathbf{E}}\left[C_T|\mathcal{F}_t\right] = v_{t,T}N\left[R + (1-\tilde{q}_{t,T})(1-R)\right],$$

where $v_{t,T}$ is the risk-free discount factor and $\tilde{q}_{t,T}$ is the risk neutral default probability associated to the debtor over the interval $(t,T]$.
The model is developed both in discrete and continuous-time cases.

In the discrete-time case, the creditworthiness of the debtor is modeled through a time-homogenous Markov chain $\{\eta_t : 0 \leq t \leq T\}$, defined over a finite state space $S \coloneqq \{1 \ldots K\}$. Each state in S represents a possible rating class (*e.g.*, Aaa, Aa, ...). Hence, the Markov chain is specified by a $K \times K$ transition matrix

$$Q \coloneqq \begin{pmatrix} q_{1,1} & q_{1,2} & \cdots & q_{1,K} \\ \vdots & & & \\ q_{K-1,1} & q_{K-1,2} & \cdots & q_{K-1,K} \\ 0 & 0 & \ldots & 1 \end{pmatrix},$$

in a given real world probability measure P equivalent to $\tilde{P}$. It holds by construction that $q_{ij} \in [0,1]$ and $\sum_j q_{ij} = 1$. Furthermore, the n-step transition matrix $Q_{0,n}$ can be obtained by composition of the uniperiodal transition matrix Q. Namely, it holds

$$Q_{0,n} = Q^n.$$

A form of $\tilde{Q}$ is needed to evaluate the creditworthiness dynamics under $\tilde{P}$. It is assumed that the risk premia adjustments are such that the credit migration process under $\tilde{P}$ satisfies

$$\tilde{q}_{ij}(t,t+1) = \pi_i(t)q_{ij},$$

for all $i \neq j$, where $\pi_i(t)$ is a deterministic function of the time and the state i_t of the debtor, such that $\tilde{q}_{ij}(t,t+1) \geq 0$ for all i,j,t and $\sum_{j:j\neq i} \tilde{q}_{ij}(t,t+1) \leq 1$ for all i,t. In matrix form it holds

$$\tilde{Q}_{t,t+1} - \mathbb{1} = \Pi(t)\left[Q - \mathbb{1}\right].$$

where $\Pi(t) \coloneqq \mathbf{diag}\{\pi_1(t) \ldots \pi_K(t)\}$. The $\Pi(t)$ matrix calibration can be achieved by observing the risky ZCB prices on the markets. In fact, given a debtor in the i-th state at time t, we have

$$\tilde{q}_{iK}(t,t+1) = 1 - \frac{c_{t,t+1} - v_{t,t+1}RN}{v_{t,t+1}(1-R)N},$$

where R is calibrated from the historical default events. Since $\pi_i(t)$ does not depend on the final state of the transition, it can be estimated considering only the probabilities of default inferred from the prices:

$$\pi_i(t) = \frac{\tilde{q}_{iK}(t,t+1)}{q_{iK}}.$$

The model can be extended to a continuous-time framework. In that case, given the same state space S introduced in the discrete-time case, the time-homogenous Markov chain $\{\eta_t : 0 \leq t \leq T\}$ is specified in terms of its $K \times K$ generator matrix

$$\Lambda := \begin{pmatrix} \lambda_{1,1} & \lambda_{1,2} & \dots & \lambda_{1,K} \\ \vdots & & & \\ \lambda_{K-1,1} & \lambda_{K-1,2} & \dots & \lambda_{K-1,K} \\ 0 & 0 & \dots & 0 \end{pmatrix},$$

where $\lambda_{ij} \geq 0$ for all i, j such that $i \neq j$ and $\lambda_{ii} = -\sum_{j:j\neq i} \lambda_{ij}$. In this framework Q can be evaluated over any interval $(t, T]$ by the relation

$$Q(t, T) = \exp\left[(T-t)\Lambda\right].$$

Furthermore, also risk premia adjustments are imposed on the generator Λ instead of considering the transition matrix Q. Namely, it is supposed that the risk-neutral generator has the form

$$\tilde{\Lambda}(t) = U(t)\Lambda,$$

where $U(t) := \mathbf{diag}\{\mu_1(t) \dots \mu_{K-1}(t), 1\}$. The first $K-1$ entries $\mu_1(t) \dots \mu_{K-1}(t)$ are strictly positive deterministic functions that satisfy the requirement $\int_0^T \mu_i(t)dt < +\infty$.

It is shown [72] that the $K \times K$ risk-neutral transition matrix $\tilde{Q}$ is given as the solution to the Kolmogorov equations

$$\partial_t \tilde{Q}(t, T) = -\tilde{\Lambda}(t)\tilde{Q}(t, T), \quad \partial_T \tilde{Q}(t, T) = \tilde{Q}(t, T)\tilde{\Lambda}(T),$$

with the initial condition $\tilde{Q}(t, t) = \mathbb{I}$, allowing for the model application to the continuous-time case.

It is worth noticing that the special case when $\mu_1 \dots \mu_{K-1}$ are positive constants is easily solved:

$$\tilde{Q}(t, T) = \exp\left[(T-t)U\Lambda\right].$$

Calibration is achieved based on the ZCB observed prices, as per the time-discrete case.

Model hypotheses and structure - CreditMetrics

CreditMetrics [73, 74] is a complete framework to assess the risk profile of a given portfolio of risky bonds. In this context, the word “complete” refers to the fact that all the main components of credit risk are addressed. In a nutshell,

- the marginal dynamics of the rating migration (including the transition-to-default event) associated with each considered obligor is described by the rating transition matrix estimated by S&P;
- the recovery rate is modeled as a function of each considered bond seniority;
- the structure of dependence that defines the multivariate dynamics of the considered firms’ creditworthiness is built through a generalization of the Merton model.

The third feature - i.e., the structure of dependence - is the "signature" of the CreditMetrics model.

In Merton model, introduced in §2.2.1, the return-on-asset r_F of a given firm F is assumed to be normally distributed

$$dr_F(t) = \mu_F dt + \sigma_F dW_t$$

and, given a projection horizon $(t_0, T]$, the firm is defaulted in T if $r_F(T|t_0) \leq X_D$, where $X_D \coloneqq \ln \frac{B_T}{V_T}$ is a threshold implied by the underlying microeconomic assumption. In CreditMetrics, the marginal distribution of F's creditworthiness dynamics over $(t_0, T]$ is defined by a given transition matrix Q. Hence, following the same notation used in §2.2.4, we can rewrite the default condition as $r_F(T|t_0) \leq X_K$ where the threshold X_K is redefined as a quantile of the r_F distribution:

$$X_K \coloneqq \Phi_F^{-1}\left(q_{jK}\right).$$

The K-th state represents the firm's bankruptcy, while the j-th state is the rating assumed by F in t_0. The final creditworthiness state assumed by F in T can be generalized from alive/defaulted the whole master scale associated to the considered transition matrix by introducing the partition

$$X_{jk} \coloneqq \Phi_F^{-1}\left(\sum_{k'=k}^{K} q_{jk'}\right), \quad k = 1, \ldots, K;$$

where $X_{jk} > X_{jk+1}$ and $X_{j1} \equiv +\infty$ by definition. Thus, the migration from j-th rating class to the k-th can be described in terms of return over the projection horizon, as

$$X_{jk+1} < r_F(T|t_0) \leq X_{jk}.$$

This change of representation allows to introduce the aforementioned dependence structure, since each $r_F(t)$ dynamics can be observed directly on financial markets or assumed to be equal to an index representing F's economic sector. Hence, the covariance $\mathbf{cov}\left[r_F, r_{F'}\right]$ between two firms F and F' can be estimated and the random vector $\mathbf{r}(T|t_0) \coloneqq (r_{F_1}(T|t_0), r_{F_2}(T|t_0), \ldots)$ can be simulated by using its multivariate normal distribution. Conditioned to the initial rating of each firm, the terminal return-on-asset values $\mathbf{r}(T|t_0)$ can be used together with the partition $\{-\infty, X_{jK}, X_{jK-1}, \ldots, X_{j1}\}$ to generate sets of correlated migration events.

Despite being a generalization of the Merton model, CreditMetrics can still be classified among the reduced-form models. In fact, the marginal probability distribution of migration and default events is defined based on an exogenous transition matrix, without further assumptions on the microeconomic mechanisms that may lead a considered firm to default or relevant creditworthiness variation.

2.2.5 Reduced-form models with default intensity dynamics: Duffie&Singleton approach

In credit risk modeling, the default event is commonly represented as the first occurrence $\tau \coloneqq \inf\{t > 0 : N(t) > 0\}$ of a Poisson process $N(t)|N(0) = 0$. The theoretical justification of this choice is discussed in §3.1.

When a reduced-form credit risk model is specified through an explicit assumption on the default intensity $\lambda_t \in \mathbb{R}_+$, the assumption has a direct consequence on the functional form of the probability of default. In the case of deterministic intensity, the probability of default in $(0, T]$ is expressed as

$$P_{0,T} := \text{Prob}\left(\tau \leq T | \mathcal{F}_0\right) = 1 - \exp\left(\int_0^T \lambda_t dt\right).$$

This expression can be further simplified in the case of time-homogenous intensity. On the other hand, λ_t can be assumed to behave as a (latent) stochastic process, leading to a doubly-stochastic model for the resulting Poisson process (a.k.a. Cox process [72, 129]). This framework has produced a widely populated class of credit risk models[15], depending on the process chosen to describe λ_t.

It is worth remarking that, in principle, these models are strongly related to the rating-migrations models introduced in §2.2.4. In fact, the finite state space $S := \{1, \ldots, K\}$, that defines the "rating" Markov chain, is the codomain of the surjective function

$$S \ni s = \begin{cases} 1 & P_{0,T} \in [0, p_1) \\ 2 & P_{0,T} \in [p_1, p_2) \\ \ldots & \\ K & P_{0,T} = 1 \end{cases}$$

where the K-th state corresponds to the default event that occurred in $t \leq 0$. Hence, the transition matrix generator $\Lambda(t)$ and the SDE that defines the $d\lambda_t$ dynamics have the same role in the two representations.

As a relevant example, in the following, we summarize the main features of the Duffie&Singleton model [130, 131], firstly introduced in 1999.

Model hypotheses and structure

The Duffie&Singleton (DS) model aims to provide a pricing framework both for plain-vanilla corporate bonds and embedded options as well (e.g., callable or puttable bonds). Let us consider a defaultable future cash flow V_T, payable in T, that is discounted in discrete-time framework (*i.e.*, $t \in \mathbb{N}$)

$$V_t = h_t e^{-r_t} \tilde{\mathbf{E}}_t \left[\varphi_{t+1}\right] + \left(1 - h_t\right) e^{-r_t} \tilde{\mathbf{E}}_t \left[V_{t+1}\right]$$

where $\tilde{\mathbf{E}}_s [\cdot]$ is the risk-neutral expectation (conditional on $\mathcal{F}_s$), h_s is the conditional risk-neutral probability of a default in $s+1$, r_s is the default-free short rate and φ_s is the recovery in units of account in case of default in s.

Basically, the DS model is characterized by the so-called "*Recovery of Market Value*" (RMV) hypothesis, that is

$$\tilde{\mathbf{E}}_t \left[\varphi_{t+1}\right] = (1 - L_t) \tilde{\mathbf{E}}_t \left[V_{t+1}\right]$$

where $L_t \in [0, 1]$ is some process representing the fractional loss in case of default. Alternate hypotheses that are also considered in the literature are "*Recovery of Treasury*" (RT) hypothesis

$$\tilde{\mathbf{E}}_t \left[\varphi_{t+1}\right] = (1 - L_t) P_t,$$

[15] See, e.g., [116] for a comprehensive review.

where P_t is a risk-free ZCB with the same notional value and maturity of V_t, and the "*Recovery of Face Value*" (RFV) hypothesis

$$\tilde{\mathbf{E}}_t \left[\varphi_{t+1}\right] = (1 - L_t) V_T.$$

Among the models presented in §2.2, JLT (see §2.2.4) assumes RT hypothesis, while CreditRisk$^+$ (see §2.2.6) is defined under RFV hypothesis.

Duffie and Singleton show that RMV has a relevant implication in terms of analytical tractability of an underlying default intensity process, allowing for the closed-form determination of the default-adjusted short rate R_t. In fact, it holds

$$e^{-R_t} \tilde{\mathbf{E}}_t \left[V_{t+1}\right] = V_t \overset{\text{RMV}}{=} \left[h_t(1 - L_t) + (1 - h_t)\right] e^{-r_t} \tilde{\mathbf{E}}_t \left[V_{t+1}\right]$$

Considering annualized rates but periods of "small" length, we have

$$R_t \overset{\text{RMV}}{=} r_t + L_t h_t.$$

Given the relation among R_t, r_t and h_t, the model can be further specified by choosing the underlying processes. Duffie and Singleton consider, *inter alia*, a multivariate CIR process that allows an explicit dependence between r_t and h_t, together with the assumption of a constant loss given default $L_t = L$.

$$\begin{aligned} r_t &= \delta_0 + \sum_{i=1}^{3} \delta_i Y_{it} \\ s_t &= \gamma_0 + \sum_{i=1}^{2} \gamma_i Y_{it} \\ dY_t &= \mathcal{K} \left(\Theta - Y_t\right) dt + \Sigma \sqrt{S_t} dW_t \end{aligned}$$

where $\mathcal{K}, \Sigma \in \mathbb{R}^{3\times 3}$, $\Theta \in \mathbb{R}^3_+$ and S_t being a multivariate stochastic process whose components are defined as weighted sums of Y_{1t} and Y_{2t}. This choice and even more complex ones (e.g. considering a jump-diffusion process to model s_t) are considered by Duffie and Singleton to investigate flexibility and tractability of the DS model applied to derivatives pricing. However, the same DS framework can be specified also considering simpler hypotheses, as done in [132] for Solvency 2 application purposes.

2.2.6 Reduced-form portfolio models of default events: CreditRisk$^+$ and Vasicek models

In this section, two classical credit portfolio models are described: CreditRisk$^+$ and the Vasicek models. These models, both developed in the 90s, share two relevant feature:

- rating migration and price variation due to creditworthiness dynamics are not modeled - the model explicitly describes only the default event;
- the dependence structure is defined by introducing latent variables representing the market experienced by the modeled debtors.

The CreditRisk$^+$ model is a portfolio model developed by Credit Suisse First Boston (CSFB) by Tom Wilde [133] and coworkers, first documented in [134] and later widely discussed in [135]. It is a model *actuarially inspired* in the sense that losses are due only to default events and not to other sources of financial risk, *e.g.* variation of the credit standing (the so-called "credit migration" effect). CreditRisk$^+$ can be classified as a *frequency-severity model*, cast in a single-period framework, with the peculiarity that a doubly stochastic process describes the frequency of default events while loss severity is deterministic. The second hypothesis can be easily relaxed at the cost of some additional computational burden. However, this issue can be neglected for what follows.

The structure of dependence of default events is described using a factor model framework, where factors are unobservable (*i.e.* latent) stochastic "market" variables, whose precise financial/actuarial identification is irrelevant since the model integrates on all possible realizations ("market scenarios"). Therefore, CreditRisk$^+$ can be further classified into the family of *factor models* and, in particular, into the sub-family of *conditionally independent* factor models, since, conditionally on the values assumed by the factors, defaults are supposed (by the model) to be independent.

The Vasicek model, disclosed in 2002 [77,80], was initially developed by O.A. Vasicek between the late 80s and the 90s [75, 76] in his research activity for KMV – the financial society mentioned in §2.2.1, where also the Moody's KMV model was developed from the Merton model. Indeed, the Vasicek model can be considered a portfolio generalization of the Merton model. However, in the Vasicek model case, PDs are exogenous parameters while the dynamics of the firms' creditworthiness is built by using a latent, systemic variable (the "market" state) that is compatible with the Merton hypothesis but still does not need any balance sheet information from the debtors to be calibrated. This feature allows classifying the Vasicek model into the same sub-family of the CreditRisk$^+$ model. The model is summarized below, considering the simplifications applied by BSCS [78–80] in developing the Basel II regulatory framework for the banking system.

Model hypotheses and structure - CreditRisk$^+$

The structure of the model can be summarised as follows. Let N be the number of different risks in a given portfolio and $\mathbb{1}_i$ the default indicator function of the i-th risk ($i = 1, \dots, N$) over the time horizon $(t, T]$. The indicator function $\mathbb{1}_i$ is a Bernoulli random variable that takes the value 1 in case of default with probability q_i and the value 0 with probability $1 - q_i$. Thus:

$$\mathbf{E}\left[\mathbb{1}_i\right] = q_i, \quad \mathbf{cov}\left[\mathbb{1}_i\right] = q_i(1 - q_i), \quad i = 1, \dots, N.$$

The "portfolio loss" L over the reference time horizon $(t, T]$ is then given by:

$$L = \sum_{i=1}^{N} \mathbb{1}_i \, E_i, \quad E_i = (EAD)_i \, (LGD)_i,$$

where $(EAD)_i$ and $(LGD)_i$ are respectively the *Exposure At Default* and the *Loss Given Default* of the i-th risk.

To ease the semi-analytic computation of the distribution of L, the model introduces a new set of variables Y_i, each replacing the corresponding indicator function $\mathbb{1}_i$ ($i = 1, \ldots, N$). The new variables Y_i are supposed to be Poisson-distributed, conditionally on the value assumed by the market latent variables.

Let K be the number of latent variables and $\mathbf{\Gamma} = (\Gamma_1, \ldots, \Gamma_K)$ the K-dimensional vector describing the "market". The latent variables are assumed to be independent and gamma-distributed, where shape and scale parameters are notated as α_k and β_k respectively ($k = 1 \ldots K$). Without loss of generality, it can be further assumed that $\mu_1 = \cdots = \mu_{\mathrm{K}} = 1$, so that $\alpha_k = \beta_k^{-1}$ and $\beta_k = \sigma_k^2$.

Conditionally on Γ, the parameter of the Poisson distribution of Y_i is assumed to be:

$$p_i(\mathbf{\Gamma}) = q_i \cdot \left(\omega_{i0} + \sum_{k=1}^{K} \omega_{ik} \Gamma_k \right),$$

where the *factor loadings* ω_{ik} are all non-negative and sum up to unity:

$$\sum_{k=0}^{K} \omega_{ik} = 1, \quad \omega_{ik} \geq 0, \quad i = 1, \ldots, N, \quad k = 0, \ldots, K,$$

so that q_i is the unconditional expected default frequency:

$$q_i = \mathbf{E}\left[p_i(\mathbf{\Gamma})\right] = \int_0^\infty \cdots \int_0^\infty p_i(\mathbf{\Gamma})\, f\left(\mathbf{\Gamma}\right) d\Gamma_1 \ldots d\Gamma_{\mathrm{K}},$$

and the identity between the expected values of the original Bernoulli variable $\mathbb{1}_i$ and the new Poisson variable Y_i is granted, *i.e.* $\mathbf{E}\left[Y_i\right] = \mathbf{E}\left[\mathbb{1}_i\right] = q_i$. According to the above hypothesis, the portfolio loss is now given by L_{Y}:

$$L_{\mathrm{Y}} = \sum_{i=1}^{N} Y_i \cdot E_i, \quad \text{where} \quad Y_i | \mathbf{\Gamma} \sim \text{Poisson}\left(p_i(\mathbf{\Gamma})\right).$$

In [134] it is shown how to compute the distribution of L_{Y} using a recursive method known as "Panjer recursion", [107]. The accuracy, stability and possible variants of the original algorithm are discussed in [135]. Numerical computation of the distribution by mean of Monte Carlo simulation is also possible and turns out to be particularly simple. Importance sampling algorithms are also available in the literature [136].

Model hypotheses and structure - Vasicek model

As anticipated, the Vasicek model is based on the Merton model hypothesis: each firm's asset value V_t can be described through a geometric Brownian motion and the default event occurs when $V_t \leq B$, where B is the debt level of the firm.

The signature of the Vasicek model is the assumption that generator of each firm's dynamics is correlated with the other generators through the same latent systemic factor Z_t:

$$\begin{aligned} dV_t^{(F)} &= \mu V_t^{(F)} dt + \sigma_V V_t^{(F)} dX_t^{(F)}, \\ X_t^{(F)} &= \sqrt{1-\rho} W_t^{(F)} + \sqrt{\rho} Z_t \end{aligned}$$

for each considered firm $F \in \mathbb{N}$, where $dW_t^{(F)}$ and dZ_t are independent Wiener processes. Thus, Z_t defines the dependence structure of the model and ρ is the correlation between the returns of each considered couple of firms F,F'.

Similarly to CreditMetrics, the Vasicek model is reduced-form, despite being based on the structural hypothesis firstly introduced in the Merton model. This is because the unconditional probability of default $p_F := \text{Prob}\,(V_t \leq B_F)$ is provided exogenously, implying that

$$\begin{aligned}
\text{Prob}\,(V_t \leq B_F|\, Z_t) &= \text{Prob}\left(X_t \leq \Phi^{-1}(p_F)\Big|\, Z_t\right) \\
&= \text{Prob}\left(\sqrt{1-\rho}W_t^{(F)} + \sqrt{\rho}Z_t \leq \Phi^{-1}(p_F)\right) \\
&= \text{Prob}\left(W_t^{(F)} \leq \frac{\Phi^{-1}(p_F) - \sqrt{\rho}Z_t}{\sqrt{1-\rho}}\right) \\
&= \Phi\left(\frac{\Phi^{-1}(p_F) - \sqrt{\rho}Z_t}{\sqrt{1-\rho}}\right).
\end{aligned}$$

Given that $Z_t \sim \mathcal{N}(0,1)$, this result allows to quantify the default probability of F in a "worst case" scenario of the systemic factor Z_t, at a given α confidence level. This is the case in the Basel II banking regulatory framework, where the unexpected loss UL of an AFG portfolio[16] is calculated as

$$UL \propto LGD \cdot \left[\Phi\left(\frac{\Phi^{-1}(p) - \sqrt{\rho}\Phi^{-1}(0.999)}{\sqrt{1-\rho}}\right) - p\right]$$

where $\alpha = 0.999$ and the proportionality coefficient depends on the risks type (*e.g.*, corporate exposures, residential mortgages). Additional hypotheses are considered regarding the homogeneity of risks: $p_F = p$ for each F and the dependence structure is the same as well since all the risk sources depend on the same risk factor Z_t through the same correlation parameter ρ.

2.2.7 Feasibility of the presented models for C&S applications

In §2.2.1−2.2.6 a selection of credit risk models is presented. The chosen examples cover all the main modeling approaches typically considered in credit risk theory, although far from representing (and not aiming to be) an exhaustive list of the models developed in the field.

Based on the models discussed so far, it is now possible to evaluate to what extent each one among the considered frameworks fits C&S modeling applications.

Structural approaches *a lá* Merton are necessarily ruled out. Both the Merton model and the KMV model are based on stock prices dynamics. However, risky buyers underlying a credit insurance policy typically belong to the SME segment and, thus, are likely not to be listed in a stock market. On the other hand, in case the risky subject is a firm belonging to the large corporate segment, other forms

[16]The "Asymptotically Fine-Grained Portfolio" (a.k.a. AFGP) hypothesis assumes that a bank can effectively describe its portfolio of risks as a large (*i.e.*, $N \to \infty$ elements) set of small risks (*i.e.*, the exposure generated by each risk is approximately equal to the others and tends to zero).

of protection from counterparty risk are typically available (*i.e.*, financial instead insurance products, such as CDSs). That said, the case of a listed corporate firm underlying a credit insurance policy is perfectly possible. However, it is reasonable to assume that only a small fraction (if any) of the buyers covered by a credit insurance company fall into this category. In the case of suretyship insurance, the underlying risk is not necessarily related to credit risk. This makes all the structural models inadequate to be applied since they explicitly link the probability of a claim to a default event.

However, among structural models, cross-sectional approaches can effectively quantify the marginal probability of a credit insurance claim that is needed in a reduced-form portfolio model. Although not being listed in the stock market, the majority of the risky buyers have some idiosyncratic information available to the insurance company, that can be considered in a scoring model. Even without an updated financial statement available, pieces of information such as the number of employees, years since foundation, and past liquidity crises (*e.g.*, observed by the insurer through past claims) are usually available to the insurance company and make a scoring model feasible. Once again, suretyship insurance may be harder to be modeled by using a cross-sectional approach when the loss event is not related to insolvency. However, if the suretyship insurance company has a database depth and rich enough, that kind of technique is applicable in principle. The diffusion of this kind of approach across C&S company is suggested, *inter alia*, by the technical specifications of the second Non-Life Comparative Study (NLCS) [81], promoted by EIOPA, where C&S insurance S2LoB has been included for the first time. In template A, EIOPA asks for a segmentation of both credit and suretyship portfolios considering a 10 notch master's scale that is in line with the conventions typically followed in the sector. The classification of risky subjects on the same master scale largely copes with the application of cross-sectional techniques. It is worth noticing that this kind of framework needs an "anchor point" probability level to be properly calibrated. This fact confirms the relevance of estimating "raw" claim frequencies on risk-homogenous clusters in C&S applications. The problem will be addressed in the next chapters 3 and 6.

Econometric models as CreditPortfolioView can be considered as well to describe claim probability in C&S insurance. However, there is a strong requirement to be satisfied: a set of macroeconomic variables $\{X_j(t)\}$ must exists, such that $\text{logit}\,(PD_i(t))$ can be modeled as a linear function of the $\{X_j(t)\}$ elements, whose parameters are independent from time. Although perfectly possible in principle, the existence of such a set seems hardly likely both for a credit insurance portfolio, given that underlying risks are mainly PMI, and for a suretyship portfolio, considering that a set $\{X_j(t)\}$ should exist for each guaranteed loss event type.

Although ratings and rating models can be used in a C&S context, an explicit representation of rating migrations is not needed in most cases. In fact, neither in credit insurance nor in suretyship insurance, a decrease in creditworthiness generates a loss to the insurer, as confirmed by the results reported later in chapter 5. This kind of model is obviously relevant in the market context where it has been originally developed (*i.e.*, obligations), where rating migrations imply steep variations of price and subsequent profit or loss on a bond that can instantly be sold at the market price. This is not the case in C&S insurance sector, where both policy pricing and management actions (when applicable - e.g., credit limit variations) are evaluated based on the current claim probability associated with the subject. This argument can be easily extended to the reduced-form models based on Cox processes: in C&S context, there is no need to explicitly model a future state of a risk source apart from the binary set *claim-generated-or-not*, because these are the only two alternate scenarios that have a different impact on the insurer's financial statement.

Despite the lack of interest for an explicit representation of creditworthiness dynamics, the idea of a latent stochastic process describing the probability-of-claim dynamics of a risky subject can be relevant to risk management applications (*e.g.*, in case the C&S premium risk is measured by an internal model in the Solvency 2 framework). In fact, although not affecting the losses distribution directly, introducing a stochastic probability of claim (*i.e.* a doubly stochastic process) may increase the kurtosis of the future losses distribution, leading to a more prudential estimation of the capital requirement. This implies that latent-factors models, such as CreditRisk$^+$ and the Vasicek model, perfectly fit the features of C&S line of business. Being reduced-form models, they are compatible with scoring models, even when the latters are applied only to some of the covered risk sources.

It is worth noticing that also the doubly stochastic models that provide an explicit description of the creditworthiness dynamics are applicable. However, their output is typically redundant for the needs of a C&S insurer. On the other hand, their calibration is a more demanding task (*i.e.* a greater amount of information is required) than the calibration of a latent-factors model.

Based on the considerations above and without pretense of completeness, we choose to consider the CreditRisk$^+$ model as a tool to model the joint claim probabilities in a C&S insurance portfolio. The problem of calibrating this model facing the information limits typical in the C&S context is addressed in chapter 4.

2.3 C&S premium and catastrophe risks in Solvency 2

In the Solvency 2 framework, the quantification of S2LoB 9 unexpected losses originated by future claims follows the same structure adopted in the Standard Formula to model every other non-life line of business.

- The *premium risk* component is quantified as a whole, considering three times the product of a given volatility parameter $\sigma_{\text{prem, s}}$ and a volume measure, where s is the considered segment number.
- Specific extreme scenarios are considered in the *catastrophe risk* component. This module applies to all the non-life LoBs (including C&S) where the prudentiality provided by the premium risk measure is considered insufficient.
- Standard Formula considers a so-called rule-based aggregation approach [122], where the same rule is used to aggregate S2LoB 9 premium risk and catastrophe risk contributions to SCR, as well as the risk contributions originated from other risk types (*e.g.* reserve risk) and other non-life LoBs.

The simplified structure of the Standard Formula does not encompass an explicit representation of C&S claim probabilities. Frequency and severity components of C&S future claims are modeled as a whole in premium risk, whose contribution to SCR is proportional to a volume measure based on premiums volume[17].

However, the presence of a man-made catastrophe risk sub-module dedicated to this LoB sheds light on the EIOPA's perspective about C&S business.

Indeed, the "*Credit and suretyship risk*" sub-module[18] is composed by two independent[19] contributions:

- *Default risk*, that is quantified as the loss arising in case of instantaneous default of the two largest exposures in Segment 6 of the insurance undertaking, after the effect of recoverable amounts from reinsurance contracts or SPVs, given the assumption that the gross loss-given-default associated to the event is equal to 10%;
- *Recession risk*, that is quantified as the effect of an instantaneous loss of an amount that, without deduction of the amounts recoverable from reinsurance contracts and SPVs, is equal to 100% of the premiums earned by the insurance undertaking during the following 12 months in Segment 6.

Hence, EIOPA identifies these two types of loss events as not adequately measured by the standard premium risk capital requirement, despite the high quantile level considered in Standard Formula[20].

Default risk has a plain interpretation as a penalty for the concentration of exposures on specific subjects, despite their creditworthiness. On the other hand, recession

[17] Solvency 2 "Commission Delegated Regulation" 2015/35 [37], Chapter V, Section 2, Article 116, paragraph 3 reads: "*For all segments set out in Annex II, the volume measure for premium risk of a particular segment s shall be equal to the following:*

$$V_{(prem,s)} = \max\left[P_s; P_{(last,s)}\right] + FP_{(existing,s)} + FP_{(future,s)}$$

where:

(a) P_s denotes an estimate of the premiums to be earned by the insurance or reinsurance undertaking in the segment s during the following 12 months;

(b) $P_{(last,s)}$ denotes the premiums earned by the insurance or reinsurance undertaking in the segment s during the last 12 months;

(c) $FP_{(existing,s)}$ denotes the expected present value of premiums to be earned by the insurance or reinsurance undertaking in the segment s after the following 12 months for existing contracts;

(d) $FP_{(future,s)}$ denotes the expected present value of premiums to be earned by the insurance and reinsurance undertaking in the segment s for contracts where the initial recognition date falls in the following 12 months but excluding the premiums to be earned during the 12 months after the initial recognition date."

C&S insurance (S9LoB 9) is a part of Segment 6 "credit and suretyship insurance and proportional reinsurance", together with S2LoB 21.

[18] See Solvency 2 "Commission Delegated Regulation" 2015/35 [38], Chapter V, Section 2, Article 134: "*Credit and suretyship risk sub-module*".

[19] Default risk capital contribution and recession risk contribution are aggregated by a plain root sum squared, implying a zero correlation coefficient.

[20] In the Solvency 2 framework, the SCR contribution originated by each sub-module represents the 0.995-th quantile of the unexpected losses distribution measured by the sub-module.

risk has a nontrivial meaning since it has been initially designed as a counter-cyclical measure.

EIOPA clarifies the original purpose of this submodule in the "Solvency II Calibration Paper" (see [83], §3.1153 p. 312), where a counter-cyclical dampening mechanism has been introduced - to be abandoned later in the 2015 release of Standard Formula technical specifications [37]:

> "[...] *The dampening mechanism is limited to a* $SCR_{CAT\_recession\_ratio\_net}$ *of 200% of the net earned premium with a net loss ratio lower than 25% and to a* $SCR_{CAT\_recession\_ratio\_net}$ *of 100% of the net earned premium with a net loss ratio higher than 125%. Within the limits the* $SCR_{CAT\_recession\_ratio\_net}$ *= 225% minus net loss ratio. This mechanism aims to ensure that at the peak of the cycle (low failure rates), the* $SCR_{CAT\_recession\_ratio\_net}$ *shall reach its highest value and C&S undertakings shall be required to have enough own funds to cover a higher SCR. On the other hand, at the trough of the cycle, SCR will be at its lowest value, so that own funds will be released. In other words, as undertakings face harder net claims ratio due to an increase of failure rates, the SCR decreases.*"

However, the original idea of recession risk as a dampening mechanism was abandoned in the final version of the Standard Formula. Furthermore, the event type considered in C&S catastrophe risk is the same as per premium risk. Hence, double-counting is possible if not prevented in calibration. Indeed, this is the view expressed by ICISA [99] as an answer to EIOPA's "second set of advice to the European Commission on specific items in the Solvency II Delegated Regulation" [20]:

> "*Calibration was done without exclusion of outliers in terms of years (only outliers in terms of undertakings have been excluded). ICISA members disagree on the a priori consideration that catastrophe events are not expected to have a major impact on the results: 2001 and 2008 crises show well that recession events are already part of our history, leading to a double-counting phenomenon in the standard formula for the LoB CS. Parameters would have been more adequately calibrated with an exclusion of these exceptional years already treated in the Cat recession module of the SF.*"

The final decision of EIOPA has been to maintain the Recession Risk sub-module unaltered after the feedback. Nonetheless, the Premium Risk volatility parameter associated with Segment 6 has been lowered in the 2019 revision of Solvency II Delegated Regulation [38].

Regardless of the technical choices made by EIOPA in defining default risk and recession risk in Standard Formula, the introduction of these two elements highlights the importance of estimating both the marginal claim probability of highly unlikely events in C&S LoB (*i.e.* default risk) and the joint claim probability across the C&S risks underwritten by the insurance undertaking, the latter needing to be modeled and calibrated in a way that encompasses systemic crises (*i.e.* recession risk).

Chapter 3

Probability estimation of absorbing events

In Chapter 1, Credit and Suretyship insurance line of business has been introduced. In Chapter 2, similarities and differences between a credit insurance policy and a suretyship contract are outlined. In particular, in section 2.1 it is discussed how a claim covered by a C&S policy - considering either a credit or a suretyship product - can be represented as an absorbing event. Such an event may be effectively described by using some of the tools and the methods initially developed to measure the financial credit risk, as shown in section 2.2.

Now, it is worth recalling the main properties of absorbing events, together with the classical inference techniques that are commonly applied to estimate their probability of occurrence. This is the aim and scope of Chapter 3. The chapter is organized as follows.

Section 3.1 recalls the concept of absorbing events, together with the basic definitions and results useful in the following.

Section 3.2 introduces the inference problem of measuring the probability of an absorbing event from historical data. The presented solutions are reached under the assumption of having access to complete historical information.

In section 3.3, the estimation error that affects the measure of an absorbing event probability is quantified. This result is then applied in section 3.4, which discusses how to handle eventual lack of information and approaches the inference problem with partial access to the historical data. While section 2.2 considers theoretical results that have been developed for financial applications, in this case a selection of classical biostatistics techniques is presented. Indeed, human mortality measurement is the context where the methods to infer the probability of absorbing events were developed when the historical information needed is only partially available. That is the case also in credit insurance.

Finally, section 3.5 discusses to what extent the techniques presented in the previous sections come in handy when estimating the probability of a C&S claim. Concepts and methods introduced in this chapter will be applied in the following chapters 4−6, where the estimation of C&S claim probabilities is investigated.

3.1 Absorbing events

Loosely speaking, an absorbing event is defined by its possibility to occur once at most across a given time horizon $(t_0, T]$. The concept of absorbing event appears in many contexts and, usually, it is not associated with nice facts. Human death, bankruptcy, and breach of undertaking are a few relevant examples.

Following the usual notation, let $\{\Omega, \mathcal{F}, \text{Prob}\}$ be a probability space, $\{S, \Sigma\}$ a measurable space and $\{\eta_t : \Omega \to S\}_{t_0 \le t \le T}$ a stochastic process defined on the state space S and the time horizon $(t_0, T]$.

The word "absorbing" is taken from the semantics of Markov chains. Indeed, being η_t a Markov process, a state $a \in S$ is absorbing if and only if

$$\text{Prob}\left(\eta_{t'>t} \neq a | \eta_t = a\right) = 0$$

The Markov chain itself is said to be "absorbing" iff every state $s \in S$ is connected with an absorbing state $a_s \in A \subseteq S$.

$$\forall s \in S \lim_{\Delta_t \to +\infty} \text{Prob}\left(\eta_{t+\Delta_t} \in A | \eta_t = s\right) = 1$$

where $A \subseteq S$ is the subset of absorbing states.

Although the "absorbing state" concept is commonly introduced in the Markov chains context, it is worth noticing that the definition of absorbing state does not need the memoryless property of Markov chains. In fact, the absorbing state definition requires only that the probability distribution of η_t depends on η_{t^-} *at least*. Indeed, we just want that the process "knows", in t, whether its state in t^- belongs to A or not, regardless of its memory length.

On these premises, an absorbing event $(a, \tau) \in A \times (t_0, T]$ satisfies the following

$$(a, \tau) : \eta_{\tau^-} = s \in S \setminus A \wedge \eta_\tau = a \in A$$

Namely, an absorbing event is the transition of a stochastic process (not necessarily a Markovian one) to an absorbing state.

Considering most financial and actuarial applications, including C&S products, we need to describe a specific absorbing event "a" only - that is the one covered by the contract. In the remainder of this section, such an event will be named "default" for the sake of brevity. However, a C&S contract may guarantee a wide variety of absorbing events, as seen in chapter 1. Hence, the following Bernoulli r.v.'s are introduced:

$$d_{t,t'} := \mathbb{I}_{\{\tau \in (t,t']\}},$$

representing the event "default occurs in $(t, t']$", and

$$\text{l}_t := \mathbb{I}_{\{\tau > t\}},$$

which stands for the event "default does not occur up to t", where $t_0 \le t < t' \le T$.

Further, let us introduce the following two valuable quantities: the cumulative default probability

$$P_{t_0,t} := \text{Prob}\left(\tau \le t | \tau > t_0\right)$$

and the hazard rate

$$h_t := \lim_{t'-t \to 0^+} \frac{1}{t'-t} \text{Prob}\left(\tau \in (t,t'] | \tau > t\right).$$

A relevant implication of the "absorbing event" definition is that $P_{t_0,t}$ can be expressed as a function of h_t.

Indeed, considering the *tower rule* of expectations, we can write

$$\begin{aligned} \mathbf{E}\left[d_{t,t'} | \tau > t_0\right] &= \mathbf{E}\left[\mathbf{E}\left[d_{t,t'} | \mathrm{l}_t\right] | \tau > t_0\right] \\ &= \mathbf{E}\left[d_{t,t'} | \mathrm{l}_t = 1\right] \text{Prob}\left(\mathrm{l}_t = 1 | \tau > t_0\right) \\ &+ \underbrace{\mathbf{E}\left[d_{t,t'} | \mathrm{l}_t = 0\right]}_{=0} \text{Prob}\left(\mathrm{l}_t = 0 | \tau > t_0\right) \end{aligned}$$

It holds $\mathbf{E}\left[d_{t,t'} | \mathrm{l}_t = 0\right] = 0$ by definition, since τ is the stopping time of an absorbing event. Thus we have

$$\mathbf{E}\left[d_{t,t'} | \tau > t_0\right] = \mathbf{E}\left[d_{t,t'} | \mathrm{l}_t = 1\right] \text{Prob}\left(\mathrm{l}_t = 1 | \tau > t_0\right) \tag{3.1}$$

It holds

$$\mathbf{E}\left[d_{t,t'} | \mathrm{l}_t = 1\right] = \text{Prob}\left(\tau \in (t,t'] | \tau > t\right) \tag{3.2}$$

by definition of $d_{t,t'}$ and

$$\mathbf{E}\left[d_{t,t'} | \tau > t_0\right] = P_{t_0,t'} - P_{t_0,t}, \tag{3.3}$$

$$\text{Prob}\left(\mathrm{l}_t = 1 | \tau > t_0\right) = 1 - P_{t_0,t} \tag{3.4}$$

by definition of absorbing event. Hence, the application of equations (3.2−3.4) to equation (3.1) leads to

$$P_{t_0,t'} - P_{t_0,t} = \text{Prob}\left(\tau \in (t,t'] | \tau > t\right)(1 - P_{t_0,t}) \tag{3.5}$$

By multiplying both the sides of the equation above by $\frac{1}{t'-t}$, in the limit $t'-t \to 0^+$ we have

$$\partial_t P_{t_0,t} = h_t (1 - P_{t_0,t})$$

that implies

$$P_{t_0,t} = 1 - \exp\left(-\int_{t_0}^{t} h_x dx\right)$$

The result above is universally true (*i.e.* model-free) when considering an absorbing event. Thus, any well-posed model aiming to describe the temporal evolution of P_t may be expressed by defining the dynamics of the locally conditioned probability density h_t.

The hazard rate h_t can be either a deterministic function of t (*e.g.* a constant) or a stochastic process as well.

It is worth introducing the survival function $S_{t_0,t}$

$$S_{t_0,t} := \text{Prob}\left(\tau > t | \tau > t_0\right)$$

By definition it holds $S_{t_0,t} = 1 - P_{t_0,t}$, hence

$$S_{t_0,t} = \exp\left(-\int_{t_0}^{t} h_x dx\right).$$

This is a noteworthy result, with applications ranging from finance to medicine. For example, it shows that the risky discount factor $v^{\star}_{t_0,T}$ and the risk-free discount factor $v_{t_0,T}$ share the same functional form

$$v^{\star}_{t_0,T} := \mathbf{E}\left[e^{-\int_{t_0}^{T} r_t + h_t dt} \middle| \mathcal{F}_{t_0}\right]; \quad v_{t_0,T} := \mathbf{E}\left[e^{-\int_{t_0}^{T} r_t dt} \middle| \mathcal{F}_{t_0}\right],$$

where r_t is the instantaneous spot risk-free rate.

3.2 Inference under a complete historical information

Now that absorbing events and their probability of occurrence $P_{t,t'}$ have been introduced, we need to estimate $P_{t,t'}$ based on historical observations to use these concepts for practical purposes. In this section we discuss how to do it, considering some convenient assumptions that will be weakened in the next sections 3.3 and 3.4.

When observing a given subject (hereafter referred to as "risk source" or "debtor"), that may generate a specific absorbing event (*e.g.* "default"), multiple observations of that event are forbidden by definition. Hence, we need to observe a set $\{j\}$ of indexed risky sources (e.g., a portfolio of risky debtors) in order to collect enough data to infer the $P_{t,t'}$ value with adequate precision.

As a toy example, let us consider a portfolio composed of personal loans, each of them taken out by a different debtor $j \in \{j\}$. At the time t, each debtor is paying the due installments on a regular basis. Thus, the number $\overline{\mathrm{L}}_t$ of "alive" (*i.e.* not yet defaulted) debtors up to t is equal to the portfolio size at t:

$$\overline{\mathrm{L}}_t := \sum_{j\in\{j\}} \mathrm{l}_t^{(j)} | \mathcal{F}_t = |\{j\}|$$

On the other hand, the number of debtors in $\{j\}$ that will default in $(t, t']$ is represented by the r.v.

$$\mathrm{D}_{t,t'} := \sum_{j\in\{j\}} d_{t,t'}^{(j)} | \mathcal{F}_t.$$

After Δ_t (*e.g.* one year), at $t' = t + \Delta_t$, some debtors are insolvent:

$$\overline{\mathrm{D}}_{t,t'} := \sum_{j\in\{j\}} d_{t,t'}^{(j)} | \mathcal{F}_{t'}.$$

In this context, the empirical default frequency

$$f_{t,t'} := \frac{\overline{\mathrm{D}}_{t,t'}}{\overline{\mathrm{L}}_t},$$

measured by t', is an intuitive way to estimate the real-world default probability

$$P_{t_0,t_0+\Delta_t} \coloneqq \mathbf{E}\left[d_{t_0,t_0+\Delta_t}|\mathcal{F}_{t_0}\right]$$

over a generic time interval $(t_0, t_0 + \Delta_t]$ for any debtor $j \in \{j\}$, alive by t_0. In the following, we discuss which assumptions are needed to use observed frequency as a proper estimator of probability (*i.e.* $\hat{P}_{t,t'} = f_{t,t'}$).

First, we introduce a three-time scheme that may be used to describe a typical real-life absorbing event, in the perspective of an observer that aims to estimate the event frequency. This scheme will come in handy later, when the frequency estimation will be affected by censoring events, in section 3.4.

τ_j^0 : the transition to the absorbing state takes place.

τ_j^1 : there is a measurable effect of the event occurred in τ_i^0.

τ_j^2 : the observer finds out that the absorbing event occurred.

It is true by definition that

$$\tau_i^0 \leq \tau_i^1 \leq \tau_i^2$$

The scheme fits well the considered "loans" toy example. In fact

τ_j^0 stands for the latent time when the j-th debtor becomes insolvent (*e.g.* j loses his job) and thus will not be able to pay the next installment;

τ_j^1 is the due date of the first unpaid installment (*i.e.* the first measurable effect of the absorbing event in the bank's perspective);

τ_j^2 is the time when the bank front office notifies to the risk management department that j did not pay an installment (*i.e.* the subject who actually measures the probability of default associated with that loans portfolio "observes" the event).

Table 3.1 summarizes this and other examples of absorbing events described by applying the three-time scheme introduced in this section.

Intuitively, the shorter the time intervals $[\tau_j^0, \tau_j^1]$ and $[\tau_j^1, \tau_j^2]$ are, the more likely it is that nothing happens in the meanwhile, that may forbid the observation in τ_j^2 of the absorbing event occurred in τ_j^0. The case when information may be lost between τ_j^0 and τ_j^2 is discussed later in section 3.4. On the other hand, in the following, we assume that every absorbing event is perfectly accessible by the observer.

Assumption 3.1 (Full observability)**.** *Each occurred absorbing event is immediately observed across the whole considered time horizon* $(t_0, t_0 + \Delta_t]$ *where the observation takes place:*

$$\tau_j^0 = \tau_j^1 = \tau_j^2 =: \tau_j$$

for each $\tau_j^0 \in (t_0, t_0 + \Delta_t]$.

Retail banking: $\boxed{\tau_j^0 \lessapprox \tau_j^1 = \tau_j^2}$	τ_j^0 a borrower loses his job; τ_j^1 an installment expires; τ_j^2 the bank is aware that the borrower is insolvent.
Oncology: $\boxed{\tau_j^0 \simeq \tau_j^1 < \tau_j^2}$	τ_j^0 an experimental therapy fails; τ_j^1 the patient dies; τ_j^2 researchers record the death event.
Suretyship: $\boxed{\tau_j^0 < \tau_j^1 \simeq \tau_j^2}$	τ_j^0 a technical issue undermines the timing of a project; τ_j^1 the builing is still in progress at the delivery date; τ_j^2 the contracting authority notifies the surety.
Trade credit insurance: $\boxed{\tau_j^0 < \tau_j^1 < \tau_j^2}$	τ_j^0 a risky buyer experiences a liquidity distress; τ_j^1 an invoices expires and the buyer is not able to pay; τ_j^2 the insured seller notifies the insurer after a first attempt of recovery *in bonis*.

Table 3.1. Examples of absorbing events described by the three-time scheme introduced in section 3.2.

Remark 3.1. *Assumption 3.1 implies that $d_{t,t'}^{(j)}$ is well defined and can be used to represent the absorbing event associated with the j-th risk source without further specification. On the other hand, in case assumption 3.1 does not hold, it is necessary to specify which is the time variable belonging to $(t, t']$ among τ_j^0, τ_j^1 and τ_j^2. Further, in that case, the possibility of absorbing events occurred but not observed introduces some issues in inferring $P_{t,t'}$ from an incomplete set of observations. These issues will be addressed in section 3.4 and chapter 6.*

In the considered toy example, assuming frequent enough installments (*e.g.* monthly) and an instantanous access of all the banks departments (*i.e.* including risk management - the observer) to the information regarding the unpaid installments, assumption 3.1 holds, as summarized in table 3.1 - case "retail banking".

Further, we can consider the following two assumptions.

Assumption 3.2 (Identically distributed r.v.'s)**.** *Each absorbing event has the same* *probability to take place.*

$$Prob\left(\tau_i \le t' | \tau_i > t\right) = Prob\left(\tau_j \le t' | \tau_j > t\right) = P_{t,t'}$$

for each $i, j \in \{j\}$ *s.t.* $i \neq j$.

Assumption 3.3 (Independency). *The considered risks are (conditionally) independent.*

$$Prob\left(\tau_i \le t' \wedge \tau_j \le t' | \mathcal{F}_t\right) = Prob\left(\tau_i \le t' | \mathcal{F}_t\right) Prob\left(\tau_j \le t' | \mathcal{F}_t\right)$$

for each $i, j \in \{j\}$ *s.t.* $i \neq j$.

In the following, $f_{t,t'}$ is shown to be a proper estimator of $P_{t,t'}$. The set of assumptions 3.1−3.3 is progressively weakened, to discuss to what extent the measured frequency $f_{t,t'}$ can be used to quantify the probability $P_{t,t'}$.

Proposition 3.1 ($f_{t,t'}$ as the ML estimator of $P_{t,t'}$). *Given assumptions 3.1-3.3,* $\hat{P}_{t,t'} = f_{t,t'}$ *is the maximum likelihood estimator of* $P_{t,t'}$. *Further,* $\hat{P}_{t,t'}$ *is unbiased.*

Proof. Assumption 3.1 implies that the j-th risk is well represented by the r.v. $d^{(j)}_{t,t'}$. Hence, assumptions 3.2 and 3.3 imply the following form of the likelihood function $\mathcal{L}$

$$\mathcal{L} = P_{t,t'}^{\overline{\mathrm{D}}_{t,t'}} \left(1 - P_{t,t'}\right)^{\left(\overline{\mathrm{L}}_t - \overline{\mathrm{D}}_{t,t'}\right)}$$

It holds

$$\partial_{P_{t,t'}} \ln \mathcal{L} = 0|_{P_{t,t'} = \hat{P}_{t,t'}} = 0,$$

and thus we have

$$\hat{P}_{t,t'} = \frac{\overline{\mathrm{D}}_{t,t'}}{\overline{\mathrm{L}}_t} = f_{t,t'}.$$

For the sake of brevity, the fact that $\hat{P}_{t,t'}$ is unbiased is shown in the proof of proposition 3.2, under a subset of the assumptions considered here. ■

Now, let us remove assumption 3.3: once again, $f_{t,t'}$ can be used as an estimator of $P_{t,t'}$, although not being an ML estimator anymore.

Proposition 3.2 ($f_{t,t'}$ as the MM estimator of $P_{t,t'}$). *Given assumptions 3.1-3.2,* $\hat{P}_{t,t'} = f_{t,t'}$ *is a moment-matching estimator of* $P_{t,t'}$. $\hat{P}_{t,t'}$ *is unbiased.*

Proof. As in proposition 3.1, assumption 3.1 implies that the j-th risk is well represented by the r.v. $d^{(j)}_{t,t'}$.
Hence, assumption 3.2 implies that $d^{(j)}_{t,t'} \sim \mathrm{Bernoulli}(P_{t,t'})\ \forall j \in \{j\}$.
Let us estimate the expectation in t by following a moment-matching approach, given the information available in $t' > t$

$$\mathbf{E}\left[\cdot|\mathcal{F}_t\right] \mapsto \langle\cdot\rangle^{\mathrm{MM}}_{t'} := \frac{1}{|\{j\}|} \sum_{j \in \{j\}} \cdot|\mathcal{F}_{t'},$$

to the first moment

$$P_{t,t'} = \mathbf{E}\left[d^{(j)}_{t,t'}|\mathcal{F}_t\right].$$

Thus we have

$$\hat{P}_{t,t'} = \left\langle d^{(j)}_{t,t'}\right\rangle^{\text{MM}}_{t'} = \frac{1}{|\{j\}|}\sum_{j\in\{j\}} d^{(j)}_{t,t'}|\mathcal{F}_{t'} = f_{t,t'}.$$

Being the *sample mean* an unbiased estimator of the *mean*, $\hat{P}_{t,t'}$ is an unbiased estimator of $P_{t,t'}$ as well. ∎

Finally, let us remove also assumption 3.2: in this case $f_{t,t'}$ cannot be used as an estimator of $P_{t,t'}$ anymore, since $P_{t,t'}$ has a different value for each $j \in \{j\}$. However, it is immediate to prove that $f_{t,t'}$ still represents the average probability of default across the considered portfolio $\{j\}$.

Proposition 3.3 ($f_{t,t'}$ as the MM estimator of the portfolio EDF)**.** *Given assumption 3.1, $\hat{P}_{t,t'} = f_{t,t'}$ is a moment-matching estimator of the expected default frequency $\langle P_{t,t'}\rangle$ of the portfolio $\{j\}$. $\hat{P}_{t,t'}$ is unbiased.*

Proof. Assumption 3.1 implies that the i-th risk is well represented by the r.v. $d^{(j)}_{t,t'}$. Hence it holds $d^{(j)}_{t,t'} \sim \text{Bernoulli}(P^{(j)}_{t,t'})\ \ \forall j \in \{j\}$, where

$$P^{(j)}_{t,t'} \coloneqq \langle P_{t,t'}\rangle + \delta P_j$$

and

$$\sum_{j\in\{j\}} \delta P_j = 0$$

by definition of $\langle P_{t,t'}\rangle$.
It holds

$$\langle P_{t,t'}\rangle = \mathbf{E}\left[\frac{1}{|\{j\}|}\sum_{j\in\{j\}} d^{(j)}_{t,t'}|\mathcal{F}_t\right]$$

Let us apply the moment-matching technique to the first moment:

$$\hat{P}_{t,t'} = \left\langle \frac{1}{|\{j\}|}\sum_{j\in\{j\}} d^{(j)}_{t,t'}\right\rangle^{\text{MM}}_{t'} = \frac{1}{|\{j\}|}\sum_{j\in\{j\}}\frac{1}{|\{j\}|}\sum_{j\in\{j\}} d^{(j)}_{t,t'}|\mathcal{F}_{t'} = f_{t,t'}$$

The first application of the operator $\frac{1}{|\{j\}|}\sum_{j\in\{j\}}$ leads to $f_{t,t'}$, completing the proof, while the second application has no effect. ∎

A bayesian approach leads to the same result: the beta-binomial model, introduced in the following proposition 3.4, implies again that $P_{t,t'}$ can be estimated by the measurable quantity $f_{t,t'}$.

Proposition 3.4 (Beta-binomial model)**.** *Let $\{t_i\}$ $(i = 1 \ldots N+1)$ be a partition of the interval $(t, T]$ s.t.*

$$t_i \coloneqq t + (i-1)\Delta_t, \quad \Delta_t \coloneqq \tfrac{T-t}{N}.$$

Let assumptions 3.1-3.3 hold for each subinterval $[t_i, t_{i+1}]$ ($i = 1 \dots N$). The observable quantity

$$\hat{P}_{t,T} := \frac{\sum_{i=1}^{N} \overline{D}_i}{\sum_{i=1}^{N} \overline{L}_i}$$

where

$$\overline{D}_i := \overline{D}_{t_i,t_{i+1}}, \quad \overline{L}_i := \overline{L}_{t_i}$$

is the limit $N \to \infty$ of the bayesian estimator for $P_{t_0,t_0+\Delta_t}$, choosing beta distribution as the prior.

Proof. Let the prior be

$$P^{(0)}_{\Delta_t} \sim \mathrm{Beta}\left(\alpha_0, \beta_0\right)$$

where $P^{(0)}_{\Delta_t}$ is a short notation for $P^{(0)}_{t_i,t_i+\Delta_t}$. Assumption 3.2 implies that the i index can be omitted without loss of generality. Further, assumptions 3.1-3.3 imply that

$$\mathrm{D}_i := \mathrm{D}_{t_i,t_{i+1}} \sim \mathrm{Bin}\left(\overline{\mathrm{L}}_i, P_{\Delta_t}\right), \quad i = 1 \dots N.$$

Since beta distribution is the conjugate prior of the binomial distribution, applying Bayes theorem to the first observation period $[t_1, t_1 + \Delta_t]$ we have

$$\mathrm{Prob}\left(P_{\Delta_t} = x | \, \overline{\mathrm{D}}_1, \overline{\mathrm{L}}_1\right) = \frac{\mathrm{Prob}\left(\overline{\mathrm{D}}_1, \overline{\mathrm{L}}_1 \middle| P_{\Delta_t} = x\right) \mathrm{Prob}\left(P_{\Delta_t} = x\right)}{\int_0^1 dx' \mathrm{Prob}\left(\overline{\mathrm{D}}_1, \overline{\mathrm{L}}_1 \middle| P_{\Delta_t} = x'\right) \mathrm{Prob}\left(P_{\Delta_t} = x'\right)},$$

where

$$\begin{aligned} \mathrm{Prob}\left(\overline{\mathrm{D}}_1, \overline{\mathrm{L}}_1 \middle| P_{\Delta_t} = x\right) &= \binom{\overline{\mathrm{L}}_1}{\overline{\mathrm{D}}_1} x^{\overline{\mathrm{D}}_1} (1-x)^{\overline{\mathrm{L}}_1 - \overline{\mathrm{D}}_1}, \\ \mathrm{Prob}\left(P_{\Delta_t} = x\right) &= \frac{x^{\alpha_0} (1-x)^{\beta_0}}{B\left(\alpha_0, \beta_0\right)}. \end{aligned}$$

Hence the posterior is

$$\mathrm{Prob}\left(P_{\Delta_t} = x | \, \overline{\mathrm{D}}_1, \overline{\mathrm{L}}_1\right) = \frac{x^{\alpha_0 + \overline{\mathrm{D}}_1} (1-x)^{\beta_0 + \overline{\mathrm{L}}_1 - \overline{\mathrm{D}}_1}}{B\left(\alpha_0 + \overline{\mathrm{D}}_1, \beta_0 + \overline{\mathrm{L}}_1 - \overline{\mathrm{D}}_1\right)}$$

Iterating the acquisition of information for all the N available observations we have

$$\mathrm{Prob}\left(P_{\Delta_t} = x | \, \{\overline{\mathrm{D}}_i, \overline{\mathrm{L}}_i\}_{i=1\dots N}\right) = \frac{x^{\alpha_N} (1-x)^{\beta_N}}{B\left(\alpha_N, \beta_N\right)}$$

where

$$\begin{aligned} \alpha_N &:= \alpha_0 + \sum_{j=1}^{N} \overline{\mathrm{D}}_i \\ \beta_N &:= \beta_0 + \sum_{j=1}^{N} \overline{\mathrm{L}}_i - \overline{\mathrm{D}}_i \end{aligned}$$

Parameter P_{Δ_t} can be estimated as the average of the beta distribution that consider all the available information

$$\hat{P}^{(N)}_{\Delta_t} = \int_0^1 x \frac{x^{\alpha_N} (1-x)^{\beta_N}}{B(\alpha_N, \beta_N)} dx = \frac{\alpha_N}{\alpha_N + \beta_N} = \frac{\alpha_0 + \sum_{i=1}^N \overline{\mathrm{D}}_i}{\beta_0 + \sum_{i=1}^N \overline{\mathrm{L}}_i}$$

In the limit $N \to \infty$ contributions of α_0 and β_0 tend to nullify, completing the proof. ■

Remark 3.2. *Once again, the observed default frequency turns out to be the "natural" estimator of the default probability.*
Indeed

$$\hat{P}^{(N)}_{\Delta_t} \simeq \frac{\sum_{i=1}^N \overline{D}_i}{\sum_{i=1}^N \overline{L}_i} = \sum_{i=1}^N w_i f_{t_i, t_{i+1}}$$

where

$$w_i := \frac{\overline{L}_i}{\sum_{i'=1}^N \overline{L}_{i'}}.$$

Namely, we are putting together the information obtained from distinct observation periods $[t_i, t_{i+1}]$ through a weighted average of the corresponding measured frequencies $f_{t_i, t_{i+1}}$.

3.3 Estimation errors

Assumptions 3.1−3.3 in section 3.2 decribe an homogenous portfolio $\{j\}$, composed by independent risky buyers $j \in \{j\}$, and observed across a given period $[t, t+\Delta_t]$.
Given any reference date t_0, we have seen how the probability $P_{t_0, t_0+\Delta_t}$ of observing the default of a given j-th buyer belonging to $\{j\}$ onto a Δ_t-long holding period can be estimated equal to the default frequency $f_{t,t+\Delta_t}$ measured across the past (*i.e.*, $t \leq t_0 - \Delta_t$) interval $(t, t+\Delta_t]$.
This is true if we suppose at least that assumption 3.1 holds, *i.e.*, that the events occurred in $[t, t+\Delta_t]$ were utterly accessible to the observer.
It is worth discussing how to quantify the error that affects the measure of frequency because this will allow us to estimate the effect of incomplete information later.

Proposition 3.5 (standard error for one-period frequency estimator)**.** *Given assumptions 3.1-3.3, the standard error for the estimator $\hat{P}_{t,t'}$ is*

$$se\left[\hat{P}_{t,t'}\right] = (1 - f_{t,t'}) \sqrt{\frac{f_{t,t'}}{\overline{L}_t (1 - f_{t,t'})}}$$

where $(t, t']$ is a Δ_t-long observation period.

Proof. Assumptions 3.1-3.3 imply that $\mathrm{D}_{t,t'} \sim \mathrm{Bin}\left(\overline{\mathrm{L}}_t, P_{\Delta_t}\right)$, where $P_{\Delta_t} = P_{t,t'} = P_{t_0, t_0+\Delta_t}$ (for any $t_0 \in \mathbb{R}$). Hence

$$\mathrm{Var}\left[\mathrm{D}_{t,t'}\right] = \overline{\mathrm{L}}_t P_{\Delta_t} (1 - P_{\Delta_t})$$

As shown in propositions 3.1 and 3.2, $\hat{P}_{\Delta_t} = f_{t,t'}$. Thus, it follows that

$$\text{se}\left[\text{D}_{t,t'}\right] = \sqrt{\overline{\text{L}}_t f_{t,t'}\left(1 - f_{t,t'}\right)}$$

Since $\overline{\text{L}}_t$ is directly observed without any estimation error, it holds that $\text{se}[\hat{P}_{\Delta_t}] = \text{se}\left[\text{D}_{t,t'}\right]/\overline{\text{L}}_t$. ∎

The survival probability over an interval $(t, t']$ can be decomposed as the product of the survival probabilities over an arbitrary partition $\{t_i\}$ $(i = 0 \ldots N)$ of the same interval

$$S_{t,t'} = \prod_{i=1}^{N} S_{t_{i-1},t_i}$$

If assumption 3.1 holds over $(t_a, t_b]$, then S_{t_a,t_b} can be estimated as

$$\hat{S}_{t_a,t_b} = 1 - f_{t_a,t_b}$$

Hence we can introduce a distinct estimator of $S_{t,t'}$ for each partition $\{t_i\}$, considering the estimator associated to each factor S_{t_{i-1},t_i} separately.

$$\hat{S}_{t,t'}|\{t_i\} = \prod_{i=1}^{N} \left(1 - f_{t_{i-1},t_i}\right)$$

It is worth noticing that the equivalence between the estimated quantities

$$S_{t,t'} = \prod_{i=1}^{N} S_{t_{i-1},t_i}$$

does not imply the equivalence between the estimators $\hat{S}_{t,t'}$ and $\hat{S}_{t,t'}|\{t_i\}$.
However, if it is also true that

$$\hat{S}_{t,t'} = \hat{S}_{t,t'}|\{t_i\}$$

their estimation errors are equivalent as well

$$\text{se}\left[\hat{S}_{t,t'}|\{t_i\}\right] = \text{constant} \quad \forall\{t_i\}$$

and the choice among the infinitely many $\hat{S}_{t,t'}|\{t_i\}$ is indifferent. Indeed, this is the case, as shown in the following proposition.

Proposition 3.6 (Equivalence between uniperiodal and multiperiodal probability estimators)**.** *Given assumption 3.1, let $\{t_i\}$ $(i = 0 \ldots N)$ be a partition of the interval $(t, t']$*

$$t \equiv t_0 < t_1 < \cdots < t_N \equiv t',$$

it holds

$$\hat{P}_{\Delta_t} = f_{t,t'} = 1 - \prod_{i=1}^{N} \left(1 - f_{t_{i-1},t_i}\right)$$

Proof. Let us consider the case $N = 2$ (*i.e.* $\{t_i\} = \{t \equiv t_0; t_1; t_2 \equiv t'\}$). It holds

$$1 - (1 - f_{t_0t_1})(1 - f_{t_1t_2}) = 1 - \frac{\overline{\mathrm{L}}_0 - \overline{\mathrm{D}}_0}{\overline{\mathrm{L}}_0}\frac{\overline{\mathrm{L}}_1 - \overline{\mathrm{D}}_1}{\overline{\mathrm{L}}_1}$$

Assumption 3.1 implies that

$$\overline{\mathrm{L}}_i - \overline{\mathrm{D}}_i = \overline{\mathrm{L}}_{i+1}$$

Applying this result with $i = 0$ implies

$$1 - (1 - f_{t_0t_1})(1 - f_{t_1t_2}) = 1 - \frac{\overline{\mathrm{L}}_1 - \overline{\mathrm{D}}_1}{\overline{\mathrm{L}}_0} = 1 - \frac{\overline{\mathrm{L}}_0 - \overbrace{(\overline{\mathrm{D}}_0 + \overline{\mathrm{D}}_1)}^{=\overline{\mathrm{D}}_{t_0,t_2}}}{\overline{\mathrm{L}}_0}.$$

Hence the proposition is verified for the case $N = 2$. Given that $\overline{\mathrm{D}}_{t_0,t_{i-1}} + \overline{\mathrm{D}}_i = \overline{\mathrm{D}}_{t_0,t_i}$, this result can be extended by induction. ■

Proposition 3.6 shows that uniperiodal and multiperiodal estimators of P_{Δ_t} are equivalent, when assumption 3.1 holds. Conversely, assumption 3.1 being not verified implies that the observed sample $\{j\}$ is not "closed" during the observation period $(t_0, t_0 + \Delta_t]$.

The set $\{j\}$ not being closed means that

- *New subjects may enter* $\{j\}$ *after* t_0. Hence, assumption 3.1 is violated because, given $i^\star, t^\star \in (t_0, t_0 + \Delta_t)$ such that

$$i^\star \in \{i\} \;\; \forall t \geq [t^\star, t_0 + \Delta_t] \quad \wedge \quad i^\star \notin \{i\} \;\; \forall t \in [t_0, t^\star),$$

  the observability of $\tau_{i^\star}$ in $(t_0, t_0 + \Delta_t]$ depends on the additional condition

$$\tau_{i^\star} \in [t^\star, t_0 + \Delta_t].$$

- *Observed subjects may exit* $\{j\}$ *before* $t_0 + \Delta_t$. Hence, assumption 3.1 is violated because, given $i^\star, t^\star \in (t_0, t_0 + \Delta_t)$ such that

$$i^\star \notin \{i\} \;\; \forall t \geq [t^\star, t_0 + \Delta_t] \quad \wedge \quad i^\star \in \{i\} \;\; \forall t \in [t_0, t^\star),$$

  the observability of $\tau_{i^\star}$ in $(t_0, t_0 + \Delta_t]$ depends on the additional condition

$$\tau_{i^\star} \in [t_0, t^\star).$$

The events of migration inside/outside the observed sample $\{j\}$ are known as (left/right) **censoring events**.

The existence of censoring events implies the possibility that

$$\overline{\mathrm{L}}_i - \overline{\mathrm{D}}_i \neq \overline{\mathrm{L}}_{i+1}.$$

Hence, it is possible as well that

$$(1 - f_{t_0t_1})(1 - f_{t_1t_2}) = \frac{\overbrace{\overline{\mathrm{L}}_0 - \overline{\mathrm{D}}_0}^{\neq\overline{\mathrm{L}}_1}}{\overline{\mathrm{L}}_0} \cdot \frac{\overbrace{\overline{\mathrm{L}}_1}^{\neq\overline{\mathrm{L}}_0 - \overline{\mathrm{D}}_0} - \overline{\mathrm{D}}_1}{\overline{\mathrm{L}}_1} \neq 1 - f_{t_0t_2},$$

implying that

$$\text{se}\ \underbrace{\left[\hat{S}_{t,t'}|\{t_j\}\right]}_{=\prod_j(1-f_{t_{j-1}t_j})} \neq \text{se}\ \underbrace{\left[\hat{S}_{t,t'}\right]}_{=1-f_{t_0,t_0+\Delta_t}} = (1-f_{t,t'})\sqrt{\frac{f_{t,t'}}{\overline{\text{L}}_t(1-f_{t,t'})}}$$

The so-called *Greenwood formula* [90] allows to (approximately) evaluate the estimation error

$$\text{se}\left[\hat{S}_{t,t'}|\{t_i\}\right]$$

in a multiperiodal framework where assumption 3.1 is violated. It is worth noticing that, given the relation $S_{t,t'} = 1 - P_{t,t'}$, it holds

$$\text{se}\left[\hat{S}_{t,t'}|\{t_i\}\right] = \text{se}\left[\hat{P}_{t,t'}|\{t_i\}\right] \quad \forall\ \{t_i\}$$

Hence, the estimation error of a survival probability or the corresponding default probability is the same.

The following relation is needed to prove the Greenwood formula.

Proposition 3.7 (Greenwood lemma)**.** *Given a random variable X, it holds*

$$\text{Var}[X] \approx (\text{E}[X])^2\ \text{Var}[\ln X]$$

Proof. Let us consider the first order Taylor series of $\ln X$, centered in $\text{E}[X]$

$$\ln X \approx \ln \text{E}[X] + (X - \text{E}[X])\,\frac{1}{\text{E}[X]}$$

Squaring both sides and taking the expectation gives

$$\text{E}[\ln^2 X] \approx \ln^2 \text{E}[X] + \frac{\text{Var}[X]}{(\text{E}[X])^2}$$

that implies the statement. ∎

By proposition 3.7, the Greenwood formula can be introduced and proved.

Proposition 3.8 (Greenwood formula)**.** *Let $\{t_i\}$ $(i = 0 \dots N)$ be a partition of the interval $(t, t']$*

$$t \equiv t_0 < t_1 < \cdots < t_N \equiv t',$$

where the probability of default onto $(t_{i-1}, t_i]$ can be estimated as

$$\hat{P}_{t_i,t_{i+1}} = f_i := \frac{\overline{D}_i}{\overline{L}_i}.$$

Given assumption 3.3, the standard error of the estimator $\hat{P}_{t,t'} = 1 - \prod_{i=0}^{N-1}(1-f_i)$ can be approximately quantified as

$$se\left[\hat{P}_{t,t'}\right] \approx (1-\hat{P}_{t,t'})\sqrt{\sum_{i=0}^{N-1}\frac{1}{\overline{L}_i}\frac{f_i}{1-f_i}}$$

Proof. The proposition can be proved by a direct computation:

$$\begin{aligned}
\mathrm{Var}\left[\hat{P}_{t,t'}\right] &= \mathrm{Var}\left[1-\hat{P}_{t,t'}\right] \\
&\overset{\text{Prop. 3.7}}{\approx} (1-\hat{P}_{t,t'})^2 \mathrm{Var}\left[\ln(1-\hat{P}_{t,t'})\right] \\
&= (1-\hat{P}_{t,t'})^2 \mathrm{Var}\left[\sum_i \ln(1-f_i)\right] \\
&\overset{\text{Hyp. 3.3}}{=} (1-\hat{P}_{t,t'})^2 \sum_i \mathrm{Var}\left[\ln(1-f_i)\right] \\
&\overset{\text{Prop. 3.7}}{\approx} (1-\hat{P}_{t,t'})^2 \sum_i \frac{\mathrm{Var}\left[1-f_i\right]}{(1-f_i)^2} \\
&= (1-\hat{P}_{t,t'})^2 \sum_i \frac{1}{\mathrm{L}_i} \frac{f_i}{(1-f_i)}
\end{aligned}$$

Taking the square root of both sides completes the proof. ∎

To sum up, given the observation period $(t, t']$, the partition $\{t_i\}$ over the observation period and the estimator

$$\hat{P}_{t,t'} = 1 - \prod_i (1-f_i)$$

we have two possible ways to quantify the estimation error of $\hat{P}_{t,t'}$.
Assumption 3.1 implies that

$$\mathrm{se}\left[\hat{P}_{t,t'}\right] = (1-\hat{P}_{t,t'})\sqrt{\frac{1}{\mathrm{L}_t}\frac{\hat{P}_{t,t'}}{1-\hat{P}_{t,t'}}},$$

as shown in propositions 3.5 and 3.6, while assumption 3.3 implies that

$$\mathrm{se}\left[\hat{P}_{t,t'}\right] \approx (1-\hat{P}_{t,t'})\sqrt{\sum_{i=0}^{N-1}\frac{1}{\mathrm{L}_i}\frac{f_i}{1-f_i}},$$

as shown in propositions 3.7 and 3.8.

3.4 Deterministic censoring events in medicine

This section introduces censoring events and their contrast with assumption 3.1, *i.e.*, complete observability of absorbing events. The problem of inferring the probability of a considered absorbing event is addressed once again, this time handling the lack of information caused by the presence of censoring events. After a practical introduction to censoring events in §3.4.1, two classical results from biostatistics are discussed: the Kaplan-Meier estimator is presented in §3.4.2, while the Cutler-Ederer estimator is described in §3.4.3.

3.4.1 Censoring events

Let us consider again the process η_t introduced in §3.1. A censoring event is defined through its "censoring time" τ_c, such that the absorbing event (a, τ) can be observed in a left neighbourhood (τ_c', τ_c) of τ^c, in the limit $\tau_c' \to \tau_c^-$, iff it cannot be observed in the right neighbourhood (τ_c, τ_c'') of τ_c, in the limit $\tau_c'' \to \tau_c^+$.

In particular, if (a, τ) can be observed in $\tau = \tau_c^+$, then τ_c is called a "left censoring time" while, on the other hand, in case (a, τ) event is not observable in $\tau = \tau_c^+$, then τ_c is called a "right censoring time". The symbols τ^ℓ and τ^r are introduced to represent left and right censoring times respectively.

Remark 3.3. *An arbitrary number of censoring events can take place in a given time interval* $(t, t']$*, with regards to the observability of the same absorbing event* (a, τ)*. However, by definition, neither two left nor two right-censoring events can occur subsequently.*

Namely, in the presence of censoring events, the observer cannot measure $d_{t,t'}$ anymore. Indeed, it is possible to observe only the variable

$$\tilde{d}_{t,t'} := \begin{cases} 1 & \tau \in (t, t'] \setminus \bigcup_j [\tau_j^\ell, \tau_j^r], \\ 0 & \text{otherwise}, \end{cases}$$

where $\tau_j^\ell > \tau_{j-1}^r$ for each j.

The examples of absorbing event considered in table 3.1 can be affected by the presence of censoring events. However, censoring events can lead to a different loss of information, depending on the considered case.

Indeed, in the "retail banking" scheme (figure 3.1), the observer is hardly interested in right censoring events, although they are possible.

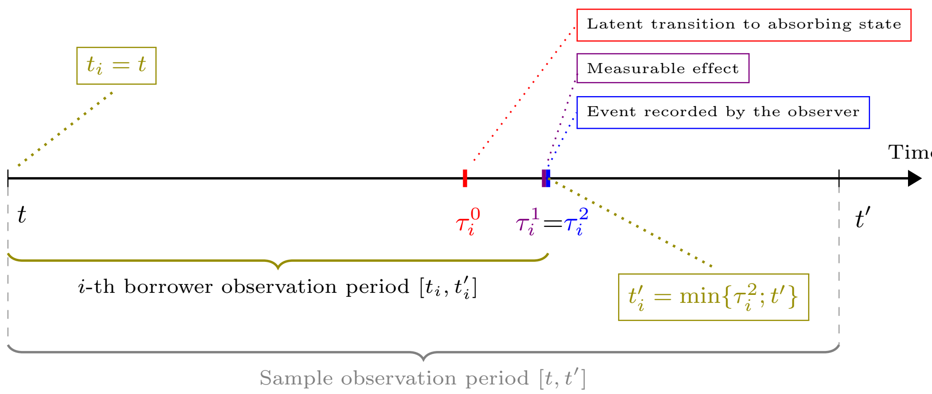

Figure 3.1. Timeline of the "retail banking" case (table 3.1): $t = t_i$ – the i-th borrower begins to pay his installments; τ_i^0 – the i-th borrower loses his job; τ_i^1 – the i-th borrower cannot pay an installment on the due date; $\tau_i^1 = \tau_i^2$ – the bank knows that the i-th borrower is insolvent.

The reasons are merely quantitative. In fact, two censoring events are possible:

$\boxed{\tau_i^r < \tau_i^0}$ *withdrawal due to prepayment* – this case is not common enough to be relevant, compared with the size of the population available to a bank (see, e.g., table 2.1);

$\boxed{\tau_i^r < \tau_i^1}$ *installment expiry after the end of the observation period* $(t, t']$ – in this case, the last expiry τ^r before t' is a right censoring time. However, in a typical real-life application installments are frequent enough that their duration (e.g., 1 month) is small, compared with the observation period length (e.g., 1 year). In such a case, this effect can be neglected with acceptable error.

Loosely speaking, large population size and frequent payments legitimate to neglect the right-censoring events in retail banking. Also, left censoring events can be ignored: considering a population large enough, the bank can afford not to consider the subjects that get their loans after the beginning t of the observation period. On the other hand, small populations or long installment durations could increase the importance of censoring events.

Let us apply the same scheme to the estimation of the survival rate of patients affected by a specific type of cancer (figure 3.2).

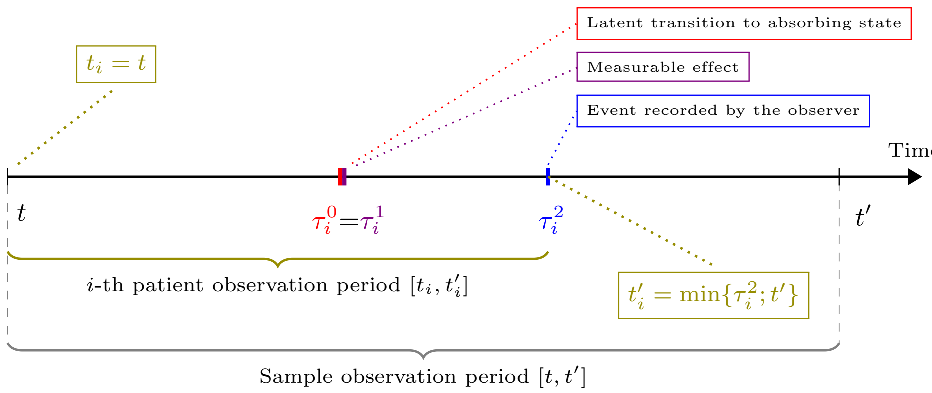

Figure 3.2. Timeline of the "oncology" case (table 3.1): t the study begins; $t = t_i$ – the i-th patient is diagnosed and becomes part of the observation sample; τ_i^0 – the i-th patient health is irreversely impaired; $\tau_i^0 = \tau_i^1$ – the i-th patient dies; τ_i^2 – researchers record the death of the i-th patient.

As the retail banking case, the observer is aware of the period $[t_i, t_i']$ when the i-th subject is actually observable. In this case, both left and right censoring events are possible. Further, the observer cannot afford to neglect them, given that the available population size is usually small. This is the reason why the first techniques [91, 92] to handle the problem of inferring an absorbing event probability were developed in this context rather than in finance.

Indeed, a left censoring event occurs when the i-th patient is diagnosed after the beginning t of the study and becomes part of the observation sample ($t < \tau_i^\ell$), as shown in figure 3.3.

In presence of a left censoring event, $t_i = \tau_i^\ell$ and $\min\{\tau_i^1; t'\}$ are available to the observer, while the health status of the i-th patient in $[t, \tau_i^\ell)$ is unknown. This

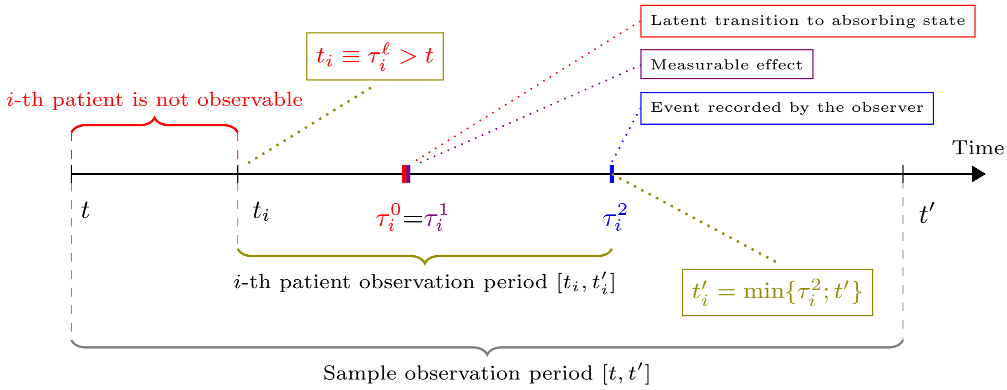

Figure 3.3. Timeline of the "oncology" case (table 3.1, figure 3.2) in presence of a left censoring event.

affects the estimation of the mortality frequency given the presence of the specific investigated illness. In fact, it is not possible to establish if and when the illness started before τ_i^ℓ.

Moreover, also right censoring events are possible, when a patient is lost to follow-up ($\tau_i^r < \tau_i^2$) as shown in figure 3.4.

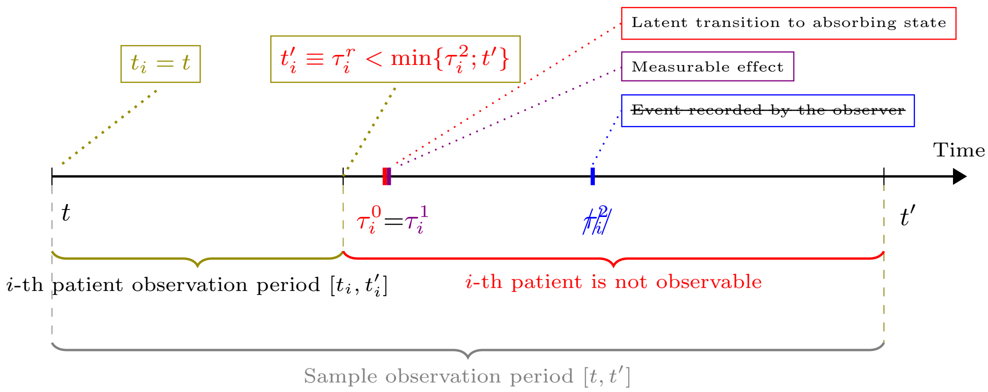

Figure 3.4. Timeline of the "oncology" case (table 3.1, figure 3.2) in presence of a right censoring event. The case $\tau_i^r > \tau_i^0$ - not represented - is possible as well.

Remark 3.4. *In this case, only a single left censoring event and a single right censoring event can occur per subject. The case of multiple left and right censoring events per subject is relevant to credit insurance and is discussed in Chapter 6.*

3.4.2 The Kaplan-Meyer estimator

Let us consider the possible occurrence of left and right censoring events. In this case, assumption 3.1 does not hold anymore.

However, let us assume that the timing of each censoring event is known. Thus, assumption 3.1 is replaced by the following.

Assumption 3.4 (Fully observed censoring events)**.** *An absorbing event generated in τ_i by the i-th subject is immediately observable over the whole considered time interval $(t_0, t_0 + \Delta_t]$, provided that the subject is observable in τ_i, for all $i \in \{i\}$.*

$$\tau_i^0 \simeq \tau_i^1 \simeq \tau_i^2 =: \tau_i$$

Further, all the censoring times $\{\tau^\ell\}$ and $\{\tau^r\}$ in $(t_0, t_0 + \Delta_t]$ are known.

Remark 3.5. *In principle, each subject can have multiple left/right censoring events.*

Introducing censoring events, we are giving up some information. However, we still assume to have a perfect, instantaneous knowledge of censoring events, that allows us to know the exact value of $\overline{\mathrm{L}}_t$ for each $t \in (t_0, t_0 + \Delta_t]$.

Kaplan and Meier [91] took advantage of this opportune condition, obtaining the following result.

Proposition 3.9 (Kaplan Meier estimator)**.** *Given assumptions 3.2–3.4, an observation period $(t_0, t_0 + \Delta_t]$ and a partition $\{t_j\}$ over $(t_0, t_0 + \Delta_t]$, the estimator*

$$\hat{P}_{\Delta_t}^{(KM)} := 1 - \prod_{\tau \in \{\tau\}} \left(1 - \frac{\overline{D}_{\tau^-,\tau}}{\overline{L}_{\tau^-}}\right)$$

is the ML estimator of P_{Δ_t}, where $\{\tau\}$ is the set of absorbing event times observed in $(t_0, t_0 + \Delta_t]$.

Proof. Given assumption 3.4, the following arrays are observable

$$\begin{aligned} \underline{\tilde{\tau}} &:= \textbf{sort}\left[\{\tau\} \cup \{\tau^\ell\} \cup \{\tau^r\}\right], \\ \underline{\tilde{\tau}}^- &:= \lim_{\delta \to 0^+} (\tilde{\tau}_1 - \delta, \tilde{\tau}_2 - \delta, \dots), \end{aligned}$$

where **sort**$[\cdot]$ is the ascending ordering operator.

At each absorbing or censoring event, we can update the value of $\overline{\mathrm{L}}_t$ by the relation

$$\overline{\mathrm{L}}_{\tilde{\tau}_k} = \overline{\mathrm{L}}_{\tilde{\tau}_{k-1}} - \overline{\mathrm{D}}_{\tilde{\tau}_k^-\tilde{\tau}_k} + \overline{\ell}_{\tilde{\tau}_k^-\tilde{\tau}_k} - \overline{\mathrm{R}}_{\tilde{\tau}_k^-\tilde{\tau}_k}$$

where $\overline{\ell}_{\tilde{\tau}_k^-\tilde{\tau}_k}$ and $\overline{\mathrm{R}}_{\tilde{\tau}_k^-\tilde{\tau}_k}$ are the number of subjects interested by a left or right censoring event in $\tilde{\tau}_k$, respectively.

In each interval $[\tilde{\tau}_{k-1}, \tilde{\tau}_k)$, $\overline{\mathrm{L}}_t = \overline{\mathrm{L}}_{\tilde{\tau}_{k-1}}$ is constant by construction.

Given assumption 3.3, we can decompose the survival probability associated with each i-th subject by using an arbitrary partition of the observation interval. We choose the partition induced by $\underline{\tilde{\tau}}^-$:

$$S_{t_0,t_0+\Delta_t} = S_{t_0,\tilde{\tau}_1^-} \cdot S_{\tilde{\tau}_1^-,\tilde{\tau}_2^-} \cdot \ldots \cdot S_{\tilde{\tau}_K^-,t_0+\Delta_t}.$$

By assumptions 3.2−3.4, the LHS is the likelihood function of the parameter P_{Δ_t}, while the factors at the RHS are likelihood functions that can be maximized separately from each other, implying

$$\hat{P}_{\Delta_t} := 1 - \prod_k \left(1 - \frac{\overline{\mathrm{D}}_{\tilde{\tau}_k^-,\tilde{\tau}_k}}{\overline{\mathrm{L}}_{\tilde{\tau}_k^-}}\right)$$

However, for each $\tilde{\tau}_k$ s.t. $\tilde{\tau}_k \notin \{\tau\}$ $\overline{\mathrm{D}}_{\tilde{\tau}_k^-,\tilde{\tau}_k} = 0$, completing the proof. ∎

Applying the Greenwood formula, we are also able to evaluate the standard error associated with the KM estimator, regardless of the presence of censoring events:

$$\begin{aligned} \mathrm{se}\left[\hat{P}^{(\mathrm{KM})}_{\Delta_t}\right] &\approx (1 - \hat{P}^{(\mathrm{KM})}_{\Delta_t})\sqrt{\sum_{\tau\in\{\tau\}} \frac{1}{\overline{\mathrm{L}}_{\tau^-}} \frac{f_{\tau^-,\tau}}{1-f_{\tau^-,\tau}}} \\ &= (1 - \hat{P}^{(\mathrm{KM})}_{\Delta_t})\sqrt{\sum_{\tau\in\{\tau\}} \frac{1}{\overline{\mathrm{L}}_{\tau^-}} \frac{\overline{\mathrm{D}}_{\tau^-,\tau}}{\overline{\mathrm{L}}_{\tau^-}-\overline{\mathrm{D}}_{\tau^-,\tau}}}. \end{aligned}$$

3.4.3 The Cutler-Ederer estimator

Cutler and Ederer [92] approached the same problem of estimating the mortality rate due to cancer by considering a lack of information greater than the one addressed by Kutler and Ederer. Indeed, their study was based on the so-called *life tables*, that is, a standard data set type where censoring and absorbing events are counted and clustered per occurrence year, without reporting the exact timing per event. Table 3.2 is the original life table considered in their study.

Unfortunately, as shown in table 3.2, the data available from life tables do not satisfy neither assumption 3.1 nor assumption 3.4. Hence, to estimate the 5 years mortality rate based on table 3.2, the following must be taken into account:

- for biological reasons, only observations at the same duration $x - t$ are comparable (*i.e.* the hazard rate is inhomogenous across the first 5 years after diagnosis);
- only $x = 1946, 1947$ cohorts have a 5 years observation period;
- the follow up dates per patient are not known.

Since only observations with the same $x-t$ can be clustered together, we can assume that only right censoring events exist. In fact, each new patient enters the observed sample at $x - t = 0$ by construction, regardless of the diagnosis year.

On the other hand, we know only the year in which the right censoring event occurred per patient - not the exact date. This is in contrast with assumption 3.4 and, thus, the estimation problem is addressed introducing the following assumption 3.5 instead.

Assumption 3.5 (Right-censoring events in life tables)**.** *The following holds true for the data available in life tables:*

Diagnosis year Cohort x	Years after diagnosis Duration $t-x$	Alive 1^{st} day of the year $\overline{\ell}_t^x$	Dead during the year $\overline{\mathrm{D}}_{t,t+1}^x$	Lost to follow up during the year $\overline{\mathrm{U}}_{t,t+1}^x$	Withdrawn alive during the last year $\overline{\mathrm{W}}_{t,t+1}^x$
1946	0	9	4	1	0
1946	1	4	0	0	0
1946	2	4	0	0	0
1946	3	4	0	0	0
1946	4	4	0	0	0
1946	5	4	0	0	4
1947	0	18	7	0	0
1947	1	11	0	0	0
1947	2	11	1	0	0
1947	3	10	2	2	0
1947	4	6	0	0	6
1948	0	21	11	0	0
1948	1	10	1	2	0
1948	2	7	0	0	0
1948	3	7	0	0	7
1949	0	34	12	0	0
1949	1	22	3	3	0
1949	2	16	1	0	15
1950	0	19	5	1	0
1950	1	13	1	1	11
1951	0	25	8	2	15

Table 3.2. Original data set considered by Cutler ad Ederer in [92].

*(i.) Absorbing events referred to different subjects have the same probability of occurrence at the same $\delta := t - x$ (*i.e.*, the same number δ of years after the respective diagnoses).*

(ii.) A diagnosis occurred in the x-th year coincides with the beginning of the illness and with Jan 1^{st} of the x-th year.

(iii.) A patient lost to follow up or withdrawn alive during the t-th year is considered either as fully observed through the whole year or lost to follow up since January 1^{st} with the same probability, thus equal to 0.5.

Remark 3.6. *Left censoring events are not considered (hyp. 3.5.i-ii) and right censoring events are simplified (hyp. 3.5.iii) by describing them through a Bernoulli r.v..*

Given 3.5, the effective number $\overline{\mathrm{L}}_t^x$ of observable patient belonging to cohort x through the t-th year is introduced.

$$\overline{\mathrm{L}}_t^x := \overline{\ell}_t^x - \tfrac{1}{2}\underbrace{\left(\overline{\mathrm{U}}_{t,t+1}^x + \overline{\mathrm{W}}_{t,t+1}^x\right)}_{\overline{\mathrm{R}}_{t,t+1}^x}$$

The associated effective frequency f_t^x, measured for a cohort/year (x, t) is

$$f_t^x := \frac{\overline{\mathrm{D}}_{t,t+1}^x}{\overline{\mathrm{L}}_t^x}$$

Let us consider a specific cohort x ($x = 1946, 1947$ - the only cohorts complete through the 5 years period). Hence, considering assumptions 3.2, 3.3, and 3.5 we have

$$\hat{P}_5^x = 1 - \prod_{t=x}^{x+4} (1 - f_t^x)\,; \quad \mathrm{se}\left[\hat{P}_5^x\right] = \left(1 - \hat{P}_5^x\right) \sqrt{\sum_{t=x}^{x+4} \frac{1}{\overline{\mathrm{L}}_t^x} \frac{f_t^x}{1-f_t^x}}.$$

Cutler and Ederer argued that assumption 3.5.i allows to consider complete and incomplete cohorts together, implying a decreased estimation error:

$$\hat{P}_5^{(\mathrm{CE})} = 1 - \prod_{\delta=0}^{4} (1 - f_\delta)\,; \quad \mathrm{se}\left[\hat{P}_5^{(\mathrm{CE})}\right] = \left(1 - \hat{P}_5^{(\mathrm{CE})}\right) \sqrt{\sum_{\delta=0}^{4} \frac{1}{\overline{\mathrm{L}}_\delta} \frac{f_\delta}{1-f_\delta}}$$

where

$$f_\delta := \frac{\overline{\mathrm{D}}_\delta}{\overline{\mathrm{L}}_\delta}, \qquad \overline{\mathrm{D}}_\delta := \sum_{x,t|x\text{-}t=\delta} \overline{\mathrm{D}}_t^x, \quad \overline{\mathrm{L}}_\delta := \sum_{x,t|x\text{-}t=\delta} \overline{\mathrm{L}}_t^x.$$

In fact, $\overline{\mathrm{L}}_t^x|_{x-t=\delta} \leq \overline{\mathrm{L}}_\delta$ by construction and assumptions 3.2, 3.3, and 3.5 imply that $\mathbf{E}[f_\delta] = \mathbf{E}[f_t^x|x - t = \delta]$. Thus, proposition 3.8 leads to

$$\frac{\mathrm{se}\left[\hat{P}_5^{(\mathrm{CE})}\right]}{\mathrm{se}\left[\hat{P}_5^x\right]} \approx \sqrt{\frac{\sum_\delta \left(\overline{\mathrm{L}}_\delta\right)^{-1}}{\sum_t \left(\overline{\mathrm{L}}_t^x\right)^{-1}}} \leq 1$$

In credit risk, cohorts are usually ignored when clustering not-defaulted debtors. However, the CE estimator comes in handy due to the assumption 3.5.iii, which leads to the expression

$$f_t^{(\mathrm{CE})} = \frac{\overline{\mathrm{D}}_{t,t+1}}{\overline{\mathrm{L}}_t - \frac{1}{2}\overline{\mathrm{R}}_{t,t+1}}$$

Indeed, in 2007, Moody's analysts disclosed that they chose this approach to handle withdrawal events in frequency time series [100].

3.5 Feasibility of the presented methods for C&S applications

In §3.2 some classic inference techniques have been reviewed to estimate the probability of a given absorbing event assuming that the observer has full access to all the relevant past events (*i.e.*, Assumption 3.1).

§3.4.2 considers a practical case where Assumption 3.1 is replaced by the weaker Assumption 3.4. In fact, in the context investigated by Kaplan and Meyer, censoring events affect the observation and destroy a part of the information that was considered to be available in §3.2. However, this fact is mitigated by the complete knowledge of when each censoring event occurred during the observation period. These pieces of information enable the maximum likelihood estimation of the frequency based on the observed absorbing events, taking into account observed censoring events as well (see Proposition 3.9).

In §3.4.3, the context that the observer experiences when collecting data is further worsened. Indeed, the timing of occurred censoring events is not known anymore, but their number per observation period (*e.g.*, per year) is still known. This fact is formalized in Assumption 3.5. It enables the definition of the Cutler and Ederer estimator, which is more precise than the "classic" frequency estimator applied to the same context.

However, the techniques mentioned above do not fit the case that past censoring events have to be represented as stochastic variables, lacking any measure about their timing or number of occurrences, that is the case of credit insurance. As displayed in figures 3.5 and 3.6, and previously discussed in §1.2, the information available to a credit insurer is diminished by both left and right censoring events that are not directly measurable by the insurer and, thus, they need to be represented in terms of random variables, even if occurred in the past.

In fact, the insolvency state of a buyer leads to a measurable loss and, thus, to a credit insurance claim, only if a covered and unexpired invoice exists such that its issue date τ_i^ℓ preceeds the transition to the buyer's default τ_0 and its due date τ_i^r follows τ_0. Hence, τ_i^ℓ is a left censoring event (see figure 3.5) and τ_i^r is a right censoring event (see figure 3.6) because, if $\tau_0 < \tau_i^\ell$ or $\tau_0 > \tau_i^r$ for each covered invoice, the absorbing event "default of the *i*-th buyer" is not observable by the insurer.

It is worth remarking that the insured seller can detect early warning signs of the financial distress experienced by a risky buyer better than the insurer due to a direct business relationship between them. Then, he or she may decide to interrupt that relationship and to stop issuing more (insured) invoices[1]. Doing so, the insured seller creates a right-censoring event that the insurer cannot be aware of (unless the insured sellers themselves request that credit limit on a given buyer is nullified - but they have no incentives to do so in most cases). Assuming this ability of the insured seller, τ_0 produces an economic effect only if it is included in the validity period of an insured invoice and not before the issue date. In fact, if insured sellers catch early signals that a buyer's transition to an insolvency state has occurred, they won't issue new insured invoices to that buyer.

The credit insurer has no access to the list of issued invoices, implying that this type of censoring event does not comply with any assumption among 3.1, 3.4, and 3.5. The issues arising from such incomplete information with regards to the estimation of the claim probability are presented in further detail and investigated in chapter 6.

On the other hand, censoring events are not possible in suretyship. Hence, there is no lack of information that may affect the estimation of claim probability. The beneficiary cannot switch on and off the exposition to the risk generated by the

[1] It is possible that an insured seller issues to a buyer only invoices that must be paid upfront (*i.e.*, the credit term is null). By doing so, they can preserve their business relationship without being exposed to the buyer's counterparty risk. However, the upfront invoices are excluded from the insurance coverage.

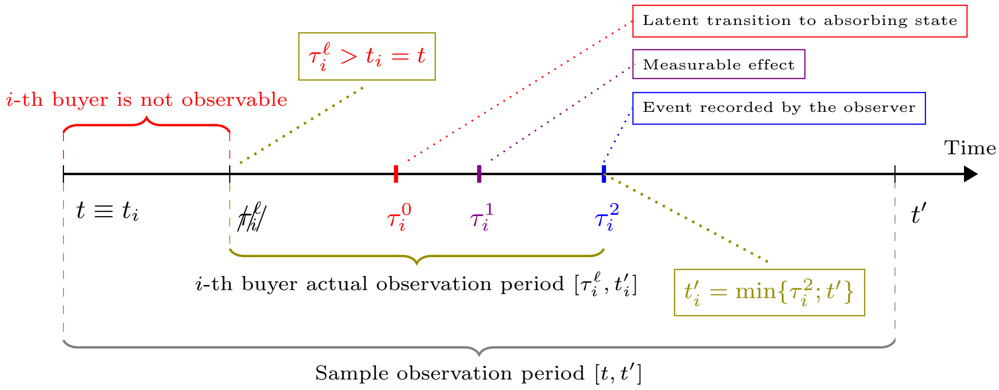

Figure 3.5. Timeline of the "credit insurance" case (table 3.1) in presence of a latent left-censoring event: $t = t_i$ – the policy is activated and a credit limit is granted on the i-th buyer; τ_i^ℓ – a covered invoice is issued and the absorbing event is observable (left censoring event); τ_i^0 – the i-th buyer defaults; τ_i^1 – the i-th buyer cannot pay the invoice by the due date; τ_i^2 – the insurance company knows that the i-th buyer is insolvent.

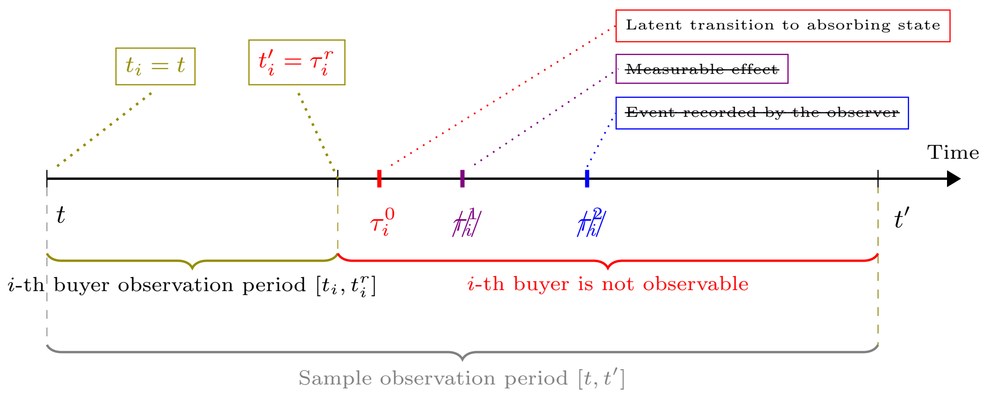

Figure 3.6. Timeline of the "credit insurance" case (table 3.1) in presence of a latent right-censoring event: $t = t_i$ – the policy is activated, a credit limit is granted on the i-th buyer and a covered invoice is issued instantaneously, enabling the default observability; τ_i^0 – the i-th buyer defaults; $t_i' = \tau_i^r$ – the early detection of liquidity distress signal from the buyer leads to the interruption of the business relationship (right censoring event); τ_i^1 – the right censoring event prevents the occurrence of measurable losses (*i.e.*, no covered invoice expires after τ_0); τ_i^2 – there are no claims notified to the credit insurance company.

principal during the lifetime of the insurance coverage. Further, the beneficiary has no incentive to avoid the claim notification if the underlying obligation is violated. Hence, given a cluster of homogeneous risks, the techniques presented in §3.2 are perfectly applicable to this class of insurance products.

Chapter 4

Multivariate probability estimation through the CreditRisk^{+} model

While the development of modern portfolio credit risk models started in the 1980-1990 decade [114] within the framework of the Basel Accords, it is with the great credit crisis of 2008 [115] that increasing attention started to be paid to the precise determination of the structure of dependence among default events. It is well established [116] that tails of the distribution of the value of asset/liabilities portfolios are dominated by the structure of dependence rather than by the other fundamental components of credit risk (*i.e.*, the marginal probability and the severity associated with each future default event). The vast research interest in modeling the structure of dependence resulted in the formalization of the so-called copula theory [117,118]. This "language" was explicitly adopted by the *second generation* of portfolio credit models to describe the dependence among loss events [119–121,128].

In this regard, the calibration issues raised by a particular structure of dependence (or, equivalently, the corresponding *copula*) can be as important as the choice of the structure itself. Generally, calibrating the dependence structure of a portfolio model is a demanding task, given the large number of parameters needed to provide a realistic description of the modeled dependencies, and considering that, on the other hand, historical data are usually not numerous enough to fill the sample space in a way sufficient for a precise estimation of the parameters.

Chapter 2 discussed the main features of the C&S claims, in order to identify a feasible approch to model their probability. In particular, in §2.2, some classical models from the credit risk literature were compared, and the CreditRisk^{+} model resulted to be especially befitting to describe the future claims arising from a C&S basket of risks. In this chapter, we address a typical real-life problem: how to choose the frequency of the historical time series of default used to calibrate the CreditRisk^{+} model, in order to provide the most accurate estimation of the structure of dependence parameters, or, in other words, how the calibration error "*scales*" with the time series frequency. The problem is especially relevant for all the cases when the debtors underlying a credit portfolio are small/medium enterprises. The lack of market information, such as CDS spread, stock price, or bond yield, forces to calibrate the model using a reduced-form approach based on historical cluster data, such as default rate time series associated with the economic sector of each debtor. This case is typical in activities such as credit insurance, suretyship, and factoring. In most cases, publicly available time series have a sampling period ranging from one to three months (e.g., [112]), while the calibrated CreditRisk$^+$ model is used on a projection horizon that is at least one year long (e.g., the unwind period required to quantify a capital requirement both in Solvency 2 and in Basel 3 regulatory frameworks).

CreditRisk$^+$ [134], disclosed in 1997, belongs to the first generation of portfolio credit risk models of "actuarial inspiration". Applications of CreditRisk$^+$ to the credit insurance sector are documented in the literature well before the 2008 financial credit crisis [85, 86], while research activity is still ongoing in the area of actuarial science [102]. At present, CreditRisk$^+$ is still one of the financial and actuarial industry standards for the assessment of credit risk in portfolios of financial loans or credit/suretyship policies.

Despite the vast research activity on this model and its calibration, the issue of using two different time scales for calibration and projection remains not investigated to date. The research done to date on the calibration of CreditRisk$^+$ [102] has addressed the issues related to the decomposition of a given covariance matrix among the time series, which is the final necessary step to complete the calibration of the model. However, the covariance matrix is obtained by the "classical" estimator, under the assumption that the sampling period of the time series and the projection horizon are equal.

This chapter shows that calibrating the model at a shorter time scale than the projection horizon is possible, nontrivial, and convenient. The internal consistency of the CreditRisk$^+$ assumptions when simultaneously imposed at different time scales is proven and guarantees that the investigated calibration mode is not ill-posed. However, the covariance estimator needed to obtain a set of parameters coherent with a specific projection horizon, using time series with a smaller sampling period, depends on the two chosen time scales. Indeed the proposed estimator coincides with the classical one only when calibration and projection time scales are equal. Finally, we show that calibrating at a smaller time scale than the projection one provides a more precise estimation of the model parameters. The estimation error and its dependence on the difference between the two time scales are discussed.

Chapter 4 is organized as follows. In section 4.1 we summarize assumptions and features of the CreditRisk$^+$ model. The same topic is already discussed above in §2.2.6. Nonetheless, it is worth recalling it hereinafter to represent the model as a set of assumptions, that will be generalized in the remainder of this chapter. In section 4.2 we discuss the internal consistency of the model assumptions when imposing them to be simultaneously true at different time horizons. The calibration of the model parameters which define the dependence structure is considered in section 4.3, while the different degree of precision of the estimators defined at increasing time scales is discussed in section 4.4. The techniques introduced in this work are applied to a real-world case study in section 4.5. The main results are summarized in section 4.6.

4.1 The CreditRisk$^+$ model

As stated above in §2.2.6, the structure of the CreditRisk$^+$ model can be summarised as follows. Let N be the number of different risks in a given portfolio and $\mathbb{1}_i$ the default indicator function of the i-th risk ($i = 1, \dots, N$) over the time horizon $(t, T]$. The indicator function $\mathbb{1}_i$ is a Bernoulli random variable such that

$$\mathbf{E}\left[\mathbb{1}_i\right] = q_i, \quad \mathbf{var}\left[\mathbb{1}_i\right] = q_i(1 - q_i), \quad i = 1, \dots, N. \tag{4.1}$$

The "portfolio loss" L over the reference time horizon (t, T) is then given by

$$L = \sum_{i=1}^{N} \mathbb{1}_i E_i \tag{4.2}$$

where each exposure E_i is supposed to be deterministic.

In order to ease the semi-analytic computation of the distribution of L, the model introduces a new set of variables Y_i, each replacing the corresponding indicator function $\mathbb{1}_i$ ($i = 1, \dots, N$). The new variables Y_i are supposed to be Poisson-distributed, conditionally on the value assumed by the market latent variables.

Assumption 4.1 (CreditRisk$^+$ distributional assumption).
Given a time horizon $(t, T]$ and a set of N risky debtors, the number Y_i of insolvency events generated by each i-th debtor over $(t, T]$ is distributed as follows:

$$Y_i \sim \text{Poisson}\left(p_i(\mathbf{\Gamma})\right), \quad p_i(\mathbf{\Gamma}) := q_i \cdot \left(\omega_{i0} + \sum_{k=1}^{K} \omega_{ik} \Gamma_k\right) \tag{4.3}$$

where $\mathbf{\Gamma} = (\Gamma_1 \dots \Gamma_K) \in \mathbb{R}_+^K$ is an array of independent r.v.'s such that

$$\Gamma_k \sim \text{Gamma}\left(\beta_k^{-1}, \beta_k\right), \quad \beta_k \in \mathbb{R}_+ \tag{4.4}$$

and the factor loadings *ω_{ik} are supposed to be all non-negative and to sum up to unity:*

$$\begin{aligned} &\omega_{ik} \geq 0, \qquad i = 1, \dots, N, \quad k = 0, \dots, K, \\ &\sum_{k=0}^{K} \omega_{ik} = 1, \quad i = 1, \dots, N. \end{aligned} \tag{4.5}$$

The $\mathbf{\Gamma}$ parameters set $\{\beta_1 \dots \beta_K\}$ is equivalent to the classical shape-scale parameterization $\{\alpha_k, \beta_k\}$ of each Gamma distributed r.v. Γ_k, after having imposed the assumption $\mathbf{E}\left[\Gamma_k\right] = 1$, that is stated in the original formulation of the CreditRisk$^+$ model. Hence, the k-th scale parameter β_k is equal to the variance σ_k^2 of Γ_k. Given the independence among Γ_k's, the covariance matrix Σ takes the form

$$\Sigma := \mathbf{cov}\left[\mathbf{\Gamma}\right] = \mathbf{diag}\left(\sigma_1^2 \dots \sigma_K^2\right) = \mathbf{diag}\left(\beta_1 \dots \beta_K\right) \tag{4.6}$$

Assumption 4.1 implies that q_i is the unconditional expected default frequency

$$q_i = \mathbf{E}\left[p_i(\mathbf{\Gamma})\right] = \int_{\mathbb{R}_+^K} p_i(\mathbf{\Gamma})\, f\left(\mathbf{\Gamma}\right) d\Gamma_1 \dots d\Gamma_{\mathrm{K}}, \tag{4.7}$$

where

$$f(\boldsymbol{x}) = \prod_{k=1}^{K} \frac{x_k^{\alpha_k - 1}}{\beta_k^{\alpha_k} \Gamma(\alpha_k)} e^{-x_k/\beta_k}, \quad x_k \geq 0, \quad \alpha_k, \beta_k > 0, \tag{4.8}$$

and that the identity between the expected values of the original Bernoulli variable $\mathbb{1}_i$ and the new Poisson variable Y_i is granted:

$$\mathbf{E}[Y_i] = \mathbf{E}[\mathbb{1}_i] = q_i. \tag{4.9}$$

The portfolio loss is now represented by the r.v. L_{Y}

$$L_{\mathrm{Y}} = \sum_{i=1}^{N} Y_i \cdot E_i, \quad \text{where} \quad Y_i | \boldsymbol{\Gamma} \sim \text{Poisson}\left(p_i(\boldsymbol{\Gamma})\right). \tag{4.10}$$

In [134] the distribution of L_{Y} is obtained by using a recursive method, further described in [107]. The accuracy, stability, and possible variants of the original algorithm are discussed in [135]. The same distribution can be easily computed through Monte Carlo simulation due to the availability of a dedicated importance sampling algorithm in [136].

Notice that, although the distributions of L and L_{Y} differ, the expected value of the portfolio loss is the same $\mathbf{E}[L] = \mathbf{E}[L_{\mathrm{Y}}]$.

In the language of copula functions, the structure of dependence implied by (4.3) corresponds [122] to a multivariate Clayton copula, *i.e.* an Archimedean copula where latent variables are Gamma-distributed (for the relation between Archimedean copula functions and factor models see, *e.g*, [128, §2.1]). The copula parameters are the factor loadings ω_{ik} and they can be gathered, taking into account the normalization condition stated in Assumption 4.1, in an $N \times K$ matrix Ω:

$$\Omega := \begin{pmatrix} \omega_{11} & \dots & \omega_{1K} \\ \vdots & \ddots & \vdots \\ \omega_{N1} & \dots & \omega_{NK} \end{pmatrix}, \tag{4.11}$$

which is, for typical values of N and K, much smaller than the $N \times N$ covariance matrix between the default indicators $\mathbb{1}$. As shown in [102], it holds

$$\mathbf{cov}[Y_i, Y_j] = q_i \, q_j \sum_{k=1}^{K} \omega_{ik} \omega_{jk} \sigma_k^2 + \delta_{ij} q_i, \tag{4.12}$$

where δ_{ij} is the Kronecker delta. Equation (4.12) allows the calibration of the factor loadings, and thus of the dependence structure of the CreditRisk$^+$ model, by matching the observed covariance matrix of historical default time series with model values. However, since the model is defined in a single-period framework, with a reference "forecasting" time horizon $(t, T]$, that is typically of 1 year, *i.e.* $T = t + 1$, it is not *a priori* evident how to use historical time series with a different frequency (*e.g.* quarterly) in a consistent way, when calibrating the model parameters. Naively, it is reasonable to expect that the larger the information provided by the historical time series (*i.e.*, the higher the frequency), the better the calibration. This issue is addressed in the next sections.

4.2 CreditRisk$^+$ using multiple unwind periods

The original CreditRisk$^+$ formulation, summarized in Assumption 4.1, defines the model in a uniperiodal framework, where only one time scale $T - t$ is considered. In this section, we discuss the internal consistency of the model assumption when imposing it more than once at distinct time scales. In this context, the expression "internal consistency" means that it is possible and well-posed imposing Assumption 4.1 to be true at two distinct time scales. The same applies also considering a slightly modified version of the CreditRisk$^+$ framework (*i.e.*, imposing Assumption 4.2, introduced in the following, instead of Assumption 4.1).

Extending the original CreditRisk$^+$ formulation to a multiperiod framework enables the calibration of the model considering a time scale different from the one chosen for its application. The results presented in this section are applied in the next section 4.3 to estimate the elements of the matrix

$$A \coloneqq \Omega^T \Sigma \Omega. \tag{4.13}$$

Estimating A is a fundamental step in order to complete the calibration of the model. In section 4.3 estimators are defined using historical series sampled with a period that is not necessarily equal to the projection horizon on which Σ and Ω are defined. Section 4.4 shows the convenience of choosing a sampling period shorter than the projection horizon in order to evaluate $\hat{A}$.

4.2.1 The single unwind period case

As discussed in section 4.1, in CreditRisk$^+$ each risk (*i.e.* debtor) is modeled by a Poisson distributed r.v. Y_i, although the Bernoulli distribution is the natural choice to represent absorbing events, such as default. Assumption 4.1 is convenient in terms of analytical tractability since L_Y distribution can be computed through a semi-analytical method. However, in order to address the problem of calibrating CreditRisk$^+$ in a "roll-over" framework, defined by an arbitrary set of time intervals, it is useful to recover the Bernoulli representation of each debtor by introducing a new r.v. $\widetilde{Y}_i \coloneqq \mathbb{1}_{Y_i>0}$.

Both the r.v. Y_i and its distribution parameter $p_i(\mathbf{\Gamma})$ can take values larger than 1. This is formally correct, given that $Y_i \sim \text{Poisson}\,(p_i(\mathbf{\Gamma}))$, despite not coping with the representation of absorbing events, that can occur at most once by definition. The so-called "Poisson approximation", introduced by substituting $\mathbb{1}_i$ with Y_i, is numerically sound as q_i approaches to zero - a condition that is well fulfilled in most real world relevant cases.

Indeed, Assumption 4.1 implies that $\widetilde{Y}_i|\mathbf{\Gamma} \sim \text{Bernoulli}(\tilde{p}_i(\mathbf{\Gamma}))$ where the distribution parameter is

$$\tilde{p}_i(\mathbf{\Gamma}) = \text{Prob}(Y_i > 0|\mathbf{\Gamma}) = 1 - \exp\left[-q_i\left(\omega_{i0} + \sum_{k=1}^{K} \omega_{ik}\,\Gamma_k\right)\right]. \tag{4.14}$$

It holds by construction

$$\mathbf{E}\left[\widetilde{Y}_i\right] = \int_{\mathbb{R}_+^K} \tilde{p}_i(\mathbf{\Gamma}) f(\mathbf{\Gamma}) d\Gamma_1 \ldots d\Gamma_K. \tag{4.15}$$

Computing the integral in (4.15) and then approximating the term $\exp[-q_i\omega_0]$ with its second order Taylor series centered at $q_i = 0$ leads to the following result.

Proposition 4.1 (Asymptotic equivalence between Bernoulli and Poisson representation of risks)**.**
Let $\widetilde{Y}_i := \mathbb{1}_{Y_i>0}$ *where* Y_i *is distributed according to Assumption 4.1. Then*

$$\tilde{q}_i := \mathbf{E}\left[\widetilde{Y}_i\right] = 1 - e^{-q_i\,\omega_{i0}} \prod_{k=1}^{K} \left(1 + q_i\,\omega_{ik}\sigma_k^2\right)^{-1/\sigma_k^2}. \tag{4.16}$$

Further,

$$\tilde{q}_i = q_i + \mathcal{O}(q_i^2) \xrightarrow{q_i \to 0^+} q_i. \tag{4.17}$$

Proposition 4.1 implies that $\mathbf{E}[L_{Y_i}] \simeq \mathbf{E}[L_{\widetilde{Y}_i}]$, provided that $q_i \ll 1$. Moreover, the same result enables also the exact satisfaction of $\mathbf{E}\left[L_{Y_i}\right] = \mathbf{E}\left[L_{\widetilde{Y}_i}\right]$, in case the stochastic parameter $\tilde{p}_i(\mathbf{\Gamma})$ is redefined through the substitution $q_i \mapsto q_i'$, where q_i' verifies the following modified version of (4.16):

$$\mathbf{E}\left[\widetilde{Y}_i(q_i')\right] = 1 - e^{-q_i'\,\omega_{i0}} \prod_{k=1}^{K} \left(1 + q_i'\,\omega_{ik}\sigma_k^2\right)^{-1/\sigma_k^2} = q_i = \mathbf{E}\left[Y_i\right]. \tag{4.18}$$

It is worth noticing that the substitution $\mathbb{1}_i \mapsto Y_i$ discussed in section 4.1 implies the preservation of the expected value $\mathbf{E}\left[L\right] = \mathbf{E}\left[L_Y\right]$ due to the fact that it is done before the introduction of the market factors $\mathbf{\Gamma}$. On the other hand, restoring the Bernoulli representation of each risk after having introduced the dependence structure requires the results presented in Proposition 4.1.

Proposition 4.1 legitimates the introduction of a slightly modified version of the CreditRisk$^+$ model that is asymptotically equivalent to the original one stated in Assumption 4.1. The equivalence between the two models is further analyzed in the next sections.

Assumption 4.2 (Modified CreditRisk$^+$ distributional assumption)**.**
Given a time horizon $(t, T]$ *and a set of* N *risky debtors, the number of insolvency events generated by each i-th debtor over* $(t, T]$ *is represented by the r.v.* $\widetilde{Y}_i \sim \text{Bernoulli}(\tilde{p}_i(\mathbf{\Gamma}))$*, where the distribution parameter* $\tilde{p}_i(\mathbf{\Gamma})$ *satisfies* (4.14)*. Assumptions on market factors* $\mathbf{\Gamma}$ *and factor loadings* Ω *remain the same stated in Assumption 4.1.*

In Assumption 4.2 the linear dependence of the parameters $p_i(\mathbf{\Gamma})$ from the latent variables has been replaced with a log link function. Thus, the modified version of CreditRisk$^+$ is also referred to as "exponential" in the following.

4.2.2 The multiple unwind periods case

This section investigates the consequences of imposing the internal consistency of Assumption 4.1 or Assumption 4.2 at distinct time scales. Assumptions 4.3 and 4.4 are introduced hereinafter, in order to specify the family of parameters that have to be considered at the distinct time intervals where the model is applied.

The following assumption guarantees the internal consistency at different time scales of the classical CreditRisk$^+$ model, defined in Assumption 4.1.

Assumption 4.3 (CreditRisk$^+$ parameters at different time scales).
Let $t \equiv t_0, t_1, \ldots, t_m \equiv T$ be a partition of the time interval $(t, T]$. Let Assumption 4.1 be satisfied by each j-th interval $(t_{j-1}, t_j]$, where $Y_i^{(j)}$ is the r.v. representing the i-th risk observed during the j-th interval and the following holds for the associated set $\{q_i^{(j)}; \mathbf{\Gamma}^{(j)}; \Omega^{(j)}\}$ of parameters and market factors:

$$\begin{aligned} q_i^{(j)} &= q_i \frac{t_j - t_{j-1}}{T - t} = \text{constant}, &(4.19)\\ \Gamma_k^{(j)} &\sim \text{Gamma}\left(\sigma_k^{-2}\, \xi_{kj}^{-1}\, \tfrac{t_j - t_{j-1}}{T-t}, \sigma_k^2\, \xi_{kj}\, \tfrac{T-t}{t_j - t_{j-1}}\right), &(4.20)\\ \Omega^{(j)} &= \Omega, &(4.21) \end{aligned}$$

where $\xi_{kj} \in \mathbb{R}_+$.

Further, the following assumption guarantees the internal consistency at different time scales of the modified version of CreditRisk$^+$ model, introduced in Assumption 4.2.

Assumption 4.4 (Modified CreditRisk$^+$ parameters at different time scales).
Let $t \equiv t_0, t_1, \ldots, t_m \equiv T$ be a partition of the time interval $(t, T]$. Let Assumption 4.2 be verified by each j-th interval $(t_{j-1}, t_j]$, where $\widetilde{Y}_i^{(j)}$ is the r.v. representing the i-th risk observed during the j-th interval. The associated set $\{q_i^{(j)}; \mathbf{\Gamma}^{(j)}; \Omega^{(j)}\}$ of parameters and market factors satisfies the same assumptions stated in Assumption 4.3.

Finally, for the sake of simplicity, the additional Assumption 4.5 is introduced, with regard to the independence among market factors considered at different times. However, being possible that real-data time series violate Assumption 4.5, this assumption is weakened in the following §4.2.3.

Assumption 4.5 (Non-autocorrelated market factors).
Given Assumption 4.3, let

$$\mathbf{cov}\left[\Gamma_k^{(j)}, \Gamma_k^{(j')}\right] = \delta_{jj'} \mathbf{var}\left[\Gamma_k^{(j)}\right]. \tag{4.22}$$

Considering the assumptions introduced above, we prove that CreditRisk$^+$ is internally consistent when extended to a roll-over framework.

Theorem 4.1 (Internal consistency of CreditRisk$^+$ in absence of autocorrelation).

*Let us consider a set of risks $\{Y_i\}$ ($i = 1 \ldots N$), observed through a time horizon $(t, T]$, and an arbitrary partition $t \equiv t_0, t_1, \ldots, t_m \equiv T$ of $(t, T]$, such that Assumptions 4.3 ("*CreditRisk$^+$ parameters at different time scales*") and 4.5 ("*non-autocorrelated market factors*") are verified with*

$$\xi_{kj} = 1 \tag{4.23}$$

*for each $k = 1 \ldots K$ and $j = 1 \ldots m$. Then $(t, T]$ verifies Assumption 4.1 ("*CreditRisk$^+$ distributional assumption*") .*
*The statement above remains true replacing 4.3 with 4.4 ("*modified CreditRisk$^+$ parameters at different time scales*") and Assumption 4.1 with 4.2 ("*modified CreditRisk$^+$ distributional assumption*"),* ceteris paribus.

The proof of Theorem 4.1 is reported in §A.1.1.
This result shows that extending the CreditRisk$^+$ model to a multiperiod framework is well-posed.

Remark 4.1. *The choice $\xi_{jk} = 1$ implies no loss of generality, since a different (positive) constant $\xi_{jk} = c$ is equivalent to redefine the variances of the market factors $c\sigma_k^2 \mapsto \sigma_k^2$.*

4.2.3 Internal consistency and autocorrelation in time series

As shown in section 4.1, the dynamics of each parameter p_i is induced by the latent Gamma factors only. Imposing Assumption 4.5 to any (arbitrarily short) time scale implies that considered time series $\{\Gamma_k^{(j)}\}_{j=1,2,\dots}$ must exhibit zero autocorrelation. Hence autocorrelation must be completely absent from the historical default frequencies too.

However, this requirement could not be satisfied by the observed time series used in calibrating the model. Indeed, we need to verify that the model can preserve its internal consistency if autocorrelation has to be considered.

The purpose of this work is to investigate whether it is possible and convenient to calibrate the CreditRisk$^+$ model at a time scale that copes with the available historical data (*i.e.*, the sampling period of the historical time series) instead of using the same time scale needed for projections (usually bigger). Hence, in case it is not possible to preserve the internal consistency of the model at each arbitrary time scale, due to the presence of autocorrelation, it is sufficient to ask that it holds up to the smallest of the two time scales of interest - the historical sampling period and the projection horizon.

Let us specialize to the constant mesh case $t_j - t_{j-1} = (T-t)/m = \delta_m$. This choice copes with a typical real case, where the sampling period δ_m of the available historical time series is constant and the considered projection horizon $T-t$ is a multiple of it. Under these premises, a weakened version of Assumption 4.5 is introduced.

Assumption 4.6 (Autocorrelated market factors)**.**
Given Assumption 4.3, for each k-th latent variable, considered at the time scale δ_m, a time-invariant ACF ϱ_{xk} exists, such that

$$\boldsymbol{cov}\left(\Gamma_k^{(j)}, \Gamma_k^{(j+x)}\right) = \varrho_{xk}\boldsymbol{var}\left(\Gamma_k^{(j)}\right). \tag{4.24}$$

Furthermore, the following closure with respect to the addition holds

$$\sum_{j=1}^{m} \Gamma_k^{(j)} \sim \text{Gamma}(\alpha_k, \beta_k) \tag{4.25}$$

for a couple α_k, β_k of shape and scale parameters.

Assumption 4.6 is considered instead of Assumption 4.5 to state the following alternate version of Theorem 4.1.

Theorem 4.2 (Internal consistency of CreditRisk$^+$ model in presence of autocorrelation)**.**
Let us consider a set of risks $\{Y_i\}$ $(i = 1 \dots N)$, observed through a time horizon $(t, T]$, and a uniform partition $\{t_j := t + j\delta_m\}_{j=1\dots m}$ of $(t, T]$, such that Assumptions 4.3 ("CreditRisk$^+$ parameters at different time scales*") and 4.6 (*"autocorrelated market factors*") are verified with*

$$\xi_{kj} = \left[1 + 2\sum_{x=1}^{m-1} \varrho_{xk}\left(1 - \frac{x}{m}\right)\right]^{-\frac{1}{2}} \tag{4.26}$$

for each $k = 1 \dots K$ and $j = 1 \dots m$. Then $(t, T]$ verifies Assumption 4.1 ("CreditRisk$^+$ distributional assumption*").*

The statement above remains true replacing 4.3 with 4.4 ("modified CreditRisk$^+$ parameters at different time scales*") and Assumption 4.1 with 4.2 (*"modified CreditRisk$^+$ distributional assumption*"),* ceteris paribus.

The proof of Theorem 4.2 is reported in §A.1.2.
Assumption 4.6 can be either well-posed or ill-posed, depending on the considered ϱ_{xk}. The trivial case $\varrho_{xk} = 0$ for each $x \in \mathbb{Z}$ copes with Assumption 4.5. Correlated Gamma variables, as well as the distributional properties of the sum of Gamma variables, have been intensively studied in the literature, and this is still an active research field [109,123–125], due to its relevance for information technology. At least in case of identically distributed Gamma variables - such as $\Gamma_k^{(j)}$ in our framework - with ACF obeying to a power-law

$$\varrho_{xk} = \rho_k^{|x|}, \quad \rho_k \in (0, 1), \tag{4.27}$$

the distribution of the sum Γ_k is known to be approximately Gamma [123], while more generical cases imply the sum to be distributed differently [124,125]. Moreover, it is known that partial sums of independent Gamma variables can be used to generate sequences of (auto)correlated Gamma variables [109].

Remark 4.2. *The exponential ACF in eq.* (4.27) *provides a non-trivial case that satisfies Assumption 4.6 and, thus, Theorem 4.2. In the following §4.3.4, Theorem 4.2 legitimates the estimation of A in presence of autocorrelated time series. Eq.* (4.27) *is then considered in §4.4.3 to investigate numerically the estimators introduced in §4.3.4. However, to date, a general framework is missing to tell whether a given ϱ_{xk} lets the partial sums $\sum_j \Gamma_k^{(j)}$ remain (approximately) Gamma distributed, with the exception of exponential ACFs.*

The estimators introduced in §4.3.4 to consider autocorrelation in time series are still applicable to an inconsistent framework, provided that at least the latent variables Γ_k (defined onto the projection horizon) are Gamma distributed and $\Gamma_k^{(j)}$ satisfy the mean and variance requirements implied by Assumption 4.6 above.

4.3 Calibration of the structure of dependence

The model is calibrated based on a partition of the risks in H homogeneous sets $c_h(t)$, $h = 1, \dots, H$. In this context "homogeneity" means that two risks belonging to the same set $c_h(t)$ have the same vector of factor loadings $\boldsymbol{\omega}^{(h)}$. The sets have an explicit time dependence since they can change by the occurrence of defaults. On the contrary, the structure of dependence, defined by $\boldsymbol{\omega}^{(h)}$ is supposed to be time-independent.

Hence, solving the calibration problem requires the evaluation of

- H factor loading vectors $\{\boldsymbol{\omega}^{(h)}\}_{h=1\ldots H}$, that link each of the homogenous clusters to the K latent variables;
- K volatilities $\{\sigma_k\}_{k=1\ldots K}$, needed to specify the distribution of each of the latent variables.

The calibration is achievable by a two-step procedure. Firstly, the matrix $A \coloneqq \Omega^T \Sigma \Omega$, introduced in section 4.2, is estimated. Then, A is decomposed under the proper constraints in order to evaluate Ω and Σ separately. This section describes a method to complete the first step, providing an estimator of A both for the single and the multiple unwind period cases, with a moment-matching approach that allows expressing $\hat{A}$ as a function of the covariance matrix among the historical frequencies of default. The second step is addressed later in section 4.5, which provides an example of calibration using a real data set.

Adopting the standard CreditRisk$^+$ Assumption 4.1, equation (4.12) can be used to link the covariance matrix among the historical frequencies of default with the matrix A. In §4.3.1, $\hat{A}$ is provided in the case of historical frequencies of default, sampled with the same tenor of the projection horizon. In §4.3.2, $\hat{A}$ is generalized to the case of historical frequencies of default sampled with an arbitrary tenor.

Furthermore, in §4.3.3, $\hat{A}$ is determined under the exponential version of the CreditRisk$^+$ framework, introduced in Assumption 4.2. Thanks to this modified assumption, the corresponding functional form of $\hat{A}$ is simpler than the one obtained in §4.3.2 based on Assumption 4.1.

Subsections 4.3.2 and 4.3.3 cope with Assumption 4.5, that implies absence of autocorrelation in time series. The final §4.3.4 uses Assumption 4.6 instead, generalizing the main results presented in this section to the case where autocorrelation must be taken into account. In this case, the simpler form of $\hat{A}$ obtained in §4.3.3 comes in handy in the generalization to the non-trivial ACF case.

4.3.1 The single unwind period case

The first case considered is that of a single unwind period $(t, T]$. For each set $c_h(t)$, let $n_h(t) \coloneqq |c_h(t)|$, $F_h \coloneqq \frac{1}{n_h(t)} \sum_{i \in c_h(t)} Y_i$ and $G_h \coloneqq 1 - F_h$. The expected values of F_h and G_h are respectively:

$$q_h \coloneqq \mathbf{E}[F_h] = \frac{\sum_{i \in c_h(t)} q_i}{n_h(t)}, \tag{4.28}$$

$$s_h \coloneqq \mathbf{E}[G_h] = 1 - \mathbf{E}[F_h]. \tag{4.29}$$

Remark 4.3. *The slight abuse of notation in* (4.28) *is done to avoid the introduction of a new symbol to represent* $\mathbf{E}[F_h]$. *However, the letters chosen for indexing risks and cluster ("i" and "h" respectively) are maintained in the following of this work, clarifying the meaning of the "q" symbol each time it is used.*

For any pair of sets of risks $\{h, h'\}$, the covariance between the default frequencies is:

$$
\begin{aligned}
\mathbf{cov}\left(F_h, F_{h'}\right) &= \mathbf{E}\left[\left(F_h - \mathbf{E}\left[F_h\right]\right)\left(F_{h'} - \mathbf{E}\left[F_{h'}\right]\right)\right] \\
&= \frac{1}{n_h n_{h'}} \mathbf{E}\left[\sum_{i \in c_h}\left(Y_i - q_i\right) \sum_{i' \in c_{h'}}\left(Y_{i'} - q_{i'}\right)\right] \\
&= \frac{1}{n_h n_{h'}} \sum_{i \in c_h} \sum_{i' \in c_{h'}} \mathbf{cov}\left(Y_i, Y_{i'}\right), \qquad (4.30)
\end{aligned}
$$

that, using eq. (4.12), becomes:

$$
\mathbf{cov}\left(F_h, F_{h'}\right) = \frac{1}{n_h n_{h'}} \sum_{i \in c_h} \sum_{i' \in c_{h'}} \left(q_i q_{i'} \sum_{k=1}^{K} \omega_{ik} \omega_{i'k} \sigma_k^2 + \delta_{ii'} q_i\right). \qquad (4.31)
$$

Eq. (4.31) shows the relation between the observed covariance of default frequencies and the factor loadings, describing the structure of dependence of the model.

Moreover, assuming that all risks in a given homogenous set share the same factor loadings, the above expression simplifies to:

$$
\mathbf{cov}\left(F_h, F_{h'}\right) = q_h q_{h'} \sum_{k=1}^{K} \omega_{hk} \omega_{h'k} \sigma_k^2 + \delta_{hh'} \frac{q_h}{n_h} \qquad (4.32)
$$

Notice that the second term in eq. (4.32) is present only when $h = h'$, and becomes quickly negligible as n_h grows (since $q_h < 1$).

Eq. (4.32) enables the estimation of A over the same time scale $T - t$ used for projections:

$$
\hat{A}_{hh'} = \frac{1}{q_h q_{h'}} \left[\widehat{\mathbf{cov}}\left(F_h, F_{h'}\right) - \delta_{hh'} \frac{q_h}{n_h}\right]. \qquad (4.33)
$$

4.3.2 The multiple unwind period case

Let us consider a set of H time series defined using a constant step $\delta_m = (T-t)/m$. As done in section 4.1, each variable introduced in §4.3.1 for the time interval $(t, T]$ can be redefined over each of the considered time intervals. Namely, in the following we use the set of observables quantities $\{F_h, G_h, q_h, s_h\}$, measured either over $(t, T]$ or $(t_{j-1}, t_j = t_{j-1} + \delta_m]$ or a generic time interval $(t, t']$. For the latter two cases we introduce the notation $\{F_h^{(j)}, G_h^{(j)}, q_h^{(j)}, s_h^{(j)}\}$ and $\{F_h(t, t'), G_h(t, t'), \dots\}$, respectively. Further, the variables

$$
\begin{aligned}
F_{mh} &:= 1 - \prod_{j=1}^{m}\left[1 - F_h^{(j)}\right], \qquad (4.34) \\
G_{mh} &:= \prod_{j=1}^{m} G_h^{(j)} = 1 - F_{mh} \qquad (4.35)
\end{aligned}
$$

are introduced.

In CreditRisk$^+$, $F_h(t,t')$ arises from a doubly stochastic process, since each absorbing event is generated conditioned to the latent systematic factors. For the sake of simplicity, we neglect the idiosyncratic uncertainty brought by each $Y_i(t,t')$. In fact, for $n_h(t)$ large enough, the Bernoulli (or Poisson) r.v.'s contributions to the variance of $F_h(t,t')$ are dominated by the contribution of $\mathbf{\Gamma}(t,t')$. This legitimates the following assumption.

Assumption 4.7 (Large clusters)**.**
For each cluster c_h ($h = 1 \dots H$) and each time interval $(t,t'] \subseteq (t,T]$ it holds

$$\mathbf{var}\left[F_h(t,t')|\mathbf{\Gamma}(t,t')\right] = 0\,.$$

Then the following holds:

Proposition 4.2 (CreditRisk$^+$ scale-invariance law)**.**
*Let us consider a set of risks $\{Y_i\}$ ($i = 1 \dots N$), observed through a time horizon $(t_a,t_b]$ and classified into a set of homogenous clusters c_h ($h = 1 \dots H$). Let Assumptions 4.3 ("*CreditRisk$^+$ parameters at different time scales*"), 4.5 ("*non-autocorrelated market factors*") and 4.7 ("*large clusters*") hold with $\xi_{kj} = 1$ for each $(t,T] \subseteq (t_a,t_b]$ and for each uniform partition $t \equiv t_0 < t_1 < \dots < t_m \equiv T$ of $(t,T]$, ($m \in \mathbb{N}^*$). Then the couple $F_h(t,T), F_{h'}(t,T)$ satisfies the conservation law*

$$\left[\mathbf{cov}\left(F_h\left(t,T\right), F_{h'}\left(t,T\right)\right) + s_h\left(t,T\right)s_{h'}\left(t,T\right)\right]^{\frac{1}{T-t}} = \text{constant}. \tag{4.36}$$

for each pair of clusters $c_h, c_{h'}$ and each $(t,T] \subseteq (t_a,t_b]$.

The proof of Proposition 4.2 is reported in §A.1.3.
Proposition 4.2 is one of the main results of this work. It allows to build an estimator of $\mathbf{cov}\left(F_h\left(t,T\right), F_{h'}\left(t,T\right)\right)$ using default frequencies $F_h^{(j)}$ defined on a different time scale δ_m. The dependence upon m of the precision of the covariance estimator is discussed in §4.4.

Indeed, applying Proposition 4.2 to eq. (4.33), it is possible to calibrate the dependence structure of the CreditRisk$^+$ model, by first determining the elements of the A matrix as

$$A_{hh'} = \frac{1}{q_h q_{h'}}\left[\left(\mathbf{cov}\left(F_h^{(j)}, F_{h'}^{(j)}\right) + s_h^{(j)}s_{h'}^{(j)}\right)^m - s_h s_{h'} - \delta_{hh'}\frac{q_h}{n_h}\right] \tag{4.37}$$

for any $j = 1,\dots,m$, and then decomposing A, thus obtaining a separate estimate of the $\{\Omega, \boldsymbol{\sigma}_\Gamma^2\}$ parameters. The SNMF decomposition can be performed, *e.g.*, by using the technique described in [102].

4.3.3 The exponential case

In this section the problem of calibrating the dependence structure is addressed using the exponential form of the model introduced in Assumptions 4.2 and 4.4. Theorem 4.1 proves that also the exponential form remains consistent when considering multiple unwind periods. Since now $\widetilde{Y}_i$ variables are used instead of the corresponding Y_i, the frequencies F_h and their complements G_h are replaced by $\widetilde{F}_h$

and $\widetilde{G}_h$, defined by the substitution $Y_i \mapsto \widetilde{Y}_i$ in F_h and G_h definitions, respectively. Furthermore, it is convenient to introduce the following

$$L_h := -\frac{q_h}{q_h^\star} \ln \widetilde{G}_h \tag{4.38}$$

where

$$q_h^\star := -\ln \frac{\sum_{i \in c_h(t)} e^{-q_i}}{n_h(t)}. \tag{4.39}$$

The notation introduced in §4.3.2 for $\{F_h, G_h, q_h, \dots\}$ is extended to the exponential case as well. Hence, the sets of symbols $\{\widetilde{F}_h(t,t'), \widetilde{G}_h(t,t'), \dots\}$ and $\{\widetilde{F}_h^{(j)}, \widetilde{G}_h^{(j)}, \dots\}$ are also used. The log link function that relates $\tilde{p}_i$ and $\mathbf{\Gamma}$ simplifies the form of the scale invariariance law presented in Proposition 4.2. Indeed, in this case the following holds.

Proposition 4.3 (Modified CreditRisk$^+$ scale-invariance law).
*Let us consider a set of risks $\{\widetilde{Y}_i\}$ $(i = 1 \dots N)$, observed through a time horizon $(t_a, t_b]$ and classified into a set of homogenous clusters c_h $(h = 1 \dots H)$. Let Assumptions 4.4 ("*modified CreditRisk$^+$ parameters at different time scales*"), 4.5 ("*non-autocorrelated market factors*") and 4.7 ("*large clusters*") hold with $\xi_{kj} = 1$ for each $(t,T] \subseteq (t_a, t_b]$ and for each uniform partition $t \equiv t_0 < t_1 < \dots < t_m \equiv T$ of $(t,T]$, $(m \in \mathbb{N}^*)$. Then $L_h(t,T), L_{h'}(t,T)$ obey to the conservation law*

$$\frac{1}{T-t}\mathbf{cov}\left[L_h(t,T), L_{h'}(t,T)\right] = \text{constant} \tag{4.40}$$

for each pair of clusters $c_h, c_{h'}$ and each $(t,T] \subseteq (t_a, t_b]$.

The proof of Proposition 4.3 is reported in §A.1.4.
Proposition 4.3 states a conservation law for the modified version of the model, likewise Proposition 4.2 in the original (*i.e.* Poisson-Gamma) CreditRisk$^+$ framework. The form obtained for the LHS of eq. (4.40) is simpler than the corrisponding LHS of eq. (4.36). In general, this framework results to be more tractable than the original model. This is especially useful when estimating A given a non-trivial ACF, as shown in the next §4.3.4.
In this case, A can be estimated as

$$A_{hh'} = \frac{1}{q_h q_{h'}}\mathbf{cov}\left[L_h, L_{h'}\right] = \frac{1}{q_h^\star q_{h'}^\star}\mathbf{cov}\left[\ln\left(1-\widetilde{F}_h\right), \ln\left(1-\widetilde{F}_{h'}\right)\right] \tag{4.41}$$

where we have neglected the contribution of $\mathbf{cov}(\widetilde{Y}_i, \widetilde{Y}_i) \propto \frac{1}{n_h(t_1)} \simeq 0$. Definition (4.38) and Proposition 4.3 imply

$$A_{hh'} = \frac{m}{q_h^{\star(j)} q_{h'}^{\star(j)}}\mathbf{cov}\left[\ln\left(1-\widetilde{F}_h^{(j)}\right), \ln\left(1-\widetilde{F}_{h'}^{(j)}\right)\right] \tag{4.42}$$

for each $j = 1 \dots m$.

4.3.4 Handling autocorrelated time series in calibration

In this section a generalization of estimators in equations (4.37) and (4.42) is provided, in case Assumption 4.5 has to be replaced with Assumption 4.6 due to the presence of autocorrelation in time series. We preliminarily report below a second order approximation that comes in handy to generalize eq. (4.37).

$$\begin{aligned}\prod_{j=1}^{m}\mathbf{E}\left[G_h^{(j)}G_{h'}^{(j)}\right] &= \prod_{j=1}^{m}\left(1-q_h^{(j)}-q_{h'}^{(j)}+\mathbf{E}\left[F_h^{(j)}F_{h'}^{(j)}\right]\right)\\ &= 1-\sum_{j=1}^{m}\left(q_h^{(j)}+q_{h'}^{(j)}\right)+\sum_{j=1}^{m}\mathbf{E}\left[F_h^{(j)}F_{h'}^{(j)}\right]\\ &+ \sum_{j<j'}\sum_{h,h'=1,2} q_h^{(j)}q_{h'}^{(j')}+\dots\end{aligned}\tag{4.43}$$

We now consider again the relation between $\mathbf{cov}\left(F_{mh},F_{mh'}\right)$ and $\mathbf{cov}\left(F_h^{(j)}F_{h'}^{(j)}\right)$ implied by Proposition 4.2, under the presence of autocorrelation for the latent variables. Unlike in §4.2.2, in this case covariance terms at delay $|j-j'|\geq 1$ cannot be nullified.

$$\begin{aligned}\mathbf{cov}\left(F_{mh},F_{mh'}\right) &= \mathbf{E}\left[\prod_{j=1}^{m}G_h^{(j)}G_{h'}^{(j)}\right]-s_hs_{h'}\\ &= 1-\sum_{j=1}^{m}\left(q_h^{(j)}+q_{h'}^{(j)}\right)\\ &+ \sum_{j<j'}\sum_{h,h'=1,2}\mathbf{E}\left[F_h^{(j)}F_{h'}^{(j')}\right]\\ &+ \sum_{j=1}^{m}\mathbf{E}\left[F_h^{(j)}F_{h'}^{(j)}\right]-s_hs_{h'}+\dots\end{aligned}\tag{4.44}$$

Replacing eq. (4.43) into eq. (4.44) we have

$$\begin{aligned}\mathbf{cov}\left(F_{mh},F_{mh'}\right) &= \prod_{j=1}^{m}\mathbf{E}\left[G_h^{(j)}G_{h'}^{(j)}\right]\\ &+ \sum_{j<j'}\sum_{h,h'=1,2}\mathbf{cov}\left[F_h^{(j)}F_{h'}^{(j')}\right]-s_hs_{h'}+O_3\end{aligned}\tag{4.45}$$

where O_3 is a compact notation for the sum of all the terms of order 3 or greater. Given that $O_3\xrightarrow{\mathbf{q}\to\mathbf{0}}0$, the approximation $O_3\approx 0$ is numerically sound in practice and implies the following generalization of $A_{hh'}$ in eq. (4.37)

$$A_{hh'}\approx\tfrac{1}{q_hq_{h'}}\left[\left(\mathbf{cov}\left(F_h^{(j)},F_{h'}^{(j)}\right)+s_h^{(j)}s_{h'}^{(j)}\right)^m+AC_{hh'}^{(L)}-s_hs_{h'}-\delta_{hh'}\frac{q_h}{n_h}\right],\tag{4.46}$$

where the autocorrelation term $AC^{(L)}$ is defined as

$$AC_{hh'}^{(L)}:=\sum_{x=1}^{m-1}(m-x)\left(\mathbf{cov}\left[F_h^{(j)}F_h^{(j+x)}\right]+\mathbf{cov}\left[F_{h'}^{(j)}F_{h'}^{(j+x)}\right]+2\mathbf{cov}\left[F_h^{(j)}F_{h'}^{(j+x)}\right]\right).\tag{4.47}$$

This completes the extension of the linear case presented in §4.3.2 to autocorrelated time series.

The exponential case - introduced in §4.3.3 - turns out to be more tractable, since the linear structure implied by Proposition 4.3 allows us to avoid approximations similar to the one applied to extend the linear case above. Indeed only the simplification implied by Assumption 4.5 must be abandoned, implying

$$\mathbf{cov}\left[L_h, L_{h'}\right] = m\ \mathbf{cov}\left[L_h^{(j)}, L_{h'}^{(j)}\right] + \sum_{x=1}^{m-1} 2(m-x)\mathbf{cov}\left[L_h^{(j)}, L_{h'}^{(j+x)}\right]. \tag{4.48}$$

This is implied by the fact that $L_h^{(j)}$ are still identically distributed for the same h but not independent. Hence the estimator in eq. (4.42) becomes

$$A_{hh'}(t,T) = \frac{m}{q_h^{\star(j)} q_{h'}^{\star(j)}}\mathbf{cov}\left[\ln\left(1-\widetilde{F}_h^{(j)}\right), \ln\left(1-\widetilde{F}_{h'}^{(j)}\right)\right] + AC_{hh'}^{(E)} \tag{4.49}$$

where

$$AC_{hh'}^{(E)} := \frac{1}{q_h^{\star(j)} q_{h'}^{\star(j)}} \sum_{x=1}^{m-1} 2(m-x)\mathbf{cov}\left[\ln\left(1-\widetilde{F}_h^{(j)}\right), \ln\left(1-\widetilde{F}_{h'}^{(j+x)}\right)\right] \tag{4.50}$$

4.4 The advantage of a short sampling period

Let us consider a Δ_t-long projection period and a set of historical time series of defaults that span a (past) time interval of length $n\Delta_t$. Typical examples can be $\Delta_t = 1$ year and $5 \leq n \leq 20$. Moreover, let the historical time series be sampled with a period δ_m, which is m times smaller than Δ_t (*i.e.* $\delta_m := \Delta_t/m$). Considering $\Delta_t = 1$ year, realistic assumptions are $m = 4$ (quarterly time series) or $m = 12$ (monthly time series). Therefore, the considered time series are defined over $m \times n$ intervals of length δ_m, defined by a schedule $t_0, \ldots, t_{m\times n}$.

This section discusses the precision improvement achievable by calibrating the model on historical default time series with a period smaller than the time horizon on which the calibrated model is applied. Indeed, the statistical error on the determination of A depends on m, *i.e.* on the sampling frequency of the observations, as shown in §4.4.1. Further, given Assumption 4.7 (“large clusters”), the statistical error can be written as a closed-form function of m, as σ_k^2 approaches to zero ($k = 1 \ldots K$). In the following, the assumption of “small” volatilities is referred to as “Gaussian regime”, because it implies $\Gamma_k \dot{\sim} \mathcal{N}(1, \beta_k)$ ($k = 1 \ldots K$), as discussed in the proof of Theorem 4.3.
As in the previous sections 4.2 and 4.3, both the standard CreditRisk$^+$ framework (Assumptions 4.1 and 4.3) and the modified “exponential” version (Assumptions 4.2 and 4.4) are discussed hereinafter.

In applications where c_h's are scarcely populated or σ_k's are not negligible, Theorem 4.3 is not guaranteed to cope with observations. This case is addressed in §4.4.2, where the robustness of the closed-form expression (4.54) is investigated by Monte Carlo simulations.

A numerical approach is maintained in §4.4.3 as well, where the estimation error of $\hat{A}$ at different time scales is measured in presence of autocorrelation, following the generalization introduced in §4.2.3 and §4.3.4. In this case, the exponential version of the model comes in handy: indeed, it is observed that the error on the estimator introduced in (4.46) (*i.e.* standard CreditRisk$^+$ version) does not decrease at increasing m, while the opposite is true for the estimator presented in (4.49) (*i.e.* exponential CreditRisk$^+$ version).

In section 4.3, the $\hat{A}$ estimator has been presented in multiple versions, depending on the considered model (standard or exponential version of CreditRisk$^+$), the chosen sampling period δ_m and the presence or absence of autocorrelation. Thus, it is worth introducing a compact notation to identify the different versions of $\hat{A}$.
The expressions for $A_{hh'}$ presented in (4.37) and (4.46) are addressed as "linear" estimators (as opposed to "exponential") in the following. In these cases the symbol $\hat{A}^{(L,m)}_{hh'}$ is used, where L stands for "linear" and $m = (T-t)/\delta_m$ is the ratio between the projection and calibration time scales.
On the other hand, the expressions for $A_{hh'}$ presented in (4.42) and (4.49) are addressed as "exponential" estimators and so the symbol $\hat{A}^{(E,m)}_{hh'}$ is used.
For the sake of brevity, when L or E is omitted, $\hat{A}^{(m)}_{hh'}$ refers to both the cases and, when m is omitted, $\hat{A}_{hh'}$ refers to the $m = 1$ case.

The improvement in statistical precision with respect to the estimate with no sub-sampling, can be quantified by the following ratio:

$$\varepsilon[\hat{A}^{(m)}_{hh'}] := \sqrt{\frac{\mathbf{var}\left[\hat{A}^{(m)}_{hh'}\right]}{\mathbf{var}\left[\hat{A}_{hh'}\right]}}. \tag{4.51}$$

Symbol $\varepsilon^{(m)}_{hh'}$ and its further specifications $\varepsilon^{(L,m)}_{hh'} := \varepsilon[\hat{A}^{(L,m)}_{hh'}]$ and $\varepsilon^{(E,m)}_{hh'}$ can be used as well. The last short notation that results to be convenient in the following is

$$c^{(Lm)}_{hh'} := \mathbf{cov}\left[F^{(j)}_h, F^{(j)}_{h'}\right] + s^{(j)}_h s^{(j)}_{h'}, \tag{4.52}$$

$$c^{(Em)}_{hh'} := \mathbf{cov}\left[L^{(j)}_h, L^{(j)}_{h'}\right]. \tag{4.53}$$

where $F^{(j)}_h$, $L^{(j)}_h$ and $s^{(j)}_h$ ($j = 1 \dots m$) are i.i.d. variables quantified using a sampling period δ_m.

Remark 4.4. *The notation "$\hat{A}$" refers to the fact the covariances involved in the definitions must be replaced with the corresponding sample estimators, when applying $A^{(m)}_{hh'}$ to historical time series. The same applies to the symbol $\hat{c}$.*

4.4.1 Precision of $\hat{A}$ at different time scales under the Gaussian regime

The following result quantifies the precision gain of performing CreditRisk$^+$ model calibration by historical time series available at increasing sampling frequencies. As anticipated, the precision of the estimated parameters increases as the sampling period decrease. This result holds under Assumption 4.7, in the limit $\boldsymbol{\sigma} \to \mathbf{0}^+$ and considering absence of autocorrelation. The cases where some n_h is small (*i.e.* it does not verify Assumption 4.7) or where some σ_k is not negligible are addressed numerically in the next §4.4.2 - showing that the precision is still increasing as shorter sampling periods are considered. The introduction of autocorrelation is addressed in §4.4.3.

Theorem 4.3 (Estimation errors under Gaussian regime)**.**
Let us consider a set of risks, observed through a time interval $(t_0, t]$ *and classified into a set of homogenous clusters* c_h $(h = 1 \dots H)$*. Let Assumption 4.3 (*"CreditRisk$^+$ parameters at different time scales*"), 4.5 (*"non-autocorrelated market factors*") and 4.7 (*"Large clusters*") hold with* $\xi_{kj} = 1$ *for a given uniform partition* $t_0 < t_1 < \cdots < t_j < \cdots < t_{m \times n} \equiv t$ *of* $(t_0, t]$*,* $(t_j - t_{j-1} = \delta_m;\ m, n \in \mathbb{N}^*)$*. Let* $\hat{A}$ *be the estimate of* A *needed to calibrate the CreditRisk$^+$ model in order to project losses over the time horizon* $(t, T]$*, such that* $(t - t_0)/(T - t) = n$ *and* $(T - t)/(\delta_m) = m$*. Then the following is true for* $\hat{A}^{(m)}_{hh'}$*:*

$$\varepsilon^{(m)}_{hh'} \xrightarrow{\boldsymbol{\sigma} \to \mathbf{0}^+} \sqrt{\frac{n-1}{m \cdot n - 1}} \tag{4.54}$$

Equation 4.54 remains true also considering Assumption 4.4 ("modified CreditRisk$^+$ parameters at different time scales*") instead of Assumption 4.3.*

The proof of Theorem 4.3 is reported in §A.1.5.

4.4.2 Beyond the Gaussian approximation: numerical simulations

In this paragraph we verify that both the estimators $\hat{A}^{(E,m)}_{hh'}$ and $\hat{A}^{(L,m)}_{hh'}$ are more precise at increasing m. The closed-form results obtained in the approximate Gaussian regime, discussed in paragraph 4.4.1, hold when the factor volatilities $\boldsymbol{\sigma}_\Gamma$ are much less than 1. Increasing σ_k $(k = 1, \dots, K)$ the Gaussian approximation becomes less satisfactory and the difference of precision amongst determinations with different values of m becomes smaller. However, the error of $\hat{A}^{(m)}_{hh'}$ remains monotonically decreasing in m, even far from the Gaussian approximation conditions.

For a numerical illustration, we considered a case study with a two-factors market $(\Gamma_k,\ k = 1, 2)$. The dynamic of the couple of systemic factors induces the dependence between two populations of risks, as per the weights reported in Table 4.1. The

Table 4.1. Matrix of weights used for the numerical simulations. The column $k = 0$ displays the factor loadings associated to the "idiosincratic" term of the dependence structure, as introduced in Assumption 4.1.

k	0	1	2
ω_{1k}	0.30	0.40	0.30
ω_{2k}	0.50	0.25	0.25

volatilities $(\sigma_k,\ k = 1, 2)$ associated to the factors are chosen according to seven different scenarios (indexed by i_σ), respectively as

$$\boldsymbol{\sigma}_\Gamma := 2^{i_\sigma} \begin{pmatrix} 2.5 \cdot 10^{-2} \\ 5.0 \cdot 10^{-2} \end{pmatrix}, \quad i_\sigma = 0 \ldots 6.$$

For each scenario, the distributions of the estimators $\hat{A}_{12}^{(E,m)}$ and $\hat{A}_{12}^{(L,m)}$ ($m = 1 \ldots 12$) have been determined using 10^5 simulations of $\{F_1(t, n_1), F_2(t, n_2)\}$ where $t \in (t_0, t_0 + n\Delta_t]$ ($n = 10$) and n_h ($h = 1, 2$) is the number of risks belonging to each cluster. For both estimators the dynamic $\mathbf{F}_h(t, n_h)$ is that reported in (4.14). All risks belonging to the same cluster are supposed to have the same unconditioned intensity of default

$$q_i(t, t + \Delta_t) = -\frac{1}{\Delta_t} \log(0.99), \quad i = 1, \ldots, n_h, \quad h = 1, 2.$$

To investigate the additional contribution to the error $\sigma\left[\hat{A}_{12}\right]$, generated by the finiteness of each cluster, different values of n_h have been considered. In particular, the number of claims per each elementary temporal step $\delta_m = 1/m$ is extracted from a binomial distribution with parameter

$$n_h \in \left\{10^3, 2.5 \cdot 10^3, 5 \cdot 10^3, 10^4, 2.5 \cdot 10^4, 5 \cdot 10^4\right\}, \quad h = 1, 2.$$

For simplicity's sake, it is assumed that each defaulted risk is instantly replaced by a new risk, keeping the population of each cluster constant in time. Finally, the case $n_h = \infty$ (absence of binomial source of randomness) is also considered.

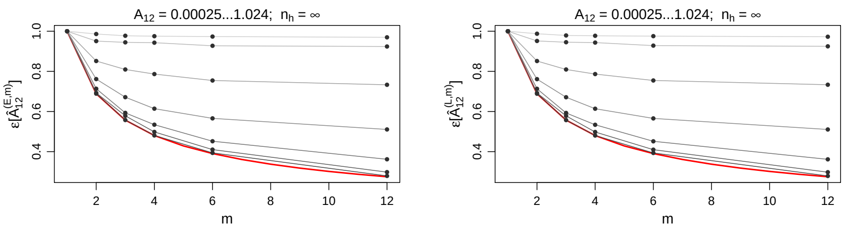

Figure 4.1. *Precision gain $\varepsilon_{12}^{(m)}$, as a function of m and i_σ.* The left and right plots show the values of $\varepsilon\left[\hat{A}_{12}^{(E,m)}\right]$ and $\varepsilon\left[\hat{A}_{12}^{(L,m)}\right]$ respectively, as a function of m, for each volatility scenario ($i_\sigma = 0, \ldots, 6$), each depicted with darker to brighter curves, in the $n_h = \infty$ hypothesis. The red curve is the theoretical value of $\varepsilon\left[\hat{A}_{12}^{(m)}\right]$ in the Gaussian regime.

Figure 4.1 shows the behaviour of $\varepsilon\left[\hat{A}_{12}^{(m)}\right]$ as a function of m, comparing various choices of $\boldsymbol{\sigma}_\Gamma$, considering both the exponential (on the left) and linear (on the right) versions of the model. In this case we are not considering yet the contribution to error due to the finite population ($n_h = \infty$ for each cluster h). Equation (4.54) (red curve) is almost perfectly verified by the least volatility scenario ($\boldsymbol{\sigma}_\Gamma = (2.5, 5.0) \cdot 10^{-2}$). At increasing volatility values (brighter curves), the gain in precision obtained at higher m is reduced, as well as the accordance with equation (4.54).

Since the transformation of $\varepsilon\left[\hat{A}_{12}^{(m)}\right]$ moving away from the Gaussian approximation (i.e.: increasing $|\boldsymbol{\sigma}_\Gamma|$) is smooth, estimating A_{12} with $m > 1$ remains convenient even for $[\boldsymbol{\sigma}_\Gamma]_k \gtrsim 1$, despite the fact that equation (4.54) is not verified anymore.

As mentioned above, comparison between the left and the right panel of figure 4.1 shows that the above argument holds both in the linear and in the exponential case. This is also verified for all the other results of this paragraph.

Figure 4.2 shows the results obtained with finite portfolio populations. The smallest size portfolio ($n_h = 10^3$) is expected to be affected by the largest binomial contribution to the error. This size is considered to be a limiting value for a realistic case. Nevertheless, even in this case, the additional error due to the finiteness of the portfolio is not relevant compared with the one generated by $\boldsymbol{\sigma}_\Gamma$ and the results are very similar to the ones previously shown in figure 4.1.

Figure 4.3 represents the distribution of the estimator $\hat{A}_{12}^{(m)}$ under a subset of choices for $\boldsymbol{\sigma}_\Gamma$ and n_h, selected amongst all possible combinations. The visual comparison between $\mathbf{E}\left[\hat{A}_{12}^{(m)}\right]$ (blue "X" symbol) and A_{12} level (red horizontal line) shows that $\hat{A}_{12}^{(m)}$ is unbiased, both in the linear and in the exponential case (equations (4.37) and (4.42) respectively). The dispersion around the mean reduces at increasing m, in agreement with both equation (4.54) and the numerical results in figures 4.1 and 4.2.

Comparing figures 4.2 and 4.3 it turns out that while the number of risks n_h does not play a relevant role (if any) in computing the ratio $\varepsilon\left[\hat{A}_{12}^{(m)}\right]$, the absolute value of the standard error $\sigma\left[\hat{A}_{12}^{(m)}\right]$ is sensitive to the size of the portfolio.

This fact is confirmed by figure 4.4, where the numerical estimates of $\sigma\left[\hat{A}_{12}^{(m)}\right]$ have been arranged as functions of n_h at fixed $\boldsymbol{\sigma}_\Gamma$ and m values. As expected, the standard error is greater when considering smaller n_h values, while the dependence on n_h of the error disappears quickly as approaching $n_h \to \infty$, even for the highest $\boldsymbol{\sigma}_\Gamma$ values tested.

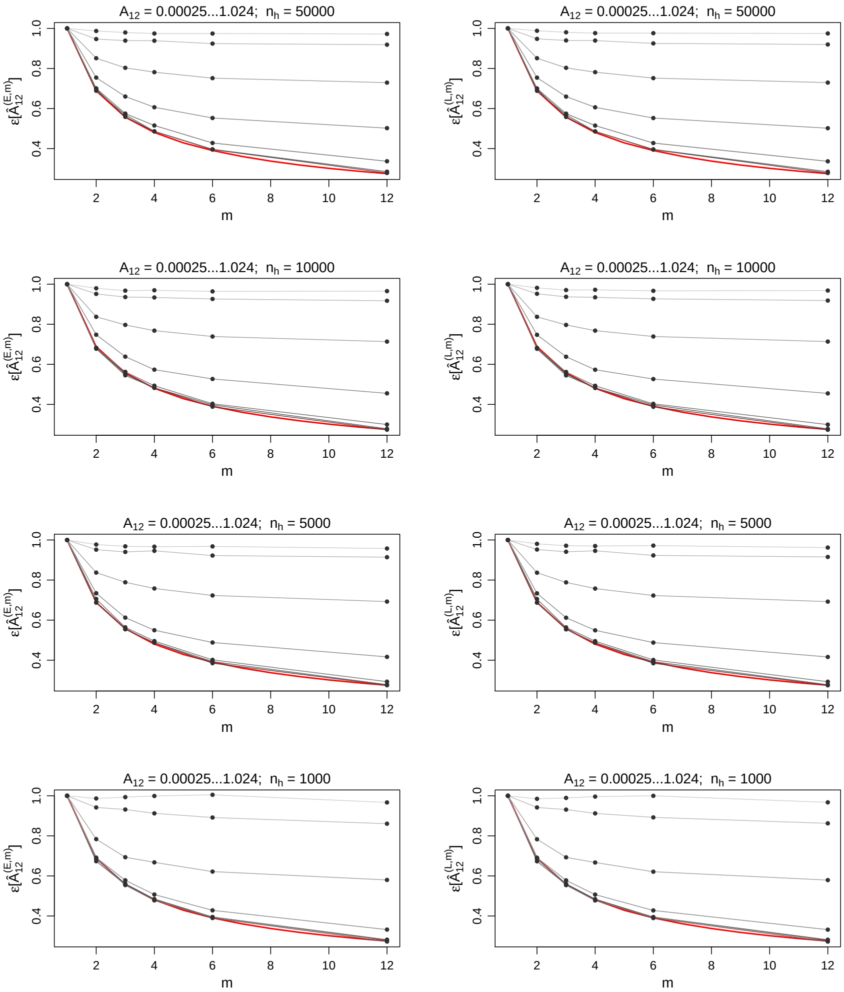

Figure 4.2. *$\varepsilon[\hat{A}_{12}^{(E,m)}]$ and $\varepsilon[\hat{A}_{12}^{(L,m)}]$ as a function of m, considering increasing i_σ (from darker to brighter curve) and decreasing values of n_h (from top to bottom).* The red curve is the theoretical value of $\varepsilon\left[\hat{A}_{12}^{(m)}\right]$ as a function of m in the Gaussian regime. For $\sigma_1, \sigma_2 \ll 1$ the analytical result is perfectly satisfied. However, $\varepsilon\left[\hat{A}_{12}^{(L,m)}\right]$ is shown to be a decreasing function of m in general. Comparing this result with the $n_h = \infty$ case (figure 4.1), we can state that $\varepsilon\left[\hat{A}_{hh'}^{(L,m)}\right]$ is almost insensitive to n_h ($h = 1, 2$), even for $n_h = 10^3$ (the minimum tested value).

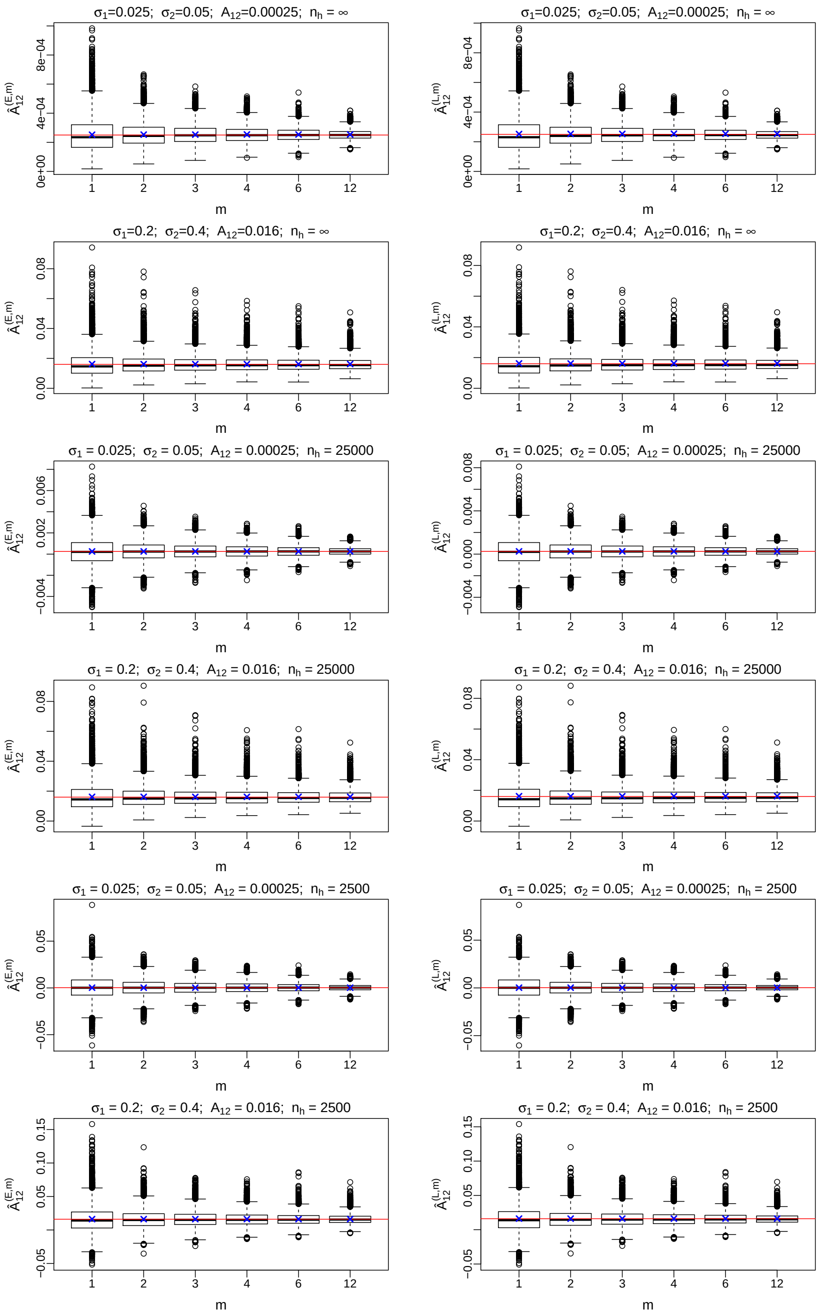

Figure 4.3. *Boxplot of* $\hat{A}_{12}^{(E,m)}$ *and* $\hat{A}_{12}^{(L,m)}$ *distributions, as a function of* m. Three choices for n_h ($h = 1, 2$) are considered, combined with two choices for $\boldsymbol{\sigma}_\Gamma$. The red horizontal line represent the true value of A_{12} and the blue X's stand for the average value of $\hat{A}_{12}^{(m)}$.

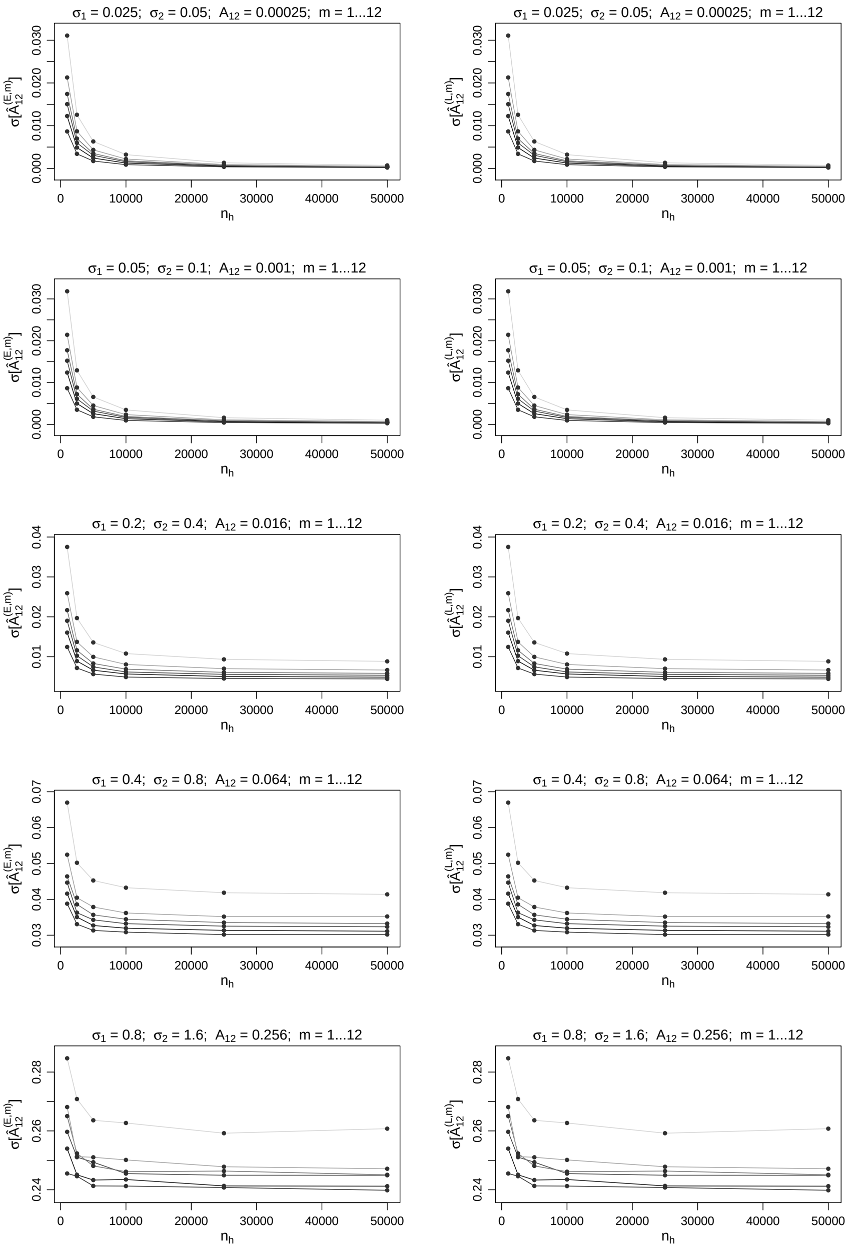

Figure 4.4. $\sigma[\hat{A}_{12}^{(E,m)}]$ *and* $\sigma[\hat{A}_{12}^{(L,m)}]$ *as a function of* n_h. Decreasing m values are considered from darker to brighter curve, as well as various choices of $\boldsymbol{\sigma}_\Gamma$.

4.4.3 Estimation error in presence of autocorrelation

In paragraph 4.4.2 the precision gain at increasing m is measured in absence of autocorrelation. In this paragraph, the same numerical simulations are re-performed introducing autocorrelation and comparing the results against the theoretical estimation of ε. The effect of autocorrelation on ε is discussed in appendix A.2.

The numerical set up introduced above in paragraph 4.4.2 has been maintained, with a further hypothesis about ACF. Indeed, we assume that each latent variable ($k = 1, 2$) obeys to the following ACF law, discussed in paragraph 4.2.3

$$\varrho_{xk} = \rho^{|x|}, \quad m = 12$$

where the considered ρ values are 0.05 and 0.5. For $m' < 12$ cases, we have considered ACF's resulting for the latent variables time series $\tilde{\Gamma}_k^{(j')}$ obtained by the clustering operation

$$\tilde{\Gamma}_k^{(j')} \equiv \frac{m'}{m} \sum_j \Gamma_k^{(j)}, \quad j = 1 + \tfrac{m}{m'}(j'-1) \ldots \tfrac{m}{m'} j'$$

given the aforementioned ACF law at $1/m$ time scale. Since the contribution of the finite population to the error has been shown to be neglectable in par. 4.4.2, simulations in presence of autocorrelation have been performed under $n_h = \infty$ hypothesis only.

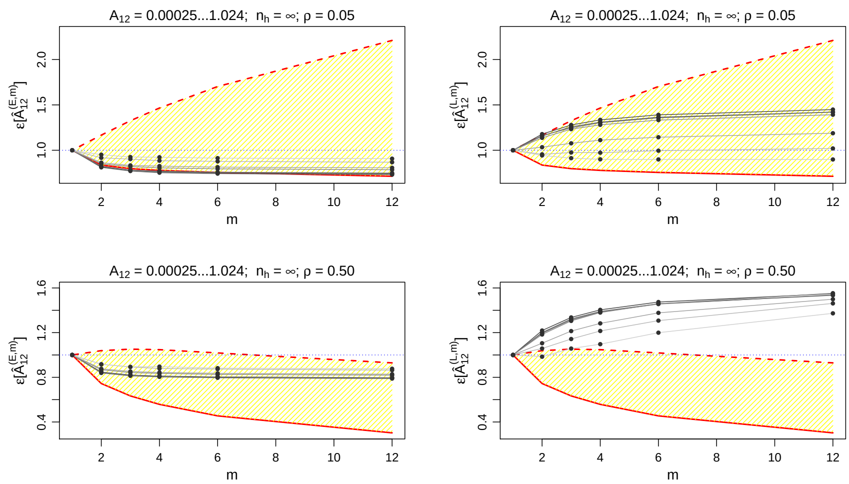

Figure 4.5. *Precision gain in presence of autocorrelation.* $\varepsilon[\hat{A}_{12}^{(E,m)}]$ (exact - left panels) and $\varepsilon[\hat{A}_{12}^{(L,m)}]$ (2^{nd} order approximation - right panels), for each volatility scenario ($i_\sigma = 0, \ldots, 6$, depicted with darker to brighter curves), for $\rho = 0.05$ (top) and $\rho = 0.5$ (bottom). The yellow area includes all the values between the maximum (dashed red line) and the minimum (solid red line) expected from the results of appendix A.2. The frontier $\varepsilon = 1$ (dotted line) allows to check the presence of a precision gain at $m > 1$.

Figure 4.5 shows that the estimator $\hat{A}_{hh'}^{(E,m)}$ remains more precise at increasing m, even in presence of autocorrelation. The analitycal results obtained in the Gaussian regime (i.e. theoretical superior and inferior estimates of ε - dashed and solid red lines in figure 4.5), discussed in appendix A.2, are in good agreement with the numerical results obtained in the considered set up. In fact, all the empirical measures of ε are included between the two theoretical limits (yellow areas).

Moreover, precision gain at $m > 1$ is also possible when using the proxy estimator $\varepsilon\left[\hat{A}_{12}^{(L,m)}\right]$, introduced in eq. (4.46). However, the approximation introduced in this case makes the estimator inefficient (i.e. $\varepsilon > 1$) in the majority of the considered configurations.

4.5 An application to market data

This section provides an example of the calibration technique applied to a real-world data set.

The calibration technique is applied to a set of historical time series of bad loan rates supplied by Bank of Italy. "Bad loan" is a subcategory of the wider class "Non Performing Loan" and it is defined as *exposures to debtors that are insolvent or in substantially similar circumstances* [126]. In particular, the chosen data set is composed by the quarterly historical series TRI30496 ($m = 4$) over a five year period (from 1 Gen 2013 to 31 Dec 2017, $n = 5$, $\Delta_t = 1$). The data are publicly available at [112]. The time series are supplied by customer sector ("*Counterpart institutional sector*") and geographic area ("*Registered office of the customer*"). The latter, in the example, is held fixed to a unique value that corresponds to the whole country (*Italy*). Tables 4.2 and 4.3 report the definition of the 6 different clusters and their main features.

Table 4.2. Definition of the clusters $h = 1, \ldots, 6$ used in data set *TRI30496*.

Cluster index h	Sector Code	Description
1	600	Consumer households
2	S11	Non-financial companies
3	S12BI7	Financial companies other than monetary financial institutions
4	S13	General government
5	S14BI4	Producer households
6	S15BI1	Non-profit institutions serving households and unclassifiable units

Table 4.3. Main features of the considered historical time series over the period 1 Gen 2013 – 31 Dec 2017. $\overline{p}_h$ ($h = 1, \dots, 6$) is the yearly average bad loan rate; σ_h is the volatility associated to each $\overline{p}_h$; $\langle n_h \rangle$ is the average number of borrowers.

h	1	2	3	4	5	6
$\overline{p}_h$	0.0119	0.0352	0.0255	0.0056	0.0259	0.0088
σ_h	0.0010	0.0042	0.0023	0.0014	0.0022	0.0010
$\langle n_h \rangle$	269515	407602	3191	5416	132179	4020

By inspection of table 4.3, it is possible to perform a rough estimate of $\boldsymbol{\sigma}_\Gamma$. Equation (4.14) implies that the following holds for coefficients of variation CV_h ($h = 1, \dots, H \equiv 6$):

$$CV_h := \frac{\sigma_h}{\overline{p}_h} \simeq \sum_{k=1}^{K} \omega_{hk} \left[\boldsymbol{\sigma}_\Gamma\right]_k$$

Furthermore, the normalization requirement over the factor loadings ω_{hk} implies

$$\sum_{k=1}^{K} \omega_{hk} \lesssim 1$$

Hence we can state that $\langle CV \rangle := \frac{1}{H} \sum_h CV_h$ has the same order of magnitude of $\frac{1}{K} \sum_k \left[\boldsymbol{\sigma}_\Gamma\right]_k$. Since $\langle CV \rangle \simeq 0.124$, results in paragraph 4.4.2 suggest that this data set is not far from the Gaussian regime and so there is an appreciable increase of precision in estimating $A(0,1)$ with $m > 1$.

$\hat{A}(0,1)$ is estimated by applying equation (4.42) over a one-year period. The results obtained for $\hat{A}^{(E,m)}(0,1)$ ($m = 1, 4$) are reported in table 4.4.

Table 4.4. Values of the matrix $\hat{A}^{(E,m)}(0,1)$ ($m = 4$ left, $m = 1$ right) obtained from the quarterly historical series *TRI30496* over the period 1 Gen 2013 – 31 Dec 2017. Results are expressed in 10^{-2} units.

0.53	0.28	0.33	0.36	0.41	0.48
0.28	0.59	0.48	0.61	0.43	0.40
0.33	0.48	0.67	0.52	0.43	0.40
0.36	0.61	0.52	7.80	0.48	0.33
0.41	0.43	0.43	0.48	0.47	0.54
0.48	0.40	0.40	0.33	0.54	1.53

0.68	0.40	0.36	1.01	0.56	0.73
0.40	1.50	0.98	1.26	0.87	-0.16
0.36	0.98	0.87	0.78	0.72	0.27
1.01	1.26	0.78	6.50	1.10	0.66
0.56	0.87	0.72	1.10	0.74	0.47
0.73	-0.16	0.27	0.66	0.47	1.35

The elementwise precision gain for $m = 4$, $\varepsilon\left[\hat{A}^{(E,4)}(0,1)\right]$, obtained under the Gaussian approximation hypothesis, is shown in table 4.5. This result is obtained applying definition (4.51) and equation (A.25) both to the cases $m = 4$ and $m = 1$. Eq. (A.25) has been shown to be valid under the Gaussian approximation, discussed in paragraph 4.4.1.
In this case, the preliminary decomposition of $\hat{A}^{(E,4)}(0,1)$, that would be needed using the Monte Carlo method discussed in paragraph 4.4.2, is not needed.

Table 4.5. The matrix of elementwise precision gains $\varepsilon\left[\hat{A}^{(E,4)}(0,1)\right]$ associated with the matrixes reported in table 4.4.

0.36	0.26	0.37	0.41	0.33	0.39
0.26	0.18	0.24	0.30	0.23	0.33
0.37	0.24	0.35	0.43	0.30	0.45
0.41	0.30	0.43	0.55	0.37	0.52
0.33	0.23	0.30	0.37	0.29	0.42
0.39	0.33	0.45	0.52	0.42	0.52

According to equation (4.54), the elements of $\varepsilon\left[\hat{A}^{(E,4)}(0,1)\right]$ reported in table 4.5 should be all approximately equal to 0.46, since they should depend only on the couple m, n ($m = 4$ and $n = 5$ in this case). However, in a real world case like the one considered, the hypothesis of zero autocorrelation is satisfied with a different precision by each time series $p_h(t)$. Furthermore, the estimated covariance matrices might need to be regularized (indeed the Higham regularization algorithm [127] was used both for $m = 1$ and $m = 4$ series). Hence, a different ratio for each element $(h, h') = 1, \dots, 6$ is justified. Nonetheless, it is worth noticing that all the ratios reported in table 4.5 have the same order of magnitude of the predicted value 0.46.

Knowledge of the historical number of risky subjects $n_h(t)$ for each cluster ($h = 1, \dots, 6$) at each observation date ($t = 1/4, 2/4, \dots, 5$) allows to take into account the binomial contribution to the error $\sigma\left[\hat{A}^{(E,m)}(0,1)\right]$, both for $m = 4$ (quarterly series) and $m = 1$ (yearly series), although the finiteness of the population does not add a relevant contribution to the error, as already observed in paragraph 4.4.2.

Table 4.6 provides Monte Carlo estimation of $\sigma\left[\hat{A}^{(E,m)}(0,1)\right]$ ($m = 1, 4$), which considers also the role of $n_h(t)$. Since the values in table 4.6 provide a measure of the error in the determination of $\hat{A}^{(E,m)}(0,1)$, it turns out that the estimates reported in table 4.4 are elementwise consistent one with the other.

Table 4.6. Matrixes $\sigma\left[\hat{A}^{(E,m)}(0,1)\right]$ ($m = 4$ left, $m = 1$ right). These are the elementwise errors of the estimators reported in table 4.4. The results above are expressed in 10^{-2} units.

0.11	0.12	0.19	0.44	0.11	0.31
0.12	0.20	0.29	0.64	0.13	0.40
0.19	0.29	1.38	1.08	0.21	0.69
0.44	0.64	1.08	5.12	0.49	1.60
0.11	0.13	0.21	0.49	0.13	0.34
0.31	0.40	0.69	1.60	0.34	3.25

0.41	0.25	0.41	1.09	0.24	0.64
0.25	0.86	0.72	1.47	0.50	0.97
0.41	0.72	1.75	2.31	0.49	1.53
1.09	1.47	2.31	9.11	1.20	3.40
0.24	0.50	0.49	1.20	0.29	0.79
0.64	0.97	1.53	3.40	0.79	4.65

The Monte Carlo estimation of $\sigma\left[\hat{A}^{(E,m)}(0,1)\right]$, as done in paragraph 4.4.2, requires the *a priori* knowledge of the true dependence structure $W, \boldsymbol{\sigma}_\Gamma$. Since this is a case study, we do not have an *a priori* parameterization of the calibrated model. Hence, we have used $\hat{W}, \hat{\boldsymbol{\sigma}}_\Gamma$ estimated from $\hat{A}^{(E,4)}(0,1)$ instead, as a proxy of the "true" model parameters. The computation of $\hat{W}, \hat{\boldsymbol{\sigma}}_\Gamma$ from $\hat{A}^{(E,4)}(0,1)$ is discussed below.

In order to complete the CreditRisk$^+$ calibration, we have to decompose $\hat{A}$ and find the factor loadings matrix $\hat{W}$ together with the vector of systemic factors variances

$\hat{\boldsymbol{\sigma}}_{\Gamma}^2$. To do so, we use the Symmetric Non negative Matrix Factorization (SNMF), an iterative numerical method to search an approximate decomposition of $\hat{A}$ which satisfies the requirements of the CreditRisk$^+$ model over $\hat{W}$ (*i.e.* all elements $\omega_{hk} > 0$ and $\sum_k \omega_{hk} = 1$). The application of SNMF to CreditRisk$^+$ is discussed in detail in [102]. In the following, we give evidence only of the implementation details necessary to address this case study.
Being an iterative method, SNMF requires an initial choice of matrixes

$$\begin{aligned} \hat{U}_0 &:= \hat{W}_U \hat{\Sigma}^{1/2}, \\ \hat{V}_0 &:= \hat{\Sigma}^{1/2} \hat{W}_V, \end{aligned}$$

such that $\hat{A} = \hat{U}_0 \hat{V}_0$. It is not required that $\hat{U}_0 = \hat{V}_0^T$, nor all the elements of $\hat{U}_0$ and $\hat{V}_0$ have to be positive. We set $\hat{U}_0, \hat{V}_0$ from the eigenvalues decomposition of $\hat{A}^{(E,4)}(0,1)$.

For the considered data set, the eigenvalues decomposition returned the set of eigenvalues and eigenvectors reported in table 4.7.

Table 4.7. Set of eigenvalues $\tilde{\sigma}_k$ and eigenvectors $\tilde{\boldsymbol{\omega}}_k$ obtained by the eigenvalues decomposition of $\hat{A}^{(E,4)}(0,1)$, as reported in Table 4.4.

$\tilde{\boldsymbol{\omega}}_k$	k=1	k=2	k=3	k=4	k=5	k=6
	0.06	-0.34	0.13	0.84	0.00	0.39
	0.10	-0.33	0.43	-0.38	-0.65	0.36
	0.09	-0.36	0.52	-0.26	0.72	0.06
	0.98	0.17	-0.07	0.01	0.01	0.00
	0.08	-0.38	0.22	0.19	-0.22	-0.84
	0.07	-0.68	-0.69	-0.20	0.06	0.07
$\tilde{\sigma}_k^2$	0.08	0.02	0.01	2.9 10^{-3}	1.4 10^{-3}	0.3 10^{-3}

We use the $\tilde{\boldsymbol{\omega}}, \tilde{\sigma}$ notation to address the quantities over which the normalization requirement of CreditRisk$^+$ has not been imposed yet.

Since more than the 95% of variance is explained by the first three eigenvectors, we reduced the dimensionality of the latent variables vector to be $K = 3$. Hence we define

$$\begin{aligned} \hat{U}_0 &= [\tilde{\boldsymbol{\omega}}_1, \tilde{\boldsymbol{\omega}}_2, \tilde{\boldsymbol{\omega}}_3] \cdot \mathrm{diag}(\tilde{\sigma}_1, \tilde{\sigma}_2, \tilde{\sigma}_3) \\ &= \begin{bmatrix} 1.80 & 5.29 & 1.15 \\ 2.78 & 5.18 & 3.71 \\ 2.51 & 5.58 & 4.46 \\ 27.79 & -2.65 & -0.57 \\ 2.33 & 5.88 & 1.93 \\ 2.07 & 10.58 & -5.96 \end{bmatrix} \cdot 10^{-2} \end{aligned}$$

and $\hat{V}_0 = \hat{U}_0^T$. In general, SNMF aims to minimize iteratively the cost function

$$\left| \hat{A} - \hat{U}\hat{V} \right|^2 + \alpha \left| \hat{U} - \hat{V}^T \right|^2$$

where $|\cdot|$ is the Frobenious norm, eventually weighted, and α is a free parameter to weight the asymmetry penality term. In our case the classical Frobenious norm is used (e.g. uniformly weighted). Application of SNMF to our case takes the form

$$\hat{V}_{s+1} = \max\left\{0, \hat{V}_s \circ \frac{\hat{U}_s^T \hat{A} + \alpha(\hat{V}_s - \hat{U}_s^T)}{\hat{U}_s^T \hat{U}_s \hat{V}_s + \epsilon \mathbf{1}_{\{K\times M\}}}\right\} \tag{4.55}$$

$$\hat{U}_{s+1} = \max\left\{0, \hat{U}_s \circ \frac{\hat{A}\hat{V}_s^T + \alpha(\hat{U}_s - \hat{V}_s^T)}{\hat{U}_s \hat{V}_s \hat{V}_s^T + \epsilon \mathbf{1}_{\{M\times K\}}}\right\} \tag{4.56}$$

where "$\circ$" is the Hadamard product and $\epsilon \ll 1$ is the iteration step. In our case $\epsilon = 10^{-6}$ and $\alpha = 1.5 \cdot 10^{-2}$. Further details on the method, such as the so called "balancing step" – that we have implemented – and the proof of equations (4.55) and (4.56) are available in [102]. The application of SNMF method, together with the normalization constraint over the factor loadings, leads to the result reported in table 4.8.

Table 4.8. The complete set of parameters $\hat{W}, \hat{\boldsymbol{\sigma}}_\Gamma^2$ necessary to specify the dependence structure in CreditRisk$^+$ model, obtained by the eigenvalues decomposition of $\hat{A}^{(E,4)}(0,1)$, as reported in Table 4.4.

k	0	1	2	3
ω_{1k}	0.67	0.04	0.29	0.00
ω_{2k}	0.07	0.07	0.27	0.59
ω_{3k}	0.00	0.06	0.28	0.66
ω_{4k}	0.13	0.87	0.00	0.00
ω_{5k}	0.63	0.06	0.31	0.00
ω_{6k}	0.29	0.04	0.67	0.00
σ_k^2		0.103	0.031	0.010

The set of parameters resulting from the calibration process described above is supported by a reasonable economic interpretation. Indeed, factor loadings associated with the "general government" sector ($h = 4$) are completely distinct from the ones of the other sectors (*i.e.* this is the only sector mainly depending on the $k = 1$ factor): this fact copes with the different nature of the public entities from the ones belonging to the other considered sectors. Furthermore, "companies" ($h = 2, 3$) share approximately the same dependence structure. The same applies when considering "households" ($h = 1, 5$). Finally, the "institutions serving households" sector ($h = 6$) shares the same latent factor ($k = 2$) but shows a different balance between idiosyncratic and systemic factor loadings compared to "households", that is coherent with the nature of a sector strongly linked to "households" sectors, despite not being completely equivalent.

Results in table 4.8 have been used to quantify the estimation errors reported in table 4.6.

4.6 Summary

In this chapter we have investigated how to calibrate the dependence structure of the CreditRisk$^+$ model, when the sampling period δ_m of the (available) default rate time series is different from Δ_t – the length of the future time interval chosen for the projections.

To address this topic, it has preliminary been shown that CreditRisk$^+$ remains self-consistent when the underlying distributional hypothesis is imposed to be simultaneously true at different time scales (theorem 4.1). The model self-consistency has been proved to be roboust against the introduction of autocorrelation, depending on the considered ACF form (theorem 4.2).
Successively, the problem has been approched in terms of moment matching, providing two (approximately) equivalent formulations for estimating the covariance matrix A amongst the systemic factors of the model (theorems 4.2 and 4.3). The choice between the two estimators of A, provided in equations (4.37) and (4.42), depends on the functional form (linear or exponential) that links the probability of claim/default and the latent variables. Both the estimators are explicitly dependent from the ratio Δ_t/δ_m, allowing for the calibration of the model at a time scale which is different from the one that is intended to be used for applying the calibrated model. Both the estimators have been generalized in order to be applied to autocorrelated time series in equations (4.46) and (4.49), although only the latter (i.e. exponential case) is an exact result, while for the linear case a second order approximation has been adopted.
Furthermore, it has been shown that calibrating the model on a shorter time scale than the projection horizon is convenient in terms of reduced estimate error on $\hat{A}$. Analytical expressions for the error are provided in the Gaussian regime (i.e.: small variances of the latent variables) by theorem 4.3, while the case of increasing variance has been investigated numerically, confirming that in general the precision of the calibration is higher when historical data with shorter tenor are employed. It has been verified that the convenience of calibrating the model at short time scales remains also in presence of autocorrelation, although this is guaranteed only in the exponential framework, where an exact correction term is available.
Finally, the proposed techniques are shown to be numerically sound when applied to a real, publicly available, data set of Italian bad loan rates.

Chapter 5

Unsustainability risk of bid bonds in public tenders

This chapter analyzes the case when the conversion of a bid bond to the corresponding performance bond turns out to be "unsustainable" to the surety, implying the generation of a claim.

The topic has been formerly introduced in §2.1 (item *vi.* - "insensitivity to creditworthiness dynamics") as the only possible exception to the general rule that a C&S claim cannot be originated by the worsening of a risky subject's credit standing. Considering a realistic setup, in the following it is shown that the bidders' creditworthiness dynamics is a negligible threat to the sustainabiliy of a public tender process. On the other hand, a poor choice of the starting price by the procuring entity plays a major role in leading to unsustainable tender outcomes. Further, the chapter discusses how a well-designed risk appetite framework may prevent these situations at all.

Sustainability is a complex and evolving concept that may include, *inter alia*, economic and financial considerations, environmental and social impacts, political and legal aspects [137]. In public works, sustainability must be considered in its broadest meaning since all the aspects are relevant to the public entity that requires the execution of a project and the citizens who benefit from its fulfillment. However, also in this context, economic sustainability remains necessary to enable all the other possible forms of sustainability. Unfair pricing of the project implies a waste of public resources (when too high) or may result in a poor - or even missed - execution (when too low). Both cases have a negative impact, at least from a social perspective but possibly also environmentally, depending on the specific situation.

Typically, the cost of a public construction project is determined by a tender promoted by the procuring entity. Nowadays, each country disciplines bid mechanisms underlying public tenders by a complex regulatory framework that guarantees fairness among participants and financial protection to the procuring entities. The leading economies share the main features of respective public bid laws (see, *e.g.*, [10–12, 138]). In particular, a system of guarantees is usually mandatory and involves insurance companies and financial institutions as sureties [3, 43, 44]. As discussed in §1.3, each participant to a public tender must underwrite a bid bond to take part in the tender. The bid bond guarantees that the contract winner will satisfy all the requirements needed to become the contractor, including acquiring a performance bond. The values of the contract and the related performance bond are subjected to stochastic variations during the bidding process due to the tender rules. Hence, when the bid bond is issued, the surety has to consider the riskiness of the guaranteed participant concerning these variations and the sustainability of issuing a subsequent performance bond in case the bidder wins the contract.

There is a negative dependency between the final value of the contract and the notional value of the performance bond. This is because the public procuring entity wants to be protected against the performance risk of the winning contractor. Intuitively, the lower is the final performance cost, the higher is the probability of poor performance or other breaches of contract. Thus, the need for financial protection of the public entity increases accordingly. The final notional value of the performance bond may be too high for the risk appetite of the surety who has issued the bid bond of the winning bidder, significantly if the bidder's creditworthiness has worsened during the tender process. If no other surety is available to issue the performance bond, the bid bond generates a claim. The surety who has issued it has to indemnify the procuring entity on behalf of the contractor, who cannot be awarded the contract.

During the last two decades, the increasing need for developing and improving public infrastructures in many countries has renewed the research interest on various financial, economic, and legal topics related to the construction industry. In particular, the attention to risk management tools and techniques in public works and large private projects has considerably increased, leading to an intense research activity [23–25,33]. In this context, surety bonds have been investigated mainly empirically, with particular reference to performance bonds and to the benefits they produce in terms of risk mitigation. In these years, performance bonds have also been investigated with regards to their legal sustainability, depending on the specific regulatory framework of a considered country [13–15]. E.g., in countries where these instruments have been introduced recently, it is worth considering the moral hazard of beneficiaries who abuse their right to call on the surety guarantees. The relevance of the problem has been assessed, and possible improvements to specific national laws are presented in [13, 14]. Surety bonds have been investigated from an actuarial perspective as well, addressing both the problems of pricing them and measuring their mitigation effect on the underlying performance risk [26, 27]. Results obtained in [26] imply that surety companies can help to mitigate the problem of contractors going bankrupt by their ability to perform a preliminary screening. Further, in [27] it is shown that contractors with a better standing are more likely to win the tender if sureties apply a risk-adjusted price to the performance bond. However, to date, the literature has focused only on investigating the performance risk and the likelihood that the winning contractor defaults during the execution of the public works.

As stated above, this chapter investigates the case that the tender process may lead to an unsustainable outcome (*i.e.*, the performance bond is not issued, and the tender process has to be reopened). This is relevant since the inefficiency of the tender process implies costs for all the subjects involved: the bidders, the sureties, and the public entity. Italian bid law [46–48] is considered to specify the tender mechanisms (*e.g.*, the functional form that links the contract pricing and the notional value of the performance bond). A risk appetite framework is proposed to model the behavior of the sureties who support the bidders, based on the Solvency II regulatory framework [36–38].

The background information considered in the following is provided in §1.3, that introduces the business and legal context of the investigated problem. The main features of the suretyship insurance business are reported therein, with a focus on the bid bond and the performance bond insurance products. Further, the Italian bidding law for public works is summarized in §1.3.1.

The chapter is organized as follows. Section 5.1 models the sustainability of the tender outcome both for the bidders and the sureties. The Solvency II Standard Formula elements needed to design the surety's risk appetite framework are introduced, and the bidder's behavior is modeled considering their appetite for a minimum profit. Section 5.2 addresses the investigated problem. After introducing the distributional assumptions needed, we measure the probability of inefficient outcomes of public tenders through numerical simulations. The main results are summarized in Section 5.3.

5.1 Sustainability of a bid bond

In this section, the risk-adjusted economic sustainability of a bid bond is investigated. §5.1.1 presents the bidder's perspective, while the risk appetite of the surety is discussed in §5.1.2. The procuring entity is protected by the mechanism described in §1.3.1, which rules the whole bidding process. In particular, equation (1.3.1) fixes the notional value of the performance bond to cope with the performance risk of the winning bid. However, the tender process is sustainable for the procuring entity only if it leads to a sustainable outcome for both the contractor and the surety, such that the tender has not to reopen. Hence the sustainability conditions of all the subjects involved in the tender process are discussed in the following, explicitly or not.

5.1.1 The bidder's perspective

Let us assume that C is the contractor's cost to fulfill the principal obligation (*e.g.*, the construction of public infrastructures). Inflation effects on the price of raw materials are negligible through the period $(t_0, t_1]$ when the tender takes place. Further, all the bidders are supposed to have access to the same liquid market to get the needed workforce, materials, and instruments. Thus, C is supposed to be equal to all the competitors and independent from time. We can safely assume that C also includes a minimum target profit required by each contractor's stakeholders. Its value remains (approximately) the same for all the bidders, even considering this additional contribution.

In a bidder's perspective, V_1 has both natural upper and lower bounds. The upper bound $\overline{V}_1 \coloneqq V_0 > V_1$ holds true by construction, while the lower bound

$$\underline{V}_1 \coloneqq C + \pi_P \leq V_1 \tag{5.1}$$

is the amount needed by the bidder to cover the expected costs, the target profit, and the risk-adjusted price $\pi_P(d_{01})$ of the performance bond. Filtered at t_1, the surety is assumed to price the performance bond according to a typical non-life actuarial pricing form [28, 29], that accounts for expected losses, costs, and a prudential loading needed to compensate the surety's risk aversion, the latter also being the surety's profit. Hence, the price π_P obeys to the equation

$$\pi_P = p_{12} L_P N_P + (r+s)\pi_P, \tag{5.2}$$

where p_{12} is the breach/default probability of the contractor over the execution period $(t_1, t_2]$ and, thus, $p_{12} L_P N_P$ is the expected loss contribution. Indeed, $L_P \in (0,1]$ is the loss-given-default mitigation coefficient. It considers both the expected recovery of the surety from the contractor after the claim and the possible reduction of N_P at the claim time due to the partial fulfillment of the guaranteed obligation. The term $r\pi_P$ is the compensation for the risk aversion of the surety's stakeholders, where r scales as the cost-of-capital rate. The contract's price π_P is approximately proportional to the contract's contribution to the solvency capital requirement (SCR) needed by the insurer to guarantee its solvency in the Solvency II Standard Formula framework. Also the surety's costs are assumed to be proportional to π_P and are taken into account by the cost ratio s term. The discount contribution of non-zero risk-free interest rates is neglected. As discussed in §1.3.1, it holds

$$N_P = \alpha_P(d_{01}) V_1, \tag{5.3}$$

where also $V_1 = d_{01} V_0$ is dependent on d_{01}. Hence, the risk-adjusted price of the performance bond is

$$\pi_P(d_{01}) = \frac{p_{12} L_P}{1-r-s} \alpha_P(d_{01}) d_{01} V_0 \tag{5.4}$$

The price π_B of the bid bond follows the same structure and assumptions (see Remark 5.2 below).

Equations (5.1) and (5.4) imply that the public tender is *sustainable* for the winner of the tender, only if the inequality

$$\left[1 - \tfrac{p_{12} L_P}{1-r-s} \alpha_P(d_{01})\right] d_{01} \geq \frac{C}{V_0} \tag{5.5}$$

is verified. It is worth noticing that d_{01} is deterministic in the bidder's perpective, since it is a bidder's decision. On the other hand, p_{12} and r are unknown to the bidder, but a non-binding offer from the surety market is usually available on request, letting the bidder to consider the provisional price

$$\hat{\pi}_P = \tfrac{\hat{p}_{12} L_P}{1-r-s} N_P(\hat{d}_{01}) \tag{5.6}$$

where $\hat{p}_{12}$ is the bidder's probability of default in $(t_1, t_2]$ estimated by the surety conditioned to the information available in t_0, and $\hat{d}_{01}$ represents the expected value of d_{01} under the same filtration $\mathcal{F}_{t_0}$. Hence, the boundaries $\overline{V}_1$ and $\underline{V}_1$ imply the determination of a compact interval where the contractor's choice of d_{01} in t_0 is rational

$$\frac{C + \hat{\pi}_P}{V_0} \leq d_{01} \leq 1 \tag{5.7}$$

Condition (5.7) confirms two intuitive facts. First, the least risky bidder can afford to offer the greatest decrease of the starting price V_0, implying that the better is the creditworthiness, the higher is the probability of winning the tender. Second, the lower is the ratio C/V_0, the smaller is the minimum sustainable d_{01} value.

Remark 5.1. *LHS of inequality* (5.7) *is a special case of inequality* (5.5)*, conditioned to the information available in* t_0*. Even if each bidder behaves rationally, placing a bid* $d_{01}(t_0)$ *in the interval defined in condition* (5.7)*, in* t_1 *it is still possible that the winner of the tender is awarded with a not sustainable contract, because satisfying* (5.7) *in* t_0 *does not imply that* (5.5) *will be fulfilled in* t_1*. This uncertainty motivates the existence of prudential bids, that are greater than the minimum rational level* $\frac{C+\hat{\pi}_P}{V_0}$*.*

Remark 5.2. *The cost* π_B *of the bid bond is negligible in the framework introduced above. In fact, it holds* $\pi_B \ll \pi_P$*, because* $\alpha_B \ll \alpha_P$ *(see §1.3.1) and* $t_1 - t_0 \ll t_2 - t_1$*. Further, the bid bond generates a claim only if the insured bidder defaults and is the winner of the tender. Hence, assuming to know the number* N *of participants involved in the tender process and considering approximately equal probabilities of being awarded among participants, the price of the bid bond can be written as*

$$\pi_B = \frac{1}{N}\frac{p_{01}L_B}{1-r-s}\alpha_B V_0. \tag{5.8}$$

under the same assumptions considered for π_P *in equation* (5.4)*. As discussed above, the probabilities of winning are not uniform among the bidders, but such precise information is not available to a surety that guarantees just one of them in most cases. In general, considering the respective durations of bid bonds and performance bonds, and the* $1/N$ *factor as well, it holds*

$$\tfrac{1}{N}p_{01} \ll p_{12}, \tag{5.9}$$

that strengthens the validity of $\pi_B \ll \pi_P$*.*

The bid bond prices should be regarded more as "generic" expenses of the contractor than costs related to specific tenders, given that each contractor has to allocate a share of economic resources to participate in tenders, to win a part of them at most.

5.1.2 The surety's perspective

The subject who acts as the surety may be either a bank or an insurance company operating in the suretyship insurance line of business. We consider the latter case in the following, assuming that the Solvency II framework regulates the surety. This assumption copes with the investigated problem (*i.e.*, the sustainability of bid bonds in Italy - a country where Solvency II is applied to the insurance market).

According to the Solvency II Directive [36] (Article 44), each insurer must define a set of rules, known as *Risk Appetite Framework* (also RAF), that aims to limit the capital absorption level below a given fraction of the own funds. This concept is then implemented in the Italian insurance law as well [30]. Since the RAF should discipline the business strategy and the management actions, the problem of the efficient capital allocation among the insurer's lines of business has been widely investigated in the actuarial literature (see, *e.g.*, [31, 32] and references therein).

However, this work focuses on the sustainability of a specific suretyship contract. Hence, our interest in a surety's RAF is limited to the subset of rules that may limit the surety's risk appetite against *Premium Risk* and the related *Catastrophe Man-Made Risk* in the Solvency II Standard Formula framework. On the other hand, the maximum acceptable amount of capital absorbed by the Suretyship line of business is assumed to be fixed. Let us consider a (sub)portfolio composed of suretyship policies only. According to the Solvency II Standard Formula [37, 38], such a portfolio exposes the insurer to three risk components of the *Underwriting Risk*, briefly summarized below:

i. The *Premium Risk*, whose Solvency Capital Requirement (SCR) is measured as

$$\begin{aligned} \mathrm{SCR}_{\mathrm{Pr}} &:= 3\sigma_{\mathrm{Pr}} V_{\mathrm{Pr}}, \\ V_{\mathrm{Pr}} &:= \max\{P_{\mathrm{Next}}; P_{\mathrm{Last}}\} + FP_{\mathrm{Existing}} + FP_{\mathrm{Future}}; \end{aligned} \tag{5.10}$$

where P_{Last} and P_{Next} are the premiums earned in the last 12 months and the premiums to be earned in the next 12 months respectively, FP_{Existing} and FP_{Future} are the expected present value of the premiums to be earned after the following 12 months for existing contracts and for contracts whose initial recognition date falls in the following 12 months[1] respectively, and $\sigma_{\mathrm{Pr}} = 19\%$ is the coefficient of variation associated to this sub-module of risk by the European regulator. The geographical diversification factor is not considered in equation (5.10), since we are considering risks arising from Italian contractors only. The effect of reinsurance is ignored as well, for this risk component and the next two listed below.

ii. The *Catastrophe Recession Risk*, whose Solvency Capital Requirement (SCR) is measured as

$$\mathrm{SCR}_{\mathrm{Rec}} := P_{\mathrm{Next}} \tag{5.11}$$

iii. The *Catastrophe Default Risk*, whose Solvency Capital Requirement (SCR) is measured as

$$\mathrm{SCR}_{\mathrm{Def}} := \mathrm{lgd}(LE_1 + LE_2) \tag{5.12}$$

where LE_i $(i = 1, 2)$ are the first and the second largest exposures in the considered portfolio lgd $= 10\%$ is a loss-given-default coefficient fixed by the European regulator.

The Standard Formula aggregation rule for the risk components listed above is

$$\begin{aligned} \mathrm{SCR}_{\mathrm{Udw}} &= \left(\mathrm{SCR}_{\mathrm{Pr}}^2 + 2\rho \mathrm{SCR}_{\mathrm{Pr}} \mathrm{SCR}_{\mathrm{Cat}} + \mathrm{SCR}_{\mathrm{Cat}}^2\right)^{\frac{1}{2}}, \\ \mathrm{SCR}_{\mathrm{Cat}} &= \left(\mathrm{SCR}_{\mathrm{Def}}^2 + \mathrm{SCR}_{\mathrm{Rec}}^2\right)^{\frac{1}{2}}, \end{aligned} \tag{5.13}$$

where $\rho = 25\%$ and $\mathrm{SCR}_{\mathrm{Udw}}$ is the Underwriting Risk measure under the assumption that all the risk components different from $i.-iii.$ are null, as futher specified in the

[1] For future contracts, premiums earned during the first 12 months after the initial recognition date are excluded from FP_{Future} contribution to volume measure.

following Remark 5.3. It is worth recalling that additional elements of the Standard Formula framework defined for the C&S Line of Business are available in §2.3.

Remark 5.3. *Equation* (5.13) *measures only a part of the* SCR_{Udw} *that each suretyship insurance company has necessarily to cover. In particular, the Reserve Risk sub-module has been ignored, since the study reported in this chapter is focused on the growth of Premium Risk due to newly underwritten contracts, that is directly related to the sustainability of the new policies.*

*This simplification can be interpreted either as the assumption of instantaneous indemnifications (*i.e.*, the surety opens and immediately closes the reserve provision associated with each claim, keeping the Reserve Risk negligible) or as the assumption that the surety's RAF disciplines the Reserve Risk capital requirement separately from the Premium and Catastrophe risks. Indeed, the latter assumption is more likely than the first one.*

Lapse Risk is ignored since it is not considered relevant to this line of business.

Loosely speaking, in this context the risk measure $\mathrm{SCR}_{\mathrm{Udw}}$ scales approximately with the size of the future earned premiums that, according to equations (5.4) and (5.8), are proportional to the notional exposures ($N_B = \alpha_B V_0$ or $N_P = \alpha_P V_1$ in case of bid bonds or performance bonds, respectively) of each bond underwritten and to the corresponding claim probabilities ($\frac{1}{N} p_{01}$ or p_{02} respectively). Further, both N_B and N_P are proportional to the initial value V_0 of the contract and the performance bond exposure N_P has also a non-linear positive dependence on d_{01}, as shown in equations (1.3.1) and (5.3).

For the sake of simplicity, let us consider a stable or expanding business, so that

$$\max\{P_{\mathrm{Next}}; P_{\mathrm{Last}}\} = P_{\mathrm{Next}}. \tag{5.14}$$

The simplification introduced in equation (5.14) implies that equation (5.13) can be rewritten as follows

$$\mathrm{SCR}_{\mathrm{Udw}} = \sqrt{9\sigma_{\mathrm{Pr}}^2 (P+F)^2 + 6\rho\sigma_{\mathrm{Pr}}(P+F)\sqrt{D^2+P^2} + D^2 + P^2} \tag{5.15}$$

where we use the compact notation $D := \mathrm{SCR}_{\mathrm{Def}}$ and $P := P_{\mathrm{Next}}$. Equation (5.15) allows to estimate the marginal contribution $\delta\mathrm{SCR}_{\mathrm{Udw}}$ to the capital requirement originated by a newly underwritten policy

$$\begin{aligned} \delta\mathrm{SCR}_{\mathrm{Udw}}(P,F,D) &= \mathrm{SCR}_{\mathrm{Udw}}(P+\delta P, F+\delta F, D) - \mathrm{SCR}_{\mathrm{Udw}}(P,F,D) \\ &= C_P\,\delta P + C_F\,\delta F + \ldots, \end{aligned} \tag{5.16}$$

where

$$C_P := \frac{1}{\mathrm{SCR}_{\mathrm{Udw}}(P,F,D)}\left[9\sigma_{\mathrm{Pr}}^2(P+F) + 3\rho\sigma_{\mathrm{Pr}}\frac{D^2+P^2+P(P+F)}{\sqrt{D^2+P^2}} + P\right], \tag{5.17}$$

$$C_F := \frac{1}{\mathrm{SCR}_{\mathrm{Udw}}(P,F,D)}\left[9\sigma_{\mathrm{Pr}}^2(P+F) + 3\rho\sigma_{\mathrm{Pr}}\sqrt{D^2+P^2}\right], \tag{5.18}$$

δP is the new policy's contribution to P_{Next}, and δF is the new policy's contribution to $FP_{\mathrm{Existing}} + FP_{\mathrm{Future}}$. D is assumed to be constant, that is generally true, unless the new policy's exposure exceeds LE_2 in equation (5.12).

The premium accrual is linear in time, although the risk generated by the policy decreases as a non-linear function of the time-to-maturity. Hence it holds

$$
\begin{aligned}
\delta P &= \pi \frac{\min\{T-t,1\}}{T-t_0}, \\
\delta F &= \pi \frac{\max\{T-t-1,0\}}{T-t_0},
\end{aligned}
\tag{5.19}
$$

where π is the bond premium, t is the observation date, t_0 and T are the recognition date and the maturity date of the bond, respectively. Abrupt variations of $\mathrm{SCR}_{\mathrm{Udw}}$ may occur in case an insured bidder wins a tender and the surety issues the performance bond as needed. In this case, the bid bond premium π_B, is replaced by the corresponding performance bond premium $\pi_P \gg \pi_B$ (see Remark 5.2). The surety's RAF should aim to prevent that the exposure "jumps" associated with the conversions of bid bonds into performance bonds may lead to a breach of the established $\mathrm{SCR}_{\mathrm{Udw}}$ threshold level.

Many policy underwriters are simultaneously and independently operating on behalf of the surety. Hence, the contribution of each issued bond to $\mathrm{SCR}_{\mathrm{Udw}}$ cannot be taken into account instantaneously. In a realistic situation, $\mathrm{SCR}_{\mathrm{Udw}}$ is likely to be updated quarterly or twice a year, while new policies are issued daily or weekly. Hence, the surety may choose to maintain $\mathrm{SCR}_{\mathrm{Udw}}$ at a safe distance from a threshold $\overline{\mathrm{SCR}}_{\mathrm{Udw}}$ by defining a maximum acceptable $\delta\mathrm{SCR}_{\mathrm{Udw}}$ caused by each newly underwritten policy. Equations (5.4), (5.8), (5.16), and (5.19) imply that the maximum acceptable exposure $\overline{E}_p$ of a new policy scales as p^{-1}, where p is the claim probability of the policy. Namely, it holds

$$
\overline{E}_p = \frac{\overline{\delta\mathrm{SCR}}_{\mathrm{Udw}}}{C_P \frac{\min\{T-t,1\}}{T-t_0} + C_F \frac{\max\{T-t-1,0\}}{T-t_0}} \frac{1-r-s}{L} p^{-1}, \tag{5.20}
$$

where $\overline{\delta\mathrm{SCR}}_{\mathrm{Udw}}$ is the maximum acceptable variation of $\mathrm{SCR}_{\mathrm{Udw}}$ due to the new risk. C_P and F_P depend on the last updated values of SCR_{Udw}, P, F, and D. Terms beyond the first order in equation (5.16) are assumed to be negligible.

Two concerns should be addressed before using equation (5.20) to define a (simplified) surety RAF. First, if the contract is a bid bond, the case the contractor wins the tender and, thus, a performance bond is needed must be considered. This issue is addressed later in Definition 5.2. Further, sureties want to limit their concentration of exposure against each contractor. Thus, a penalty term due to existing exposures that the same contractor generates should be considered.

To address the latter issue, the threshold $\overline{\delta\mathrm{SCR}}_{\mathrm{Udw}}$ is lowered by the first-order contribution to SCR_{Udw} of the policies already underwritten by the same contractor. Applying equation (5.16) once again, we have the new threshold

$$
\overline{\delta\mathrm{SCR}}_i := \max\left\{0, \overline{\delta\mathrm{SCR}}_{\mathrm{Udw}} - C_P \sum_{j\in\{i\}_t} \delta P_{ij} - C_F \sum_{j\in\{i\}_t} \delta F_{ij}\right\} \tag{5.21}
$$

where $\{i\}_t$ is the sub-portfolio of policies existing in t and underwritten before t by the i-th contractor, and

$$\begin{aligned} \delta P_{ij} &= \pi_{ij}\frac{\min\{T^{(ij)}-t,1\}}{T^{(ij)}-t_0^{(ij)}}, \\ \delta F_{ij} &= \pi_{ij}\frac{\max\{T^{(ij)}-t-1,0\}}{T^{(ij)}-t_0^{(ij)}} \end{aligned} \tag{5.22}$$

are the contributions to P and F of the j-th policy in $\{i\}$, given the same notation used in equation (5.19). It is worth noticing that the new threshold can be equal to zero, in case the concentration level on the i-th contractor has already exceeded the surety's risk appetite.

To handle the first concern on equation (5.20), the surety's RAF can be defined as follows.

Definition 5.1 (Backward-looking Surety's RAF)**.** *The RAF is specified by the function $\Psi : (0,1) \to \mathbb{R}_+$, defined as follows. $\Psi(p_{iJ})$ is the maximum increment of exposure $\overline{\delta E_i}$ that the surety is allowed to guarantee against the i-th risky contractor ($|\{i\}_t| = J-1$), by issuing the new iJ-th bond whose claim probability is equal to p_{iJ}. $\Psi : p_{iJ} \mapsto \overline{\delta E_i}$ has the form*

$$\Psi(p_{iJ}) = \frac{\overline{\delta SCR_i}}{C_P\frac{\min\{T^{(iJ)}-t,1\}}{T^{(iJ)}-t_0^{(iJ)}} + C_F\frac{\max\{T^{(iJ)}-t-1,0\}}{T^{(iJ)}-t_0^{(iJ)}}}\frac{1-r-s}{L}p_{iJ}^{-1}, \tag{5.23}$$

Thus, the surety refuses to underwrite each J-th contract such that $\delta E_{iJ} > \Psi(p_{iJ})$.

The RAF in Definition 5.1 is *backward-looking* in the sense that the acceptance or rejection of a given contract depends only on the contribution of the contract to the last SCR measured. As anticipated, Definition 5.1 does not offer solution to the first issued raised above (*i.e.*: bid bonds which cope with Definition 5.1 may lead to performance bonds that exceed the frontier defined in equation (5.23) in the future). Definition 5.2 handles also this issue.

Definition 5.2 (Forward-looking Surety's RAF)**.** *The RAF is specified by the couple $\{\Psi(\cdot); p_\Psi\}$. The function $\Psi : (0,1) \to \mathbb{R}_+$ defines the maximum increment of exposure $\overline{\delta E_i}$ that the surety is allowed to guarantee against the i-th risky contractor ($|\{i\}_t| = J-1$), by issuing the new iJ-th bond whose claim probability is equal to p_{iJ}. $\Psi : p_{iJ} \mapsto \overline{\delta E_i}$ has the form stated in equation* (5.23)*. The tolerance $p_\Psi \in (0,1)$ is the maximum admissible probability that a risk underwritten in t implies a breach of the boundary $\{p_{i'j'}, \overline{\delta E_{i'}}\}$ at some $t' > t$ for all i', j'. Namely, the iJ-th bond can be underwritten in t only if*

$$\mathbb{P}\left[\Psi\left(p_{i'j'}(t')\right) < E_{i'}(t')\middle|\, \mathcal{F}_t\right] < p_\Psi \quad \forall t' > t,\, i', j' \in \mathbb{N}. \tag{5.24}$$

where $\mathbb{P}$ is real-world probability measure available to the surety. A bond that satisfies both conditions (5.23) *and* (5.24) *by t is* sustainable *in the surety's perspective.*

The Solvency II Standard Formula is based on some simplifying assumptions that also affect equation (5.23).

First, premiums as a volume measure establish a link between the riskiness of each risk source (*i.e.* the contractor in this case) and the capital requirement.

However, in case the premium rate is fixed at the issuing date t_0, it is related to the contractor's standing at t_0, but it may not be representative of the contractor's riskiness when SCR_{Udw} is evaluated.

Further, the fraction of premium to be earned by the surety decreases linearly over time. Hence, the negative dependence between risk and residual time-to-maturity is taken into account. However, the non-linear decreasing of risk by time - as shown, *e.g.*, in equation (5.6) - is replaced by a linear dependency.

Despite these limitations, the Standard Formula represents a breakeven point between simplicity and effectiveness. Being an established standard in the European insurance industry, it is worth considering it when defining the RAF used to investigate the sustainability of a given bid bond. Internal model approaches are possible as well and are not affected by the limitations mentioned above. However, in this chapter, we are interested in investigating possible paradoxes arising in a standard context. Hence we chose to use the Standard Formula exclusively.

Definition 5.1 introduces a maximum exposure-at-risk $\overline{E}_i$ per contractor, implicitly. The value of $\overline{E}_i$ depends on the standing of the i-th contractor and the remaining time-to-maturity of each underwritten contract, in agreement with intuition. On the other hand, Definition 5.2 also forbids less trivial cases.

A bid bond that satisfies condition (5.23) in t_0 may still not comply with condition (5.24), in case it holds

$$\mathbb{P}\left[\Psi^{-1}\left(E_{iJ+1}(t_1)\right) < p_{iJ+1}(t_1) \middle| \mathcal{F}_{t_0^{(iJ)}}\right] > p_\Psi, \tag{5.25}$$

where the bid bond is the J-th policy underwritten with the i-th contractor and the subsequent performance bond (in case the contractor wins the tender) is the $J+1$ policy.

It is worth noticing that equation (5.24) implies restrictions stronger than the one stated in equation (5.25). Let us consider the iJ-th bid bond mentioned above, assuming that it satisfies condition (5.23) and does not have the problem in equation (5.25). Evenu in such a case, the bond could still not satisfy condition (5.24) due to portfolio issues. In fact, when the number of simultaneously active bid bonds is large enough, the probability that one of them results in a future performance bond not compliant with condition (5.23) exceeds p_Ψ, even if the last underwritten bid bond complies with the RAF when considered stand-alone.

The "global" sustainability of a bid bond (*i.e.* in the context of the surety's portfolio of underwritten bonds) is addressed in the next §5.2 numerically.

5.2 Measuring and managing the unsustainability scenarios in public tenders

This section addresses the sustainability issues introduced in §5.1 by implementing the surety's RAF proposed in Definitions 5.1 and 5.2. A model is introduced in §5.2.1 to simulate the tenders. Each simulation considers three alternate versions of the surety: without a RAF, adopting a backward-looking RAF as per Definition 5.1, and adopting a forward-looking RAF as per Definition 5.2. The results obtained by the Monte Carlo simulations are presented in §5.2.2, §5.2.3, and §5.2.4, respectively.

5.2.1 Simulation of tenders in a surety's perspective

In the following, the model employed to simulate the tenders of public works is described. The model aims to highlight that unsustainable requests for a performance bond are possible when considering realistic dynamics of default (or breach) probabilities associated with each bidder. Further, the model is employed to investigate the effectiveness of the RAF strategies implemented by the surety.

A realistic model of the Italian public works market should include some aspects not considered in the following, such as the actual number of tenders per year where the considered surety guarantees at least a bidder and the distribution of the public works costs C. However, data needed to calibrate such a model are non publicly available, and the model itself would not fit better for out purpose than the toy model introduced here.

Let us consider a discrete-time framework where the elementary time step δt is a quarter long. We consider a surety with access to 10^3 public tenders per quarter, issuing a bid bond to at least a participant per tender since there is no need to simulate tenders where the considered surety has no business. We assume that the initial price of each tender is a uniform r.v.

$$V_0 \sim \mathrm{Unif}_{[C,\,150\%C]}. \tag{5.26}$$

Both the boundaries $\min V_0$ and $\max V_0$ are admitted to represent a possible misjudgment of the procuring entity. In fact, $V_0 = C$ leaves no room to lower the initial price or to aim for an extra profit, implying that no bidder is joining the tender. Further, in case $V_0 = 150\%\,C$, equation (1.3.1) implies that E_P/E_B can reach a value of 45 and above, increasing the probability that the required performance bond violates equation (5.23) and, thus, that sureties reject the (unsustainable) winner's request for a performance bond.

Thus, it is natural assuming a positive dependency between V_0/C and the number of potential bidders $\tilde{N}$ interested in joining the tender. We choose the form $\tilde{N} = \left\lfloor 100 \left(e^{V_0/C-1} - 1 \right) \right\rceil$ that implies a realistic range $\tilde{N} \in \{0 = N_{V_0=C}, \ldots, N_{V_0=1.5\cdot C} = 65\}$. However, the $N \leq \tilde{N}$ constructors who actually join the tender are the ones able to make a bid according to the condition (5.7), depending on the values of $\hat{\pi}_P$ and V_0.

The surety can issue more than a bid bond per tender (up to N), increasing the probability that one among its insured bidders wins the contract and is required to underwrite the corresponding performance bond. However, considering the competition in the surety market, we assume that the number of bidders n joining the same tender and insured by the same surety is distributed as a shifted Poisson r.v.. Namely, $\tilde{n} - 1 \sim \mathrm{Pois}(\lambda_B)$ where $\lambda_B = 0.1$ and $n = \min\{\tilde{n}; N\}$.

The parameter α_B is assumed to be distributed as a categorical r.v. with probability mass function

$$f(\alpha_B) = \begin{cases} 0.2, & \alpha_B \in \{1\%; 3\%; 4\%\}; \\ 0.4, & \alpha_B = 2\%; \\ 0.0, & \text{otherwise.} \end{cases} \tag{5.27}$$

where the mode is fixed at 2%, as anticipated in §1.3.1.

To estimate claim probabilities and their dynamics, we consider historical time series of performing ("PL") and non-performing loans ("NPL") [126], publicly available from the Bank of Italy [112]. This choice is justified by the assumption that the claim probability of a contractor is completely correlated to its creditworthiness. This is true in the extreme case of bankruptcy, which implies the contractor's inability to be operating. In general, it is a fair approximation, although other elements of technical nature (*e.g.*, unforeseen geological features of the building location) may contribute to the performance risk in specific cases.

Time series PL_t and NPL_t are quarterly available by ATECO 2007 economic sector (*i.e.*, our data are restricted to the "constructors" sector), size of loan s (3 clusters) and geographical location g of the Italian debtor (5 clusters). Hence, dynamics of claim probability can be specified by considering 15 bivariate time series $\{\mathrm{PL}_t; \mathrm{NPL}_t\}_{sg}$, where PL_t is the number of performing loans at the first day of the t-th quarter, while NPL_t is the number of loans become past due during the t-th quarter.

Since we need to represent a significant number of contractors by introducing a parsimonious number of parameters, we choose to apply the CreditRisk$^+$ model [133, 134] to describe the dependence structure among the claims and the marginal volatility of each cluster probability of default. The CreditRisk$^+$ model defines the dependence among defaults (or other absorbing events, such as breaches of contracts) through an array of latent market factors $\mathbf{\Gamma} \in \mathbb{R}_+^K$ where $\Gamma_k \sim \mathrm{Gamma}(\sigma_k^{-2}, \sigma_k^2)$, $(k = 1, \dots, K)$. It holds $\mathbf{E}[\Gamma_k] = 1$ and $\mathrm{Var}[\Gamma_k] = \sigma_k^2$ by construction. The market factors alter the parameter's value of the r.v. $Y_i(t, t')$, that represents the occurence of a claim generated by the i-th contractors in the time interval $(t, t']$. In its original formulation, the model [134] is defined in a single-time-scale framework and $Y_i \dot{\sim} \mathrm{Pois}(p_i)$, where

$$p_i(\mathbf{\Gamma}) := q_i \cdot \left(\omega_{i0} + \sum_{k=1}^{K} \omega_{ik} \Gamma_k \right) \tag{5.28}$$

and the factor loadings ω_{ik} are supposed to be all non-negative and to sum up to unity:

$$\begin{aligned} &\omega_{ik} \geq 0, \qquad i = 1, \dots, N, \quad k = 0, \dots, K, \\ &\sum_{k=0}^{K} \omega_{ik} = 1, \quad i = 1, \dots, N. \end{aligned} \tag{5.29}$$

We consider the model's generalization discussed in chapter 4 and [103], which is also proposed for applications to Credit insurance [104]. The advantage of this choice is the possibility to calibrate the model by using the quarterly time series available and using it to estimate both the bid bond claim probabilities $p_{01}(t, h(i))$ and the corresponding performance bond claim probabilities $p_{12}(t, h(i))$, where $h = 1, \dots, 15$ labels the cluster of the i-th bidder.

We assume that each bid bond has a 3 months coverage period on the interval $(t = t_0, t_1 = t + \delta t]$, while each performance bond has a 5 year coverage period on the interval $(t_1, t_1 + 20\delta t]$. This simplification is a part of the toy framework that we are defining since public works duration depends on each project's features and size. However, it is numerically sound since a tender process takes a few months to close, while a public works project typically lasts a few years. Hence, a two-time-scales parameterization is needed to price both the bid and the performance bond.

As discussed in §2.2.6, §4, and [103], the claim event in CreditRisk$^+$ framework can be modeled as $Y_i(t,t') \sim \text{Bernoulli}(p_i(t,t'))$, where the parameter p_i has an exponential dependency on the latent factors. Namely, under our set of assumptions it holds

$$p_{01}(t_0,h) = 1-\exp\left[-q_h\left(\omega_{h0}+\sum_{k=1}^{K}\omega_{hk}\Gamma_k(t_0)\right)\right], \tag{5.30}$$

$$p_{12}(t_1,h) = 1-\exp\left[-20q_h\left(\omega_{h0}+\frac{1}{20}\sum_{\tau=0}^{19}\sum_{k=1}^{K}\omega_{hk}\Gamma_k(t_1+\tau\,\delta t)\right)\right], \tag{5.31}$$

where $\tau \in \mathbb{N}$ is the index used to label each quarter. Further, assuming that the surety has developed an internal rating model such that reliable estimates of

$$\{\mathbf{\Gamma}(t) : t = t_0,\ldots,t_0+19\delta t\}\,|\mathcal{F}_{t_0},$$

we can represent the possible fluctuations in performance bond pricing by considering $\hat{p}_{12} = p_{12}\,(t_0,h(i)) = p_{12}\,(t_1-\delta t,h(i))$, while the (correct) estimate p_{12}, that allows the computation of π_P, shall not be available until $t=t_1$.

Let us consider the time series $\{\text{PL}_t;\text{NPL}_t\}_{sg}$ available in [112], from 1^{st} quarter of 2016 to 1^{st} quarter of 2021, to calibrate the CreditRisk$^+$ model. The generalized covariance estimator introduced in equation (4.37), the decomposition technique introduced in [102], and the standard regularization technique described in [127] return the result in table 5.1.

Remark 5.4. *Claims are not simulated in our setting, although the CreditRisk$^+$ framework is explicitly designed to do it. In fact, they are not relevant to the part of the surety's RAF addressed in this analysis. Occurred claims affect mainly the reserve provision and the Reserve Risk capital requirement. On the other hand, they may lead to a slight decrease of SCR_{Cat} or SCR_{Pr}, since claims generated by suretyship insurance products are* absorbing events. *Each policy may generate one claim at most during the coverage period, implying the zeroing of both the corresponding exposure and future premiums (if any).*

The framework is completed by associating each i-th bidder to its $h(i)$-th cluster. This is achieved by modeling the position of the public works underlying each tender as a categorical random variable. The probability associated with each g-th area ($g=1,\ldots,5$) is proportional to the number of performing borrowers belonging to that area observed in the construction sector by the 1^{st} quarter of 2021. That is equivalent to assume a correspondence between demand and offer in this economic sector (*i.e.*, the presence of many constructors implies a relevant number of public tenders and *vice versa*). Assuming that all the bidders belong to the same area where the public works must be executed, their distribution among the three loan classes is modeled in a similar way, considering a categorical variable per geographical area g, where probabilities are proportional to the number of performing borrowers observed in the cluster sg ($s=1,2,3$), conditioned to g.

s	g	h	q_h	**k** =	**0**	**1**	**2**	**3**	**4**	**5**
1	North-West	1	0.0050	$\omega_{hk} =$	0.516	0.246	0.025	0	0.212	0
1	South	2	0.0095		0.473	0.220	0.019	0.216	0	0.072
1	Islands	3	0.0098		0.559	0.231	0	0.210	0	0
1	North-East	4	0.0040		0.658	0.254	0.088	0	0	0
1	Center	5	0.0079		0.026	0.234	0.070	0.432	0	0.238
2	North-West	6	0.0079		0.509	0.228	0	0.028	0.235	0
2	South	7	0.0137		0.651	0.292	0.057	0	0	0
2	Islands	8	0.0158		0.455	0.266	0	0.036	0.237	0.005
2	North-East	9	0.0073		0.628	0.281	0.086	0.006	0	0
2	Center	10	0.0126		0.590	0.281	0.057	0.016	0	0.056
3	North-West	11	0.0130		0.091	0.235	0	0.134	0.443	0.096
3	South	12	0.0147		0.467	0.236	0.048	0	0.127	0.122
3	Islands	13	0.0185		0.619	0.333	0	0	0	0.049
3	North-East	14	0.0134		0.611	0.265	0.050	0	0.068	0.005
3	Center	15	0.0178		0.325	0.251	0.099	0	0	0.326
				$\sigma_k^2 =$		2.417	0.157	0.052	0.049	0.044

Table 5.1. The complete set of parameters $\hat{\Omega}, \hat{\boldsymbol{\sigma}}_\Gamma^2$ necessary to specify the dependence structure in CreditRisk$^+$ model applied to the Italian "*Constructors*" economic sector.

It is worth recalling that the public works tenders in Italy can be classified as first-price, sealed descending bid auctions[2]. The bid domain is compact, and the winner is the author of the lowest bid in a set of non-identically distributed bids. Hence, we cannot use the Fisher-Tippett-Gnedenko theorem [141, 142], which is commonly applied to model the distribution of the winning bid in the ascending bid auctions (see, *e.g.*, the recent paper [143], where a Weibull distribution is considered). Thus, we need to determine the winning bid numerically, considering that each rational bidder chooses its d_{01} according to condition (5.7).

A non-uniform distribution is assumed over each i-th domain, to take into account the appetite of each bidder for obtaining the contract. Namely, it holds

$$\frac{d_{01}V_0 - C - \hat{\pi}_P(i)}{V_0 - C - \hat{\pi}_P(i)} \sim \text{Beta}(\alpha, \beta) \tag{5.32}$$

where multiple specifications have been tested for the parameters set (α, β)as represented in figure 5.1: mode[pdf(d_{01})] tends to $(C + \hat{\pi}_P)/V_0$ at increasing bidder's appetite for winning the tender, while it tends to V_0 at increasing appetite for profit. However, no relevant effect of the (α, β) choice is observed on the results presented in §5.2.2-5.2.5. Hence, only the choice $(\alpha, \beta) \equiv (1.80, 7.20)$ is considered in figures 5.2-5.5.

[2]For complete classification of auctions and a deep theoretical discussion, see, *e.g.*, [139] and [140].

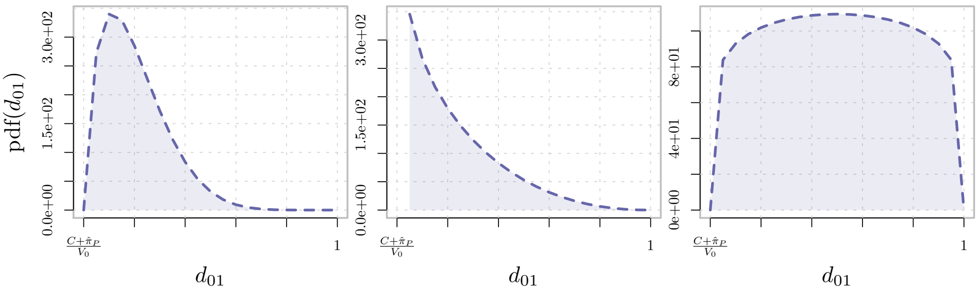

Figure 5.1. Different behaviors of the bidders modeled by alternative parameterizations of the Beta distribution in equation (5.32). *Left panel*: $\alpha = 1.80$, $\beta = 7.20$. *Central panel*: $\alpha = 0.80$, $\beta = 3.20$. *Right panel*: $\alpha = 1.16$, $\beta = 1.16$.

The winner of each tender is the lowest simulated bid per tender/scenario. The bidders are indexed in simulations ($i = 1, \ldots, N$). Doing so, the case when the winning bidder is among the ones guaranteed by the considered surety is explicitly represented.

5.2.2 Dynamics of the capital requirement without taking management actions

Let us consider a suretyship insurance company that operates as a surety in the framework introduced in §5.2.1. The surety is supposed to start operating in $t = 0$. It is worth recalling that the duration of a bid bond is established to be equal to three months, while each performance bond is assumed to expire after five years.

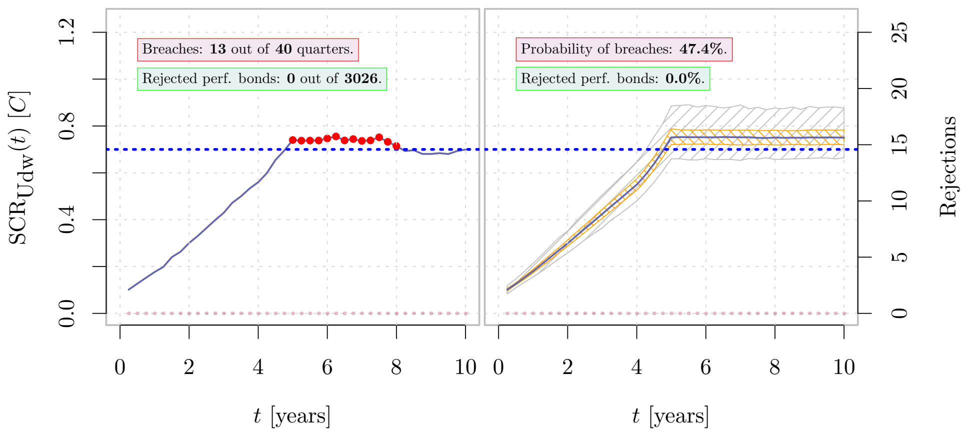

Figure 5.2. Dynamics of $\mathrm{SCR}_{\mathrm{Udw}}(t)$. *Left panel*: example of Monte Carlo simulation (single scenario), where the horizontal dashed line represent the $\overline{\mathrm{SCR}}_{\mathrm{Udw}}$ threshold that is fixed in the RAF, while red dots corresponds to the simulated breaches of $\mathrm{SCR}_{\mathrm{Udw}}(t)$ above the $\overline{\mathrm{SCR}}_{\mathrm{Udw}}$ level. Red columns at 0 level (right y-axis scale) represent the count of rejected performance bonds per quarter (zero since no management action is taken). *Right panel*: simulated distribution of $\mathrm{SCR}_{\mathrm{Udw}}(t)$ (10^3 Monte Carlo scenarios). The median (solid line), $0.25 - 0.75$ quantiles (orange area), and $0.01 - 0.99$ quantiles (grey area) are plotted. Red columns at 0 level (right y-axis scale) represent the average number of rejected performance bonds per quarter.

Thus, as expected, the surety $\mathrm{SCR}_{\mathrm{Udw}}$ - as defined in equation (5.15) - reaches an equilibrium after five years, considering a stable flow (on average) of new contracts per year (figure 5.2).

Without loss of generality, we choose $\overline{SCR}_{\mathrm{Udw}} = 0.7C$, that is below the equilibrium level $\mathrm{SCR}_{\mathrm{Udw}}(t > 5) \approx 0.75C$ obtained numerically. The surety is supposed to increase its sales volume until its risk appetite level is reached. Then, a risk appetite framework is introduced to discipline the underwriting process, as discussed in §5.1.2. Hence, in a liquid market, a surety with a higher risk appetite or a larger amount of available own funds than the one considered in our simulations would reach the same equilibrium state at a different $\mathrm{SCR}_{\mathrm{Udw}}$ level.

5.2.3 Dynamics of the capital requirement adopting a backward-looking RAF

In §5.2.2 the surety has reached an equilibrium state that is slightly above its risk appetite level. Thus, a RAF is needed to prevent the occurrence of breaches $\mathrm{SCR}_{\mathrm{Udw}}(t) > \overline{\mathrm{SCR}}_{\mathrm{Udw}}$. The same simulations presented in figure 5.2 are re-performed, applying the management actions implied by Definition 5.1, *ceteris paribus*.

The level $\overline{\delta\mathrm{SCR}}_{\mathrm{Udw}}$ is needed to specify $\Psi(\cdot)$. It has to be as high as possible to refuse the minimum number of contracts per unit of time, conditioned to avoid breaches or, at least, make them improbable enough according to the surety's risk appetite. In the example we chose

$$\overline{\delta\mathrm{SCR}}_{\mathrm{Udw}}(t) = \max\left\{0; \min\left\{\overline{\mathrm{SCR}}_{\mathrm{Udw}} - \mathrm{SCR}_{\mathrm{Udw}}(t-\delta t); \delta\mathrm{SCR}_{\mathrm{Udw}}\left[\Phi^{-1}_{\pi_P}(0.95)|\mathcal{F}_t\right]\right\}\right\}$$

where $\Phi_{\pi_P}(\cdot)$ is the cumulative distribution function of the performance bond prices π_P and $\delta\mathrm{SCR}_{\mathrm{Udw}}\left[\cdot|\mathcal{F}_t\right]$ is the marginal contribution of a given contract underwritten in t to $\mathrm{SCR}_{\mathrm{Udw}}$. $\delta\mathrm{SCR}_{\mathrm{Udw}}$ is evaluated by applying the linear approximation stated in equation (5.16).

Namely, when the last measure $\mathrm{SCR}_{\mathrm{Udw}}(t-\delta t)$ done until t is far enough from the threshold $\overline{\mathrm{SCR}}_{\mathrm{Udw}}$, we aim not to reject more than 5% of the performance bond requested by the insured bidders who win their respective tenders. In case the distance $\overline{\mathrm{SCR}}_{\mathrm{Udw}} - \mathrm{SCR}_{\mathrm{Udw}}(t-\delta t)$ approaches zero or negative values, the RAF constraint become stronger up to block the acquisition of new contracts at all, until an acceptable $\mathrm{SCR}_{\mathrm{Udw}}$ level is restored. Figure 5.3 shows the effectiveness of this approach. Breaches are observable in tail scenarios, almost only in the region ($t \simeq 5$) where the SCR regime is changing from *expansion* to *equilibrium*. The small number of breaches and the RAF's reaction implies a fraction of performance bonds rejected slightly above the 5% target level.

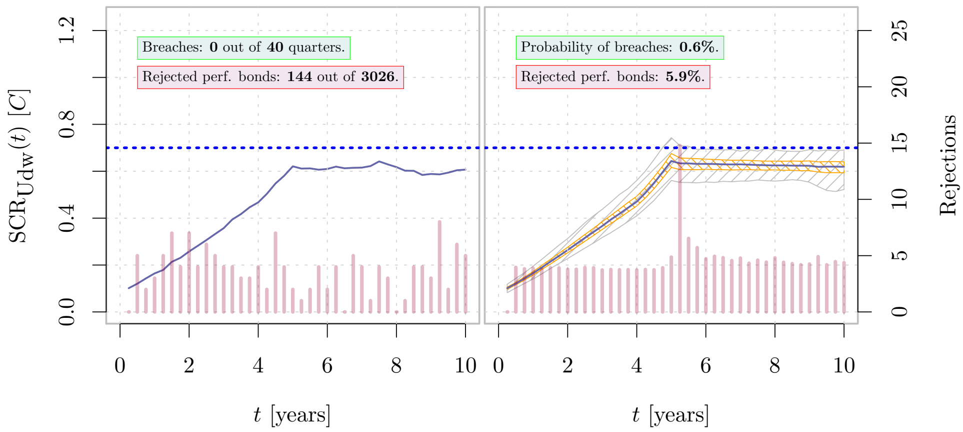

Figure 5.3. Dynamics of $\mathrm{SCR}_{\mathrm{Udw}}(t)$, given the notation introduced in figure 5.2. RAF introduced in Definition 5.1 is applied to estabilish the management actions taken.

5.2.4 Dynamics of the capital requirement adopting a forward-looking RAF

As discussed above, the surety should avoid the rejection of a performance bond as much as possible. This case opens the possibility that the tender must be reopened if no other surety is willing to issue the performance bond instead of the surety that has issued the bid bond to the winning contractor. Further, in a market where sureties are comparable, the rejection of a request by a company implies that the other companies in the same market are likely to do the same, leading to a claim generated by the bid bond.

Hence, it is worth addressing this issue by implementing the forward-looking RAF introduced in Definition 5.2. It is not necessary to establish p_Ψ explicitly. The probability that a performance bond generates a breach is an increasing function of V_0/C. Loosely speaking, a higher starting price implies that the winning bid - always close to C - corresponds to a greater discount $1 - d_{01}$. Thus, a higher α_P can be expected as well, increasing the probability of a breach $\overline{\delta \mathrm{SCR}}_{\mathrm{Udw}}$.

Even without knowing the analytical form of the dependencies described above, the qualitative picture is enough to implement the constraint (5.24) as

$$\Phi^{-1}_{V_0/C}(0.75) > \left(\frac{V_0}{C}\right)_k \tag{5.33}$$

where $\Phi_{V_0/C}(\cdot)$ is the cumulative distribution function of $\frac{V_0}{C}$ and $\left(\frac{V_0}{C}\right)_k$ is the ratio observed in the k-th tender. The percentile 0.75 has been chosen numerically, with the aim of minimizing both the number of rejected performance bond requests and the frequency of SCR breaches. Bid bonds that do no cope with equation (5.33) are rejected, preventing a possible unsustainable request for a performance bond (in case the bidder wins).

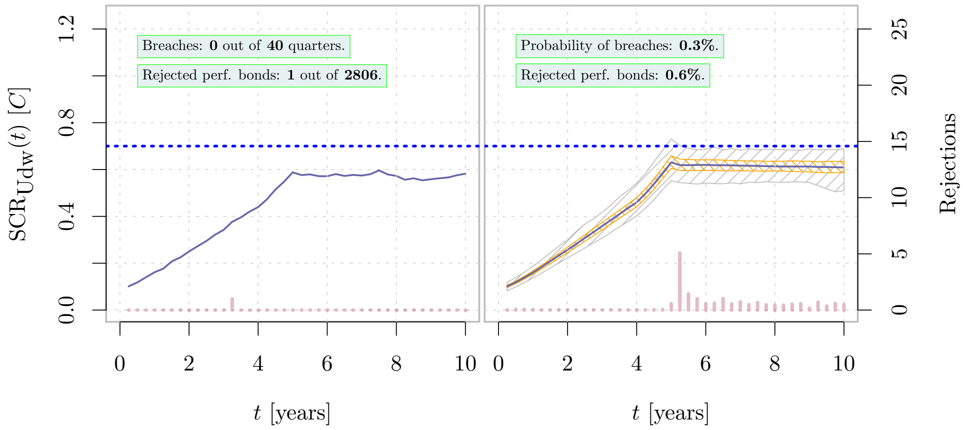

Figure 5.4. Dynamics of $\mathrm{SCR}_{\mathrm{Udw}}(t)$, given the notation introduced in figure 5.2. RAF introduced in Definition 5.2 is applied to estabilish the management actions taken.

Given the additional constraint, we can weaken the other introduced in §5.2.3, passing from 0.95 to 0.99 (*i.e.*, we aim to reject up to 1% of performance bonds, to limit both the claims arising from the corresponding bid bonds and the surety's reputational risk). Results are exposed in figure 5.4, where the number of rejected performance bonds decreases.

5.2.5 The role of the procuring entity

As shown in §5.2.4, a surety that implements a forward-looking RAF can prevent the majority of unsustainable tender outcomes, avoiding rejecting performance bond requests from the winners and the possible subsequent need for reopening the tender.

However, unsustainability issues originate from a poor choice of starting price by the public procuring entity. Left panel of figure 5.5 shows how a starting price near the breakeven level (*i.e.* $V_0 \simeq C$) disincentives constructors to join the bid, due to the constraints introduced in condition (5.7). It is worth noticing that this result is independent of the assumption made about the dependency $\tilde{N}(V_0)$. A tender that does not attract participants is clearly unsustainable from an economic perspective. The resources invested in promoting it are wasted, and the public works cannot be executed. Further, also the opposite case (*i.e.* $V_0 \gg C$) implies the economic unsustainability of the tender, as the tender outcome implies $N_P/N_B \gg 1$ and, thus, an excessive risk for the surety (*i.e.* an unsustainable cost for the winning bidder or the inability to underwrite the - mandatory - performance bond). The right panel of figure 5.5 shows the results of our simulations in this perspective: the fraction of requests for a performance bond rejected by given surety increases from $\approx 0\%$ to $\approx 50\%$ as V_0 passes from $\approx 1.25C$ to $\approx 1.45C$.

Remark 5.5. *Figure 5.5 shows that the tender is almost surely sustainable, depending on the procuring entity's proper choice of the* V_0 *value. This fact implies that the case* $\pi_P \gg \hat{\pi}_P$ *due to the worsening of a bidder's creditworthiness during the tender process has a negligible impact. The numerical evidence copes with the* *intuition: since the tender lasts a few months, a relevant change in a bidder's credit standing is unlikely during such a short period.*

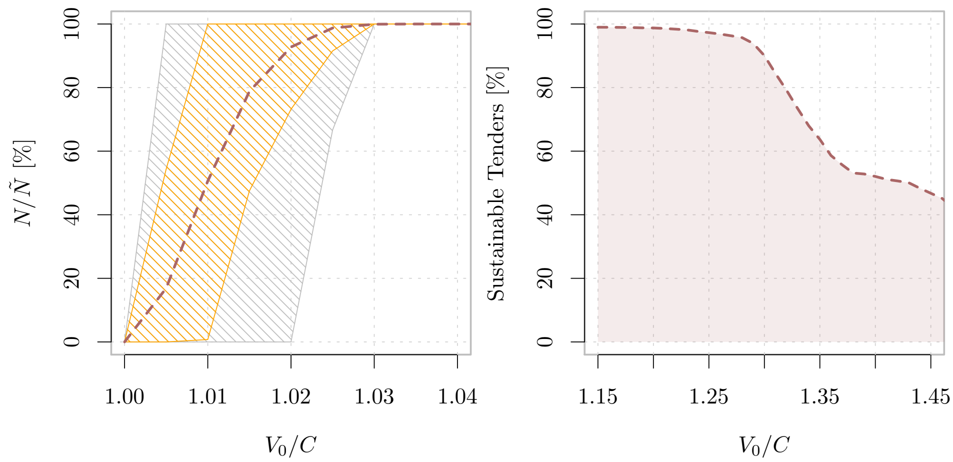

Figure 5.5. Sustainability as a function of the starting price V_0 *Left panel*: fraction of constructors who can afford to make a bid, as the starting price approaches the breakeven level C (*dashed line*: average; *orange area*: confidence interval within ± 1 Standard Deviation; *grey area*: confidence level within 0.01-0.99 quantiles). *Right panel*: fraction of concluded tenders whose requested performance bond is not rejected by the surety (average level) at increasing starting prices.

5.3 Summary

This chapter highlights the existence of an economic unsustainability risk for the tenders that award public works contracts, as suggested above in §2.1. The Italian bidding law and the Solvency II regulatory framework are explicitly considered to model the behavior of the three subjects involved in the tender process: the procuring entities, the bidders, the sureties. Numerical simulations show that this risk can be mitigated and prevented by the proper choices of both the surety and the procuring entity.

In §5.1 sustainability conditions are stated for both the bidders and the sureties. In particular, sureties can protect their SCR target levels by applying a RAF, as requested by the Solvency II Directive. In §5.1.2 we propose two simplified RAFs, both based on the linearization of the Solvency II Standard Formula. The first one (*i.e.*, "backward-looking" - Definition 5.1) aims only to protect the surety's SCR level, regardless the effects on the tender process. A boundary $(p; \overline{E}_p = Kp^{-1})$ is shown to separate sustainable new exposures $E \leq \overline{E}_p$ from the unsustainable ones ($E > \overline{E}_p$), depending on the default probability p associated to the considered contractor. A closed-form expression for K is provided from the Standard Formula prescription to evaluate the Non-life Underwriting risk module for the S2LoB 9 - *Credit & Suretyship Insurance*. We show that the sureties can protect both themselves and the tender process by applying a "forward-looking" RAF, like the one proposed in Definition 5.2. A numerical comparison among the two RAFs is presented in §5.2.3 and §5.2.4, showing that the latter can actually protect all the three subjects involved in the tender process.

While the surety can mitigate the unsustainability risk, the public procuring entity can prevent it by establishing an acceptable starting price for the tender. Our simulations suggest that $V_0/C \in [1.05, 1.25]$ is the most sustainable choice in a realistic setting, where the Italian constructors' default probability is modeled by applying the CreditRisk$^+$ framework to recent historical data.

The framework proposed in this chapter can be further investigated and improved in future studies. In particular, the Standard Formula approach can be replaced by a Partial Internal Model to define the two RAFs. The specific features of other regulatory frameworks diverse from the Italian bidding law may be investigated as well, provided that historical information to calibrate the model is publicly available for each considered country. Further, two limitations of this study are reported in the following. They can be addressed in further studies as well. First, both the procuring entity and the surety are assumed to perform error-free estimates. The procuring entity could actually choose V_0 poorly because of a bad strategy or the error affecting its C measure. The surety may perform a poor estimate of p_{12} as well, implying the missed identification of an unsustainable tender outcome due to a mistaken π_P evaluation. A second limitation of the study arises from the simplifications made in simulating the solvency balance sheet of the surety. Our model could consider a dynamical reserve risk and a non-zero market risk (generated by the surety's assets) in addition to premium and catastrophe risk components to provide a more realistic representation. Although future studies can adequately address these limitations, it is worth noticing that our simplified framework is consistent with the features that the considered system should have according to [26, 27], as summarized in the chapter's introduction. These limitations do not diminish the practical conclusions of the study about the strategies that the surety and the procuring entity can implement to mitigate the investigated unsustainability risk - namely, the adoption of the RAF introduced in Definition 5.2 and the choice of $V_0/C \in [1.05, 1.25]$, respectively.

Finally, the results reported in this chapter support the conclusions of chapter 2. In fact, considering the setup outlined above in §5.2.1, the numerical evidence reported in §§5.2.2−5.2.5 suggests that the main cause of unsustainable tender outcomes - and related claims as well - is the tender pricing misspecification, as highlighted in Remark 5.5. On the other hand, in the extreme case of a "blind" procuring entity (as assumed in equation (5.26) in our numerical framework), the contribution to bid bond claim probability due to unsustainability can be almost fully prevented by a surety who follows a proper underwriting strategy, as shown in §5.2.4.

In this picture, there is no evidence of non-negligible contribution to the claim probabilities coming from a realistic bidders' creditworthiness volatility, as the one simulated considering parameters in Table 5.1. Hence, the conclusions reported in §2.2.7 are confirmed: the choice to model C&S claim probabilities without modeling any explicit creditworthiness dynamics seems appropriate.

Chapter 6

Information asymmetry in Credit Insurance

Estimating the frequency and probability associated with the occurrence of absorbing events is a central issue in finance, medicine and actuarial science [90–94,98,100]. Human death, bankruptcy, and breach of undertaking are relevant examples.

Over the past century, frequency estimation has also been addressed in cases where only incomplete information is available. This type of problem highlights the limits of the classical frequency estimator, that is, the maximum likelihood estimator for the Bernoulli distribution parameter. The seminal papers of Kaplan and Meier [91] and Cutler and Ederer [92] introduced two frequency estimators designed to handle data incompleteness to a certain extent. Both were developed to measure the human mortality rate, given right-censoring events, that is, when some individuals belonging to an observed sample cease to be observable before the end of a given measurement period [101]. Hence, the observer cannot establish their state (dead or alive) at the end of the period. These results are among the foundations of the contemporary methods [93–97] used to process historical data on human mortality and related topics.

The Cutler–Ederer estimator has also been applied in finance due to its simplicity and flexibility [98, 100] to measure the default frequency observed for a population of risky debtors, and is still in use. The classical frequency estimator is commonly used to measure default frequency [102, 107, 108], given the complete availability of the information needed in most cases of financial interest.

Credit insurance modeling also requires the estimation of default frequencies because the underlying risk covered by a policy is the credit risk generated by a pool of debtors [39, 40, 85–87]. However, the insurer does not know if and when each debtor receives goods or services from the insured against a delayed payment. Moreover, the payment maturity is unknown unless the debtor is insolvent at the due date and the insured seller notifies the insurer of the claim. Hence, credit insurers cannot distinguish between the time when they can "observe" a default event because of an existing commercial credit that could not be paid at the due date or the time when a default of the same debtor would not affect them (i.e., it would be "unobservable") because the insured seller is not exposed. In the latter case, the seller could stop providing goods or services to the debtor, implying no further exposition of the insurer to the debtor's default risk. Therefore, multiple left- and right-censoring events occur during the coverage period: each debtor goes from being observable to being not observable and *vice versa*, depending on the unknown (i.e., stochastic) behavior of the insured seller.

This chapter addresses the problem of estimating the default frequency of a pool of risky debtors from a credit insurance company's perspective. As the main frequency estimators in the literature are inadequate for this purpose, a new parametric estimator is proposed.

Improved estimation of default probabilities brings relevant benefits to credit insurance companies. First, credit insurers have the right to reduce or cancel previously granted coverages. Nonetheless, taking the right management actions requires identifying which clusters of covered debtors are the most likely to generate losses in the future. The censoring events mentioned above may impair this ability, and the proposed estimation technique can overcome this issue. Further, the precision of the claim probability estimation affects policy pricing and provisioning dramatically. These are two core activities to any insurance company and are well-known to drive the profitability of an insurance business. Finally, a credit insurance company can benefit from precise default probabilities estimation also from a regulatory perspective. In the Solvency 2 framework, high precision default probabilities are needed to develop an internal model to measure Credit&Suretyship underwriting risk, as the European supervisory authority EIOPA compares the forward-looking estimates with the realized losses across several years for each authorized Credit&Suretyship internal model [81].

The remainder of this chapter is organized as follows. Section 6.1 introduces the problem that a credit insurer has to handle when estimating claim and default frequency. Section 6.2 presents a model describing the insured seller's behavior. This section also addresses the calibration of the proposed model and develops a new default frequency estimator as a function of the model parameters. Section 6.3 numerically tests the model against real-world historical default frequencies, showing the possibility of precisely estimating the "true" number of defaults, including those not observed by the insurer. Section 6.4 summarizes the main results and draws conclusions.

6.1 Information available to insurers in trade credit insurance

This section introduces the information asymmetry in credit insurance between the insurers and the insured, together with its consequences regarding claim frequency estimation.

§6.1.1 shows how credit insurers experience an important lack of information that does not allow the direct estimation of the exact number of insolvency/default events occurring for a basket of risky debtors. A consistent estimation of insolvency frequency is possible by applying a dedicated estimation method to compensate for the bias introduced by the poor quality of the available information. The following sections propose and apply this method.

§6.1.2 discusses the application of the principal frequency estimators available in the statistics literature, showing they are not suitable in the credit insurance context.

It is worth recalling that §1.2.1 outlines the main features of a credit insurance policy, introducing the main elements of credit insurance relevant to this research problem and the proposed solution. Moreover, §§3.2−3.4 introduce the three classic frequency estimators considered in §6.1.2.

6.1.1 Counting default events in a homogenous set of buyers

This section and the next §6.1.2 introduce the research problem: the insurer's underestimation of the default frequency for a homogeneous population of risky buyers due to the behavior of the insured sellers. Under the framework presented in §1.2.1, an insured seller can potentially detect early warning signs of a buyer's default because their ongoing business relationship provides reliable and updated information about the buyer's creditworthiness (*e.g.*, the payment behavior of the buyer regarding previous invoices). Considering buyers belonging to the small and medium-sized enterprise (SME) segment, for which publicly available information is scarce, the seller may evaluate the likelihood of a buyer's default better than the insurer. If the buyer's debt payable to the seller is zero when the buyer's credit standing worsens, the seller could decide against issuing new invoices. Thus, the buyer's default does not lead to a claim; hence, the insurer cannot observe it, despite a credit limit granted for the defaulted buyer. This is an immediate advantage for the insurer; however, it can also induce the insurer to underestimate the default probability of comparable buyers.

As previously noted, the analysis considers buyers in the SME segment, whose default probability must be estimated by measuring the default frequency for a cluster of comparable risks. Three preliminary specifications are worth considering before introducing the definitions and assumptions needed to address this problem.

- *Absorbing events.* Although a buyer that misses a payment is not necessarily insolvent, we only consider default events because the temporary liquidity crisis of the buyer is supposed to end with a full recovery of expired debt. If a claim has been notified to the insurer, the recovery notification will follow, enabling the distinction of that claim from those generated by actual default events. The buyer's financial distress may also cause a temporary interruption of its business relationship with the seller, thus preventing the insurer from observing a claim. However, a buyer's temporary distress does not affect the quality of the default frequency estimated by the insurer in both cases. Thus, from the insurer's perspective, only the missed payments associated with absorbing events are worth considering.

- *Recurrence of sellers.* In principle, the same insurer can serve multiple sellers who trade with the same buyer. This situation would increase the probability of observing at least one claim if the buyer defaults; otherwise, all insured sellers should interrupt their business relationships with the buyer before the default event. However, it is only reasonable to expect the relevant recurrence of insured sellers over the same buyer if the buyer is a *large* company that needs goods or services from many distinct (insured) providers, such that at least some are served by the same credit insurance company. That is, such a buyer is likely to belong to the corporate segment. As we consider only the SME segment, we can safely assume that each risky buyer is brought to the credit insurer by only one seller, apart from a negligible number of exceptions.

- *Need for an unbiased estimation.* As previously observed, avoiding a claim is an advantage for the insurer. In a perfectly stationary framework, estimating the fraction of latent default events is unnecessary, as they remain constant over time. However, given the common need to model the dynamics of default probabilities, it is necessary to separate them from the effects of the sellers' behavior over the default time series available to the insurer. This situation is further discussed in §§6.3.3-6.3.5, which provide some numerical applications.

Let us consider a credit insurance company (CIC) that protects an insured seller (ISE) against losses from a given pool of buyers $\{b\}$. Let $b_j \in \{b\}$ be the j-th buyer of the pool. We assume that the CIC has granted a credit limit to the ISE against losses from b_j. Let $[t_j, T_j]$ be the validity period of the credit limit, which implies that the ISE is indemnified by the CIC for the overdue invoices issued in $[t_j, T_j]$, provided the exposition over b_j does not exceed the approved credit limit. However, as the research problem refers to how to estimate the expected default frequency of buyers belonging to a homogenous subset $\{b'\} \subseteq \{b\}$, neither the amount of the credit limit nor the ISE exposure on b_j are relevant. The same applies to any other amount normally involved in a credit policy life-cycle (*e.g.*, the taxable amount of each insured invoice).

The ISE may issue an arbitrary number of invoices to b_j during $[t_j, T_j]$. Each invoice can be identified by double index jk, where index $k = 1 \dots K_j$ is used to sort the invoices issued to the j-th buyer in chronological order. In this context, the jk-th invoice is described by three subsequent instants: issue date t^0_{jk}, due date t^1_{jk}, and end of the extension period t^2_{jk}, where $t^0_{jk} < t^1_{jk} < t^2_{jk}$. If the jk-th invoice is overdue at t^1_{jk}, the ISE can wait until t^2_{jk} before notifying the CIC of the claim. The extra time can be used to fully recover the outstanding credit from b_j. After t^2_{jk}, even if the k-th invoice issued to the j-th buyer is still overdue, the ISE loses the right to notify the claim to the CIC.

An explicit notation is also introduced for the waiting time between two subsequent invoices:

$$\Delta^0_{jk} \coloneqq t^0_{jk} - t^0_{jk-1}, \tag{6.1}$$

and for the durations

$$\Delta^x_{jk} \coloneqq t^x_{jk} - t^{x-1}_{jk}, \quad x = 1, 2, \tag{6.2}$$

where Δ^1_{jk} and Δ^2_{jk} are the "credit term" and "extension period" of the k-th issued invoice, respectively. Extension period Δ^2 is defined in the policy form and is usually equal for all insured invoices; hence, $\Delta^2_{jk} = \Delta^2$. The credit term can be different for each invoice, as established by the ISE for the jk-th invoice at t^0_{jk}. However, the jk-th invoice is guaranteed by the CIC only if $0 < \Delta^1_{jk} \leq \Delta^1$ holds, where Δ^1 is the maximum credit term established in the policy form. It thus holds that $\Delta^x_{jk} > 0$ $(x = 0, 1, 2)$ by construction.

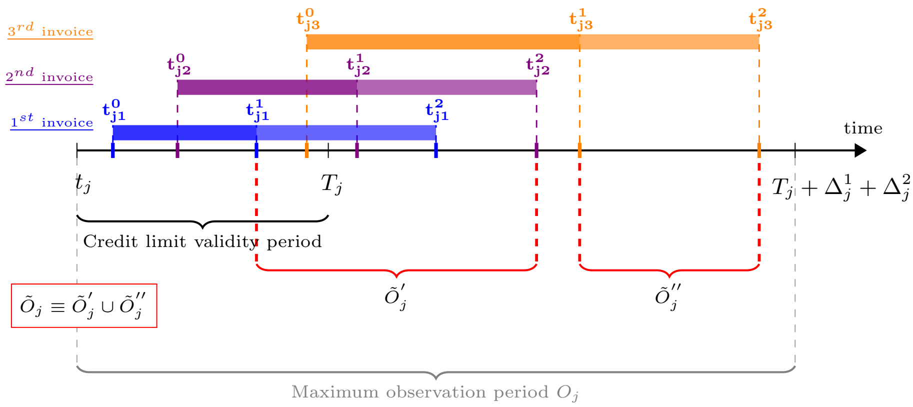

Figure 6.1. Example of credit insurance coverage. Three invoices are issued to the j-th buyer of a credit risk policy during the credit limit validity period. In this example, neither the period when the default of b_j can generate a claim or the period when the claim can be notified are connected intervals.

The example in Figure 6.1 displays a set of invoices issued to the same buyer. As previously discussed, given a credit limit validity period $[t_j, T_j]$, claims can occur over interval $[t_j, T_j + \Delta^1]$ and can be notified until $T_j + \Delta^1 + \Delta^2$.
By construction, a claim occurs only if:

$$\exists k \leq K_j \text{ s.t. } \tau_j \leq t^1_{jk}, \tag{6.3}$$

where τ_j is the default time of b_j. An invoice can somewhat be overdue only if the buyer becomes insolvent before the due date. When the jk-th invoice generates a claim and it is submitted, t^0_{jk} and t^1_{jk} are reported to the CIC. Otherwise, a paid invoice remains unsubmitted even after T_j, unless the CIC inspects the entire ISE turnover.
Let ISE detect a criticality when the default of b_j occurs at τ_j, implying the interruption of their commercial relationship. The following assumption then follows

Assumption 6.1 (Supply disruption in case of default)**.** *Let τ_j be the time when b_j transitions to an absorbing default state. Thus,*

$$\nexists k \leq K_j \textit{ s.t. } \tau_j \leq t^0_{jk}.$$

That is, in the case of a default, the insurer receives a claim only if at least one invoice has already been issued before τ_j. Assumption 6.1 implies that the default of b_j can be observed by the CIC only if:

$$\tau_j \in \mathcal{O}_j := \bigcup_{k=1}^{K_j} \left(t^0_{jk}, t^1_{jk}\right], \tag{6.4}$$

where $\mathcal{O}_j$ is a disjoint union of intervals, whose positions and lengths are stochastic for the CIC. The following is the estimation problem to address in this context: the number of default events $\hat{D}_{t,t'}$ occurring in a given period $(t,t']$ and observed by CIC is likely to be less than the actual number, $D_{t,t'}$, of all default events that occurred in the same period. In fact, it holds by construction that:

$$D_{t,t'} := \sum_{j:b_j\in\{b\}} \mathbb{1}_{\{\tau_j\in(t,t']\}} \geq \sum_{j:b_j\in\{b\}} \mathbb{1}_{\{\tau_j\in(t,t']\cap\mathcal{O}_j\}} =: \hat{D}_{t,t'}. \tag{6.5}$$

Furthermore, the CIC can receive a claim notification only in $\cup_{k=1}^{K_j}(t^1_{jk}, t^2_{jk}]$, implying a delay in the observation of the default event. However, given that t^0_{jk}, t^1_{jk} of the defaulted invoice are always notified to the CIC together with the claim occurrence, t^2_{jk} and Δ^2 are not relevant to what follows.

6.1.2 Classical frequency estimators and the credit insurance problem

Given a sample $\{b\}$ of N_t buyers not defaulted in t, each being observed in $(t, \min\{\tau_j, t'\}]$, where $j = 1 \dots N_t$ if each default event is directly observable, default frequency $f_{t,t'}$ over a given interval $(t,t']$ can be estimated as:

$$\hat{f}_{t,t'} := \frac{D_{t,t'}}{N_t}, \tag{6.6}$$

which is the maximum likelihood estimator of the Bernoulli distribution parameter, given a set of independent and identically distributed (IID) realizations. However, as discussed in §6.1.1, $D_{t,t'}$ is not directly available to the CIC. Hence, the form of the "classical" Bernoulli estimator is not applicable here because of the bias implied by Equation (6.5):

$$f_{t,t'} = \mathbf{E}\left[\hat{f}_{t,t'}\right] > \mathbf{E}\left[\hat{f}^{\mathrm{CIC}}_{t,t'}\right], \quad \hat{f}^{\mathrm{CIC}}_{t,t'} := \frac{\hat{D}_{t,t'}}{N_t}. \tag{6.7}$$

Assume N_s $(s \in (t,t'])$ may change because the subjects cease to be observable (right-censoring events or "losses") or become observable (left-censoring events) at some $\{s_1, s_2, \dots\} \subset (t,t']$. As we are introducing new causes to explain the N_s variations over time, other than default events, the value N_t at the beginning of the observation period does not consider these variations adequately. Kaplan and Meier [91] introduced a nonparametric estimator designed to address this problem:

$$\hat{f}^{\mathrm{KM}}_{t,t'} := 1 - \prod_{\tau^-\in\{\tau\}} \left(1 - \frac{N_{\tau^-} - D_\tau}{N_{\tau^-}}\right), \tag{6.8}$$

where $\{\tau\} \subset (t,t']$ is the set of default times observed in considered period $(t,t']$ and D_τ is the number of default events observed in τ. This estimator can be applied only if the observer knows the set of time values $\{\tau\}$ when at least one default occurs and the number of (left)right-censoring events occurring between two subsequent elements of $\{\tau\}$ to compute N_{τ^-}. Unfortunately, neither requirement is met for credit insurance. The CIC does not observe the "true" default time, τ_j, but only the nearest t^1_{jk} following it (the due date of the first overdue invoice). Moreover, left- and right-censoring events depend on the policyholder behavior regarding the invoices issued to b_j. As discussed in §1.2.1 and §6.1.1, this information is generally not available to the CIC.

When the exact timing of right-censoring events or default events occurring in period $(t, t']$ is available, the nonparametric estimator introduced by Cutler and Ederer [92] can be considered:

$$\hat{f}^{\mathrm{CE}}_{t,t'} := \frac{D_{t,t'}}{N_{t'} + D_{t,t'} + \frac{1}{2}R_{t,t'}} = \frac{D_{t,t'}}{N_t - \frac{1}{2}R_{t,t'}}, \tag{6.9}$$

where $R_{t,t'}$ is the number of right-censoring events occurring in $(t, t']$. Hence, each subject that exits from the sample before t' is approximately counted as observed until $t + \frac{1}{2}(t' - t)$. Once again, the requirements for applying this estimator do not correspond with the information available to the CIC. It is, thus, necessary to assume that only right-censoring events occur in $(t, t']$, and their number, $R_{t,t'}$, is *a posteriori* known in t'. As discussed, the observability of each b_j is affected by multiple stochastic left- and right-censoring events and their number remains unknown to the insurer, even after T_j.

Hence, the lack of relevant information in credit insurance does not allow for inferring default frequency $f_{t,t'}$ by applying one of the "classical" estimators developed for this purpose. However, although the $\hat{f}^{\mathrm{CIC}}_{t,t'}$ introduced in Equation (6.7) is a biased estimator of $f_{t,t'}$, it is also an unbiased estimator of the claims frequency that the CIC experiences under a given unknown invoice distribution (i.e., an unknown "average" ISE behavior). This situation explains when it is appropriate to use the Bernoulli estimator in credit insurance modeling, as is the case in the recent research by [102–104]. Assuming that the latent ISE behavior can be represented by a stationary distribution, there is no reason to identify and compensate bias $\mathbf{E}[\hat{f}^{\mathrm{CE}}_{t,t'}] - f_{t,t'}$, given that CICs are more interested in measuring their own claim frequency instead of the latent $f_{t,t'}$, which includes unobserved default events.

However, the ISE behavior experienced by the CIC may change over time due to the evolution of either the credit insurance demand (e.g., an increasing number of policyholders belonging to an economic sector that was formerly rare among ISEs) or the credit insurer selection criteria (e.g., a decreasing number of policyholders belonging to a cluster considered too risky during a certain time interval). In this case, it is possible to infer $f_{t,t'}$ and the bias due to the invoice distribution separately for a correct forecast of the claim probability.

Section 6.2 introduces a method to achieve the unbiased inference of $f_{t,t'}$ using the information available to the credit insurer. Section 6.3 discusses numerical examples of the stationary versus the evolving policyholder behavior and the possible estimation of the hidden $D_{t,t'}$ using the estimator proposed in this chapter.

6.2 Behavioral model for the policyholder and estimation of latent default events

As discussed in §6.1.2, classical frequency estimators are not suitable for measuring the default frequency from the information available to a credit insurer because of stochastic left- and right-censoring events.

The Kaplan–Meier and Cutler–Ederer estimators handle the presence of censoring events by partitioning the time interval and decreasing the effective number of observed subjects, respectively. Both strategies are conditioned by the availability of some information on timing or, at least, the number of censoring events occurring during the considered time interval. Given the lack of this kind of information experienced by the credit insurer, this section introduces a new estimator, which addresses the problem by increasing the number of observed default events. Namely,

$$\tilde{f}_{t,t'} := \frac{\tilde{D}_{t,t'}\left(\underline{\pi}\right)}{N_t} = \frac{\frac{1}{\tilde{P}(\underline{\pi})}\hat{D}_{t,t'}}{N_t}, \tag{6.10}$$

where $\tilde{P} := \text{Prob}(\tau_j \in \mathcal{O}_j | \mathcal{F}_{t_j})$ is the probability of observing a credit insurance claim conditioned to the buyer's default. Hence, $\tilde{D}_{t,t'}$ is an estimate of the actual (latent) number of default events $D_{t,t'}$, whereas $\hat{D}_{t,t'} \leq \tilde{D}_{t,t'}$ is the number of observed default events (i.e., number of claims). The policyholder's behavior can be described by the model introduced in §6.2.1, which is completely specified by parameter set $\underline{\pi}$.

The estimation of $\tilde{P}$ is addressed in §6.2.2. The theorem proven in §6.2.2.1 provides nonparametric expressions for the lower and upper bound of $\tilde{P}$. In §6.2.2.2, the corresponding parametric expressions are developed from the distributional assumptions introduced in §6.2.1. Finally, §6.2.3 describes a calibration method for estimating $\hat{\underline{\pi}}$.

6.2.1 Definitions and assumptions

Let us consider buyer b_j to be a risk underlying a credit risk policy. Assume the CIC grants a positive credit limit to the ISE against the default of b_j and let $[t_j, T_j]$ be the credit limit validity period. Following the notation introduced in §6.1.1, this section introduces a set of assumptions to describe the covered invoices and provides the distributions of the issue dates $\{t^0_{j1}, t^0_{j2}, \dots\}$ and due dates $\{t^1_{j1}, t^1_{j2}, \dots\}$.

When the k-th invoice is issued, k-th issue date t^0_{jk} and k-th credit term Δ^1_{jk} are simultaneously fixed by the ISE, despite being unknown to the CIC. Hence, a stochastic bivariate process is introduced.

$$\left(t^0_{jk}, \Delta^1_{jk}\right)_{k\in\mathbb{N}}, \tag{6.11}$$

using the conventional value $\left(t^0_{j0} \equiv t_j, \Delta^1_{j0} \equiv 0\right)$ for $k = 0$.

The filtration associated to the process is written as:

$$\mathcal{F}_{jk} := \sigma\left(t^0_{jk'}, \Delta^1_{jk'} \middle| k' \leq k\right); \quad \mathbb{F}_j := \left(\mathcal{F}_{jk}\right)_{k\in\mathbb{N}}. \tag{6.12}$$

The distributions needed to fully specify this process are introduced in Assumptions 6.4 and 6.5. However, it is useful to introduce two basic features of this process.

Assumption 6.2 (Markovianity). *The process defined in* (6.11) *is a stationary Markov chain, namely:*

$$\mathrm{Prob}\left(t^0_{jk},\Delta^1_{jk}\middle|\, t^0_{j0},\Delta^1_{j0},\dots,t^0_{jk-1},\Delta^1_{jk-1}\right)=\mathrm{Prob}\left(t^0_{jk},\Delta^1_{jk}\middle|\, t^0_{jk-1},\Delta^1_{jk-1}\right)$$
$$\mathrm{Prob}\left(t^0_{jk+1},\Delta^1_{jk+1}\middle|\, t^0_{jk},\Delta^1_{jk}\right)=\mathrm{Prob}\left(t^0_{jk},\Delta^1_{jk}\middle|\, t^0_{jk-1},\Delta^1_{jk-1}\right)$$

holds for each $k\in\mathbb{N}^*$.
Moreover, $(\Delta^1_{jk})_{k\in\mathbb{N}^*}$ *are assumed to be IID random variables, whereas only the distribution of* t^0_{jk} *depends on the realization of* $\left(t^0_{jk-1},\Delta^1_{jk-1}\right)$.

The issue date of a new invoice to the j-th buyer may depend on when the last invoice to the same buyer has expired (been paid). The buyer may decide to order a new supply of goods or services only after paying for the previous order. Alternatively, the seller may require receiving all outstanding credit from the debtor before satisfying new orders. Moreover, the timing of the new buyer's order may only be related to the buyer's needs, regardless of the debt situation. However, the buyer–seller decision on the timing of a new invoice issue is unlikely to depend on issues older than the last issue. Hence, we observe that (t^0_{jk},Δ^1_{jk}) depends on $(t^0_{jk-1},\Delta^1_{jk-1})$, which can at most imply Assumption 6.2.
Assumption 6.1 can be equivalently reformulated as:

$$\mathrm{Prob}_{jk}\left(t>t^0_{jk+1}\,|t>\tau_j\right)=0, \tag{6.13}$$

where notation $\mathrm{Prob}_{jk}(a|b)$ represents the probability associated with event a, conditioned to b, and filtered at time t^0_{jk}. However, we also introduce the following

Assumption 6.3 (ongoing business relationships). *Provided* b_j *continues to be solvent and granted by a credit limit, the ISE will almost certainly keep providing goods or services:*

$$\lim_{t\to\infty}\mathrm{Prob}_{jk}\left(t>t^0_{jk+1}\middle|\, t<\tau_j\wedge(t^0_{jk},t]\subseteq[t_j,T_j]\right)=1.$$

It is thus reasonable to assume that the existence of a credit limit on b_j implies an ongoing commercial relationship between the ISE and b_j.

6.2.1.1 Distributional assumptions

The following two assumptions are introduced to specify the probability density functions of t^0_{jk} and Δ^1_{jk}, respectively. Let

$$\rho^x_{jk|k'}(t)\,dt \;\coloneqq\; \mathrm{Prob}_{jk}\left(t^x_{jk}=t\right) \tag{6.14}$$
$$\rho^{\Delta x}_{jk|k'}(t)\,dt \;\coloneqq\; \mathrm{Prob}_{jk}\left(\Delta^1_{jk}=t\right), \tag{6.15}$$

where $x=0,1$. Assumptions 6.4 and 6.5 are then used to determine the joint distribution of issue dates t^0_{jk} $(k=1\dots K_j)$.

Credit terms Δ^1_{jk} are defined over the interval $\left(0, \Delta^1_j\right]$. However, in practice, Δ^1_{jk} is likely to belong to a discrete and finite subset of its domain. The due date is typically settled after an integer number of quarters, months, or, less commonly, weeks after the issue date. Hence, the effective domain of Δ^1_{jk} is $\{d^1, 2d^1, \ldots, N_\Delta d^1 \equiv \Delta^1\} \subset (0, \Delta^1]$, where d^1 is equivalent to a quarter, month, or week. The associated probability distribution is also discrete, as described by the following assumption.

Assumption 6.4 (Credit term distribution)**.** *Credit terms Δ^1_{jk} are IID categorical random variables, distributed according to the probability density function:*

$$\rho^{\Delta 1}_{jk|k'}(u) = \sum_{n=1}^{N_\Delta} w_n \delta\left(t - u_n\right) \quad k > k', \tag{6.16}$$

where $N_\Delta = \Delta^1/d^1$ is an integer, $u_n := nd^1$ and $\sum_{n=1}^{N_\Delta} w_n = 1$, where $w_n > 0$ $(n = 1 \ldots N_\Delta)$.

The temporal distribution of issue dates depends on the commercial relationship between the policyholder and buyer. A gamma distribution is a natural choice for modeling waiting times [110] and can be considered in this context. Assumption 6.5 introduces a three-parameter gamma distribution considering a location parameter to better represent the possible dependence of t^0_{jk} on t^1_{jk-1}.

Assumption 6.5 (Issue date distribution)**.** *The k-th issue date is distributed as a gamma random variable*

$$\rho^0_{jk|k-1}(v) = \begin{cases} \frac{1}{\Gamma(\mathcal{K}_k)\theta^{\mathcal{K}_k}} \left[v - \gamma_k\right]^{\mathcal{K}_k - 1} \exp\left(-\frac{v-\gamma_k}{\theta}\right) & v \geq \gamma_k, \\ 0 & \textit{otherwise.} \end{cases}, \tag{6.17}$$

where γ_k is a location parameter that depends on both the issue date and the credit term of the previous (i.e., $k-1$-th) invoice:

$$\gamma_k := t^0_{jk-1} + \xi_1 \Delta^1_{jk-1}, \tag{6.18}$$

and $\mathcal{K}_k$ is a shape parameter that depends on the credit term of the previous invoice.

$$\mathcal{K}_k := \frac{\xi_2 \Delta^1_{jk-1}}{\theta}. \tag{6.19}$$

However, scale parameter θ is assumed to be fixed at a constant value for all considered invoices. The same applies to ξ_1, ξ_2. The distribution is fully specified by the knowledge of (ξ_1, ξ_2, θ).

It is worth noting that Equation (6.18) in Assumption 6.5 introduces a dependence of the k-th location parameter γ_k on credit term Δ^1_{jk-1} of the previously issued (k-1)-th invoice. This introduction is required to represent business relationships where a service is provided on demand, without considering whether the previous service has already been paid for, as well as business relationships where there is a rigid alternation of services and payments. The flexibility provided by this assumption is further discussed in §6.2.1.2.

Assumption 6.5 states, *inter alia*, that the probability density function of variable t^0_{jk} can be represented as a function of difference $t^0_{jk} - t^0_{jk-1}$. Hence, it holds by definition that:

$$t^0_{jk} = t^0_{j0} + \sum_{\kappa=1}^{k} \Delta^0_{j\kappa}. \tag{6.20}$$

Equation (6.20) implies that, given the filtration $\mathcal{F}_{jk-1}$, Δ^0_{jk} is gamma distributed with the same scale and shape parameters of t^0_{jk} but a different location parameter:

$$\gamma'{}_k \coloneqq \xi_1 \Delta^1_{jk-1}. \tag{6.21}$$

Given that gamma variables with the same θ scale are closed under the "+" operation, Equation (6.20) allows writing densities $\rho^0_{jk|k'}(\cdot)$ and $\rho^{\Delta 0}_{jk|k'}(\cdot)$, $(k' < k)$. Given the same functional form for $\rho^0_{jk|k-1}(\cdot)$, the shape and location parameters are:

$$\mathcal{K}_{k|k'} = \frac{\xi_2}{\theta} \sum_{\kappa=k'}^{k-1} \Delta^1_{j\kappa} \tag{6.22}$$

$$\gamma_{k|k'} = t^0_{jk'} + \xi_1 \sum_{\kappa=k'}^{k-1} \Delta^1_{j\kappa}, \tag{6.23}$$

where we adopt the same convention, $t^0_{j0} \equiv t_j$, chosen for the process defined in Equation (6.11). Mathai and Moschopoulos [109] determined the multivariate gamma density function associated with the joint probability of random variables (RV), which can be expressed as the partial sums of gamma-distributed RV with the same scale parameter. Since their results fit perfectly in our context, they are stated without modifying anything but the semantics of the considered variables; thus, Assumption 6.5 can be considered without being rewritten.

Theorem 6.1 (Mathai–Moschopoulos multivariate gamma)**.** *Let us consider $k \in \mathbb{N}$ invoices, whose issue dates $\underline{t}^0_j \coloneqq \left(t^0_{j1}, \dots, t^0_{jk}\right)$ are conditionally distributed according to Assumption 6.5. Therefore, $\rho_{MM}(\cdot)$ is the joint probability density function conditional on $\mathcal{F}_{j0}$, to the realization of corresponding credit terms $\underline{\Delta}^1_j \coloneqq \left(\Delta^1_{j1}, \dots, \Delta^1_{jk}\right)$.*

$$\rho_{MM}\left(\underline{t}^0_j \middle| \underline{\theta}, \underline{\Delta}^1_j, \mathcal{F}_{j0}\right) \coloneqq \tag{6.24}$$
$$\frac{\left(t^0_{j1} - t^0_{j0}\right)^{\mathcal{K}_1 - 1}}{\theta^{\sum_{\kappa=1}^{k} \mathcal{K}_\kappa} \prod_{\kappa=1}^{k} \Gamma\left(\mathcal{K}_\kappa\right)} \left[\prod_{\kappa=2}^{k} \left(t^0_{j\kappa} - t^0_{j\kappa-1} - \gamma'_\kappa\right)^{\mathcal{K}_\kappa - 1}\right] e^{-\left(t^0_{jk} - \gamma_{k|0}\right)/\theta}$$

, where $\underline{\theta} \coloneqq (\theta, \mathcal{K}_1, \dots, \mathcal{K}_k, \gamma'_1, \dots, \gamma'_k)$ is the parameter array and

$$\gamma_{k|0} = t^0_{j0} + \sum_{\kappa=1}^{k} \gamma'_\kappa \tag{6.25}$$

by construction.

Moreover, covariance matrix Σ_j^0 *of* $\underline{t}_j^0$ *has the following form:*

$$\Sigma_j^0 = \begin{pmatrix} \sigma_1^2 & \sigma_1^2 & \dots & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2+\sigma_2^2 & \dots & \sigma_1^2+\sigma_2^2 \\ \vdots & \vdots & & \vdots \\ \sigma_1^2 & \sigma_1^2+\sigma_2^2 & \dots & \sigma_1^2+\dots+\sigma_k^2 \end{pmatrix}, \tag{6.26}$$

where $\sigma_\kappa^2 = \mathcal{K}_\kappa \theta^2$.

Assumptions 6.2, 6.4, and 6.5, together with Theorem 6.1, outline a behavioral model for the commercial relationship between the insured and buyers. The number of shape and location parameters $\mathcal{K}_k$ and the γ_k, associated with each k-th invoice, may be arbitrarily large, depending on the number of invoices issued over period $[t_j, T_j]$. However, the model is parsimonious because the only parameters to be calibrated are

$$\underline{\pi} \coloneqq (\xi_1, \xi_2, \theta, \underline{w}) \in \mathbb{R}_+^3 \times [0,1]^{N_\Delta}. \tag{6.27}$$

Triplet $\{\mathcal{K}_k, \gamma_k, \theta\}$ (needed to define the probability density associated with each t_{jk}^0) is determined as a function of $\{\xi_1, \xi_2, \theta\}$ and realized couple $t_{jk-1}^0, \Delta_{jk-1}^1$, as stated in Assumption 6.5.

6.2.1.2 Model flexibility

Despite the low dimensionality of $\underline{\pi}$, the model can represent a variety of commercial relationship types between the policyholder and a given risky buyer. Table 6.1 shows four examples of the possible $\underline{\pi}$ choices, each representing a common type of commercial relationship. The corresponding panel in Figure 6.2 shows the realization of process $(t_{jk}^0, \Delta_{jk}^1)$, generated by using the chosen parameterization.

A *regular supply* (case *a*) can be represented by $\xi_2 \ll 1$. A small ξ_2 value combined to a constant credit term induces the almost deterministic issuing of subsequent invoices (i.e., the one just after the expiry of the former). The further specification $\xi_1 = 1$ implies a continuous trade between the seller and the buyer:

$$E\left[t_{jk}^0\right] \gtrsim t_{jk-1}^1, \quad \sigma\left[t_{jk}^0\right] \ll 1. \tag{6.28}$$

Case	ξ_1	ξ_2	θ	$\underline{w}$
a)	1.000	0.001	0.500	$(0.00, 0.00, 0.00, 1.00)$
b)	5.000	0.001	0.500	$(0.00, 1.00, 0.00, 0.00)$
c)	0.000	1.714	0.800	$(0.00, 0.00, 0.50, 0.50)$
d)	0.000	1.000	0.050	$(0.25, 0.00, 0.00, 0.75)$

Table 6.1. Parameter choice for the examples shown in Figure 6.2, given $d^1 = \frac{1}{12}$.

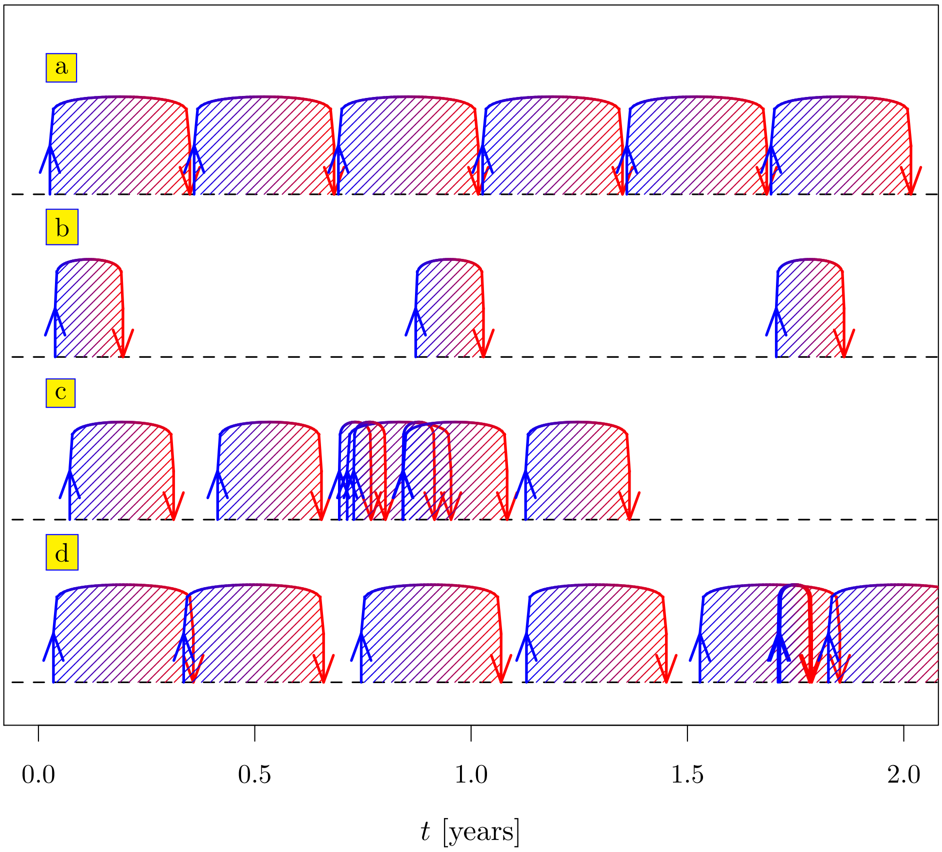

Figure 6.2. Examples of insured issuing behaviors described in §6.2.1.2: *a*) regular/continuous; *b*) regular/seasonal; *c*) irregular; and *d*) a possible intermediate case. Each arc represents interval $[t^0_{jk}, t^1_{jk}]$ associated with a given invoice. The upward arrows represent the issue dates t^0_{jk} and the downward arrows the corresponding due dates. The examples are obtained by using Assumptions 6.4 and 6.5 with different parameters choices - detailed in table 6.1.

A greater ξ_1 value leads to *seasonal trade* (case *b*), with the expected interval between two subsequent invoices being much longer than the credit term:

$$E\left[t^0_{jk} - t^0_{jk-1}\right] \simeq E\left[t^0_{jk+1} - t^0_{jk}\right] \gg \Delta^1_j, \quad \sigma\left[t^0_{jk}\right] \ll 1. \tag{6.29}$$

However, if the supply of goods or services is completely *aperiodic* (case *c*), $\xi_1 = 0$ allows the next invoice to be issued immediately after the previous one. Moreover, if the credit term is fixed to a constant value, Δ^1_j, for all considered invoices, choice $\xi_2 = \frac{\theta}{\Delta^1_j}$ provides a shape parameter $\mathcal{K} = 1$, imposing the mode of the distribution to be equal to t^0_{jk-1}. This behavior can be approximated even in the case of random credit terms by replacing the fixed value of Δ^1_{jk-1} with its average value. In this case, it holds that:

$$\text{mode}\left[t^0_{jk}\right] \gtrsim t^0_{jk-1}, \quad \sigma\left[t^0_{jk}\right] \simeq \theta^{\frac{1}{2}}. \tag{6.30}$$

Intermediate cases (e.g., case *d*) can be effectively represented by choosing $\xi_1 = 0$

and $\xi_2 = 1$, which implies that:

$$E\left[t^0_{jk}\right] = t^1_{jk-1}, \quad \sigma\left[t^0_{jk}\right] = \theta^{\frac{1}{2}}, \tag{6.31}$$

allowing for issue times both before and after the due date of the previous invoice, with the dispersion controlled by scale parameter θ.

6.2.2 Application to the inference problem in credit insurance

This section shows how the model introduced in §6.2.1 can be used to infer $\tilde{P}$ in Equation (6.10). In §6.2.2.1, the expressions for the upper and lower bound of $\tilde{P}$ are provided, without applying the distributional assumptions introduced in §6.2.1.1. Assumptions 6.4 and 6.5 are employed in §6.2.2.2 to quantify $\tilde{P}(\underline{\pi})$.

6.2.2.1 Nonparametric estimation of $\tilde{P}$

The results in this section and the next consider the expected values of Δ^0_{jk}, Δ^1_{jk}, and their functions. Index k can be used in expected values without loss of generality because all couples $\Delta^0_{jk}, \Delta^1_{jk}$ are assumed to be identically distributed, given Assumption 6.2.

In the following equation, for a compact notation, we write: $\mathbf{E}_k[\cdot] := \mathbf{E}[\cdot|\mathcal{F}_{jk}]$.

Theorem 6.2 (Nonparametric upper and lower bounds of $\tilde{P}$.)**.** *Given a j-th buyer underlying a credit insurance policy, and considering Assumptions 6.1-6.3 in limit $T_j - t_j \to \infty$, it holds that:*

$$\tilde{P}_{inf} \leq \tilde{P} \leq \tilde{P}_{sup},$$

where

$$\begin{aligned} \tilde{P}_{inf} &:= \frac{\mathbf{E}_{k-1}\left[\min\{\Delta^0_{jk+1}, \Delta^1_{jk}\}\right]}{\mathbf{E}_{k-1}\left[\Delta^0_{jk+1}\right]} \\ \tilde{P}_{sup} &:= \tilde{P}_{inf} + \frac{\mathbf{E}_{k-1}\left[\max\{0, \Delta^1_{jk} - \Delta^0_{jk+1} - \Delta^1_{jk+1}\}\right]}{\mathbf{E}_{k-1}\left[\Delta^0_{jk+1}\right]}. \end{aligned}$$

Proof. Assumption 6.1 implies that the default of the j-th buyer can generate a claim that becomes observable only if $\tau_j \in \mathcal{O}_j$, which is introduced in Equation (6.4). Hence, the probability of observing a claim conditioned to the presence of a default event is:

$$\tilde{P} = \mathbf{E}_0\left[\frac{\mu(\mathcal{O}_j)}{T_j - t_j}\right],$$

where $\mu(\cdot)$ is the standard Lebesgue measure of the considered interval.

Assumption 6.3 allows for partitioning the credit limit validity period as follows:

$$\lim_{T_j \to \infty} (t_j, T_j] \equiv \lim_{K_j \to \infty} \bigcup_{k=1}^{K_j} (t^0_{jk}, t^0_{jk+1}],$$

given that $t^0_{j0} = t_j$. Hence,

$$\tilde{P} = \lim_{K_j \to \infty} \mathbf{E}_0 \left[\frac{\mu\left(\bigcup_{k=1}^{K_j} (t^0_{jk}, t^1_{jk}]\right)}{\sum_{k=1}^{K_j} \mu\left((t^0_{jk}, t^0_{jk+1}]\right)} \right].$$

Assumption 6.2 implies that:

$$\sum_{k=1}^{K_j} \mu\left((t^0_{kj}, t^0_{jk+1}]\right) \xrightarrow{K_j \to \infty} K_j \mathbf{E}_{k-1}\left[\Delta^0_{jk+1}\right].$$

Hence,

$$\tilde{P} = \lim_{K_j \to \infty} \frac{\mathbf{E}_0\left[\mu\left(\bigcup_{k=1}^{K_j} (t^0_{jk}, t^1_{jk}]\right)\right]}{K_j \mathbf{E}_{k-1}\left[\Delta^0_{jk+1}\right]}.$$

Moreover,

$$\begin{aligned} \mu\left(\bigcup_{k=1}^{K_j} (t^0_{jk}, t^1_{jk}]\right) &= \sum_{k=1}^{K_j} \left[\mu\left((t^0_{jk}, t^1_{jk}]\right) - \mu\left(\bigcup_{k'>k} (t^0_{jk}, t^1_{jk}] \cap (t^0_{jk'}, t^1_{jk'}]\right)\right] \\ &\geq \sum_{k=1}^{K_j} \left[\mu\left((t^0_{jk}, t^1_{jk}]\right) - \mu\left(\bigcup_{k'>k} (t^0_{jk}, t^1_{jk}] \cap (t^0_{jk'}, \infty)\right)\right]. \end{aligned}$$

Since it holds that $t^0_{j1} \leq t^0_{j2} \leq \ldots$, it follows that:

$$\begin{aligned} \mu\left(\bigcup_{k=1}^{K_j} (t^0_{jk}, t^1_{jk}]\right) &\geq \sum_{k=1}^{K_j} \left[\mu\left((t^0_{jk}, t^1_{jk}]\right) - \mu\left((t^0_{jk}, t^1_{jk}] \cap (t^0_{jk+1}, \infty)\right)\right] \\ &= \sum_{k=1}^{K_j} \mu\left((t^0_{jk}, t^1_{jk}] \setminus (t^0_{jk+1}, \infty)\right) \\ &= \sum_{k=1}^{K_j} \min\{\Delta^0_{jk+1}, \Delta^1_{jk}\}. \end{aligned}$$

We can apply Assumption 6.2 once again, implying that:

$$\sum_{k=1}^{K_j} \min\{\Delta^0_{jk+1}, \Delta^1_{jk}\} \xrightarrow{K_j \to \infty} K_j \mathbf{E}_{k-1}\left[\min\{\Delta^0_{jk+1}, \Delta^1_{jk}\}\right].$$

Therefore, we have:

$$\tilde{P} \geq \frac{\mathbf{E}_{k-1}\left[\min\{\Delta^0_{jk+1}, \Delta^1_{jk}\}\right]}{\mathbf{E}_{k-1}\left[\Delta^0_{jk}\right]},$$

thereby completing the proof of the first inequality in the theorem.

The remaining inequality can be derived, considering that it holds that:

$$\begin{aligned}
\mu\left(\bigcup_{k=1}^{K_j}(t^0_{jk},t^1_{jk}]\right) &= \sum_{k=1}^{K_j}\left[\mu\left((t^0_{jk},t^1_{jk}]\right)-\mu\left(\bigcup_{k'>k}^{K_j}(t^0_{jk},t^1_{jk}]\cap(t^0_{jk'},t^1_{jk'}]\right)\right]\\
&\leq \sum_{k=1}^{K_j}\left[\mu\left((t^0_{jk},t^1_{jk}]\right)-\mu\left((t^0_{jk},t^1_{jk}]\cap(t^0_{jk+1},t^1_{jk+1}]\right)\right]\\
&= \sum_{k=1}^{K_j}\mu\left((t^0_{jk},t^1_{jk}]\setminus(t^0_{jk+1},t^1_{jk+1}]\right).
\end{aligned}$$

Although $t^0_{jk} \leq t^0_{jk+1}$ by construction, it is possible that $t^1_{jk} > t^1_{jk+1}$. Hence,

$$\sum_{k=1}^{K_j}\mu\left((t^0_{jk},t^1_{jk}]\setminus(t^0_{jk+1},t^1_{jk+1}]\right) = \sum_{k=1}^{K_j}\left[\min\{\Delta^0_{jk+1},\Delta^1_{jk}\} + (t^1_{jk}-t^1_{jk+1})^+\right],$$

thus completing the proof. ∎

6.2.2.2 Parametric estimation of $\tilde{P}$

The following result allows using the model introduced in Assumptions 6.1-6.5 to easily compute the lower and upper bounds of $\tilde{P}$, given the nonparametric expressions in Theorem 6.2. The expressions for $\tilde{P}_{\text{inf}}(\underline{\pi})$ and $\tilde{P}_{\text{sup}}(\underline{\pi})$, provided by the theorem below, only require the computation of gamma and incomplete gamma functions, whose numerical implementations are widely available. Hence, although gamma and incomplete gamma functions are usually classified as special functions, they can also be classified as somewhat “well-known” functions. Therefore, the expressions provided in the theorem are claimed to be “closed-form”.

Theorem 6.3 (Parametric upper and lower bounds of $\tilde{P}$)**.** *Given a j-th buyer underlying a credit insurance policy, and considering Assumptions 6.1-6.5 in limit $T_j - t_j \to \infty$, it holds that*

$$\tilde{P}_{inf}(\underline{\pi}) \leq \tilde{P} \leq \tilde{P}_{sup}(\underline{\pi})$$

$\tilde{P}_{inf}(\underline{\pi})$, and $\tilde{P}_{sup}(\underline{\pi})$ are two closed-form functions of model parameters $\underline{\pi}$.

$$\begin{aligned}
\tilde{P}_{inf}(\underline{\pi}) &= \frac{1}{\xi_1+\xi_2}\\
&+ \frac{\theta}{(\xi_1+\xi_2)\underline{w}\cdot\underline{u}}\sum_n\frac{w_n}{\Gamma(\mathcal{K}_{k+1})}\gamma\left(\frac{\max\{0,u_n(1-\xi_1)\}}{\theta},\mathcal{K}_{k+1}+1\right)\\
&- \frac{1-\xi_1}{(\xi_1+\xi_2)\underline{w}\cdot\underline{u}}\sum_n\frac{w_n u_n}{\Gamma(\mathcal{K}_{k+1})}\gamma\left(\frac{\max\{0,u_n(1-\xi_1)\}}{\theta},\mathcal{K}_{k+1}\right),\\
\tilde{P}_{sup}(\underline{\pi}) &= \tilde{P}_{inf}(\underline{\pi})\\
&+ \frac{1}{(\xi_1+\xi_2)\underline{w}\cdot\underline{u}}\sum_{n\check{n}}\frac{w_n w_{\check{n}}}{\Gamma(\mathcal{K}_{k+1})}[u_n(1-\xi_1)-u_{\check{n}}]\gamma\left(\frac{\max\{0,u_n(1-\xi_1)-u_{\check{n}}\}}{\theta},\mathcal{K}_{k+1}\right)\\
&- \frac{\theta}{(\xi_1+\xi_2)\underline{w}\cdot\underline{u}}\sum_{n\check{n}}\frac{w_n w_{\check{n}}}{\Gamma(\mathcal{K}_{k+1})}\gamma\left(\frac{\max\{0,u_n(1-\xi_1)-u_{\check{n}}\}}{\theta},\mathcal{K}_{k+1}+1\right),
\end{aligned}$$

where $\gamma(\cdot,\mathcal{K})$ is the lower incomplete gamma function with shape parameters $\mathcal{K}$, and $\Gamma(\cdot)$ is the ordinary gamma function.

Proof. The proof comprises a direct application of Assumptions 6.4 and 6.5 to the statement of Theorem 6.2. It holds that:

$$\mathbf{E}_{k-1}\left[\Delta^0_{jk+1}\right] = \sum_n w_n \int_{t^0_{jk}+\xi_1 u_n}^{+\infty} dv \rho_\Gamma \left(v - t^0_{jk} - \xi_1 u_n | \mathcal{K}_{k+1}, \theta\right)\left(v - t^0_{jk}\right),$$

where $\rho_\Gamma(\cdot|\dots)$ is the probability density function (PDF) of a gamma-distributed RV, given a couple of shape and scale parameters. Substitution $y_n \coloneqq v - t^0_{jk} - \xi_1 u_n$ implies that:

$$\begin{aligned}\mathbf{E}_{k-1}\left[\Delta^0_{jk+1}\right] &= \sum_n w_n \int_{\mathbb{R}_+} dy_n \rho_\Gamma\left(y_n|\mathcal{K}_{k+1},\theta\right)\left(y_n + \xi_1 u_n\right) \\ &= \left(\xi_1 + \xi_2\right) \underline{w} \cdot \underline{u}.\end{aligned}$$

Furthermore, given the notation above, we have:

$$\begin{aligned}\mathbf{E}_{k-1}\left[\min\{\Delta^0_{jk+1}, \Delta^1_{jk}\}\right] &= \sum_n w_n \int_{\mathbb{R}_+} dy_n \rho_\Gamma\left(y_n|\mathcal{K}_{k+1},\theta\right) \min\{u_n, y_n + \xi_1 u_n\} \\ &= \sum_n w_n \Bigg[\underbrace{\int_0^{\max\{0,u'_n\}} \dots (y_n + \xi_1 u_n)}_{y_n \le u'_n} + \underbrace{\int_{\max\{0,u'_n\}}^{+\infty} \dots u_n}_{y_n > u'_n}\Bigg] \\ &= \sum_n w_n \left[u_n + \int_0^{\max\{0,u'_n\}} \dots [y_n - u_n(1-\xi_1)]\right],\end{aligned}$$

where $u'_n \coloneqq u_n(1-\xi_1)$. Moreover, we omit $dy_n \rho_\Gamma\left(y_n|\mathcal{K}_{k+1},\theta\right) \equiv \dots$ for a compact notation. Hence,

$$\begin{aligned}&\mathbf{E}_{k-1}\left[\min\{\Delta^0_{jk+1}, \Delta^1_{jk}\}\right] = \\ &\underline{w} \cdot \underline{u} + \frac{1}{\Gamma(\mathcal{K}_{k+1})} \sum_n w_n \left[\theta\gamma\left(\frac{\max\{0,u'_n\}}{\theta}, \mathcal{K}_{k+1}+1\right) - u'_n \gamma\left(\frac{\max\{0,u'_n\}}{\theta}, \mathcal{K}_{k+1}\right)\right],\end{aligned}$$

thus calling for proving the expression of $\tilde{P}_{\text{inf}}$.

Therefore, to prove the expression of $\tilde{P}_{\text{sup}}$, we must apply Assumptions 6.4 and 6.5 to the additional term to be considered, as per Theorem 6.2:

$$\begin{aligned}&\mathbf{E}_{k-1}\left[\max\{0, \Delta^1_{jk} - \Delta^1_{jk+1} - \Delta^0_{jk+1}\}\right] \\ &= \sum_{n\check{n}} w_n w_{\check{n}} \int_{\mathbb{R}_+} \dots \max\{0, u_n(1-\xi_1) - u_{\check{n}} - y_n\} \\ &= \sum_{n\check{n}} w_n w_{\check{n}} \int_0^{\max\{0,u''_{n\check{n}}\}} \dots (u''_{n\check{n}} - y_n),\end{aligned}$$

where $\check{n}$ is the auxiliary index used to express the PDF of Δ^1_{jk+1} and $u''_{n\check{n}} \coloneqq u_n(1-\xi_1) - u_{\check{n}}$. Hence,

$$\begin{aligned}&\mathbf{E}_{k-1}\left[\max\{0, \Delta^1_{jk} - \Delta^1_{jk+1} - \Delta^0_{jk+1}\}\right] \\ &= \frac{1}{\Gamma(\mathcal{K}_{k+1})} \sum_{n\check{n}} w_n w_{\check{n}} \left[u''_{n\check{n}} \gamma\left(\frac{\max\{0,u''_{n\check{n}}\}}{\theta}, \mathcal{K}_{k+1}\right) - \theta\gamma\left(\frac{\max\{0,u''_{n\check{n}}\}}{\theta}, \mathcal{K}_{k+1}+1\right)\right],\end{aligned}$$

completing the proof. ∎

Theorem 6.3 provides closed-form expressions for the upper and lower bounds of $\tilde{P}$. However, we require an effective value of $\tilde{P}(\underline{\pi}) \in [\tilde{P}_{\inf}(\underline{\pi}), \tilde{P}_{\sup}(\underline{\pi})]$ to estimate $\tilde{D}_{t,t'}(\underline{\pi})$ as per Equation (6.10).

Although the correct choice depends on the $\underline{\pi}$ value, the following generally holds for a realistic setup, such as the examples in Table 6.1 and Figure 6.2.

$$\tilde{P}(\underline{\pi}) \simeq \tilde{P}_{\inf}(\underline{\pi}) \tag{6.32}$$

Theorem 6.3 implies that positive contributions to $\delta\tilde{P} := \tilde{P}_{\sup}(\underline{\pi}) - \tilde{P}_{\inf}(\underline{\pi})$ are possible only if $\Delta^1_{jk+1} < \Delta^1_{jk}$ and $\Delta^0_{jk} < \Delta^1_{jk}$. The joint probability of these two conditions can be reasonably expected to be "small", implying that $\delta\tilde{P} \simeq 0$.

Remark 6.1 (Weakening Assumption 6.1)**.** *Assumption 6.1 can be easily weakened, thereby preserving the results of Theorems 6.2 and 6.3. It is possible to introduce a probability $P_{obs} < 1$ associated with the capability of the ISE to detect the default of b_j and interrupt their business relationship instantaneously. Therefore, the probability that the CIC receives a claim in case of default becomes $\tilde{P}' = \tilde{P}P_{obs} + 1 - P_{obs}$, implying that $\lim_{P_{obs}\to 0} \hat{D}_{t,t'} = D_{t,t'}$, ceteris paribus.*

6.2.3 Model calibration

Let us consider a set of debtors, $j = 1 \dots J$. For each j-th debtor, a set of past invoices, $\kappa_j = 1 \dots K_j$, issued from the same ISE is available. Data from multiple sellers may also be used simultaneously in the calibration, provided they are homogenous regarding the commercial relationships with their buyers. However, this case is supposedly rare and, hence, negligible as stated in §1.2.1. For each k_j-th invoice, set $\{t^0_{jk_j}, \Delta^1_{jk_j}\}$ is known. Information about Δ^2_{jk} is also available but is irrelevant to our scope.

For each j-th debtor, we consider only invoices $k_j \geq 2$ because we must use the value of $\Delta^1_{jk_j-1}$ in the calibration process, which is not available for $k_j = 1$.

In the following, we omit the explicit dependence of the invoice index on the debtor index for a compact notation (i.e., $k_j \equiv k$).

Δ^0_{jk} $(j = 1 \dots J,\ k_j = 2 \dots K_j)$ are distributed as:

$$\Delta^0_{jk} \sim \text{Gamma}\left(\mathcal{K}_k, \theta, \gamma'_k\right), \tag{6.33}$$

where the shape, scale, and location parameters are defined above, given Assumption 6.5 and Equation (6.21). Theorem 6.1 implies that the likelihood associated with the joint observation of a historical data set $\{\Delta^0_{jk}, \Delta^1_{jk-1}\}$ is:

$$\mathcal{L} = \prod_{jk} \frac{(\Delta^0_{jk} - \gamma'_k)^{\mathcal{K}_k - 1}}{\theta^{\mathcal{K}_k}\Gamma(\mathcal{K}_k)} \exp\left(-\frac{\Delta^0_{jk} - \gamma'_k}{\theta}\right). \tag{6.34}$$

For a compact notation, let us consider operator

$$\langle\cdot\rangle := \frac{\sum_{j=1}^{J}\sum_{k=2}^{K_j} \cdot}{\sum_{j=1}^{J}(K_j - 1)}. \tag{6.35}$$

Without considering normalization coefficient $\left(\sum_{j=1}^{J} K_j - 1\right)^{-1}$, the log-likelihood can be written as:

$$\ln \mathcal{L} = \left\langle (\mathcal{K}_\kappa - 1) \ln \left(\Delta^0_{j\kappa} - \gamma'_\kappa \right) - \mathcal{K}_\kappa \ln \theta - \frac{\Delta^0_{j\kappa} - \gamma'_\kappa}{\theta} - \ln \Gamma \left(\mathcal{K}_\kappa \right) \right\rangle . \tag{6.36}$$

Recalling that $\mathcal{K}_k$ is a function of θ, let us simplify it by considering alternate parameterization $\xi'_2 := \xi_2/\theta$ as follows: condition

$$\partial_\theta \ln \mathcal{L}|_{(\xi_1, \xi'_2, \theta) \equiv (\hat{\xi}_1, \hat{\xi}'_2, \hat{\theta})} = \frac{\langle \Delta^0_{jk} \rangle - \hat{\xi}_1 \langle \Delta^1_{jk-1} \rangle}{\hat{\theta}^2} - \frac{\hat{\xi'}_2 \langle \Delta^1_{jk-1} \rangle}{\hat{\theta}} = 0 \tag{6.37}$$

implies

$$\hat{\theta} = \frac{\langle \Delta^0_{jk} \rangle - \hat{\xi}_1 \langle \Delta^1_{jk-1} \rangle}{\hat{\xi'}_2 \langle \Delta^1_{jk-1} \rangle} . \tag{6.38}$$

Nullifying the other two components, $(\partial_{\xi_1}, \partial_{\xi'_2})$, of the gradient and substituting Equation (6.38), we have:

$$0 = \frac{\langle \Delta^1_{jk-1} \rangle}{\hat{\theta}} - \left\langle \frac{\hat{\xi}'_2 \Delta^1_{jk-1} - 1}{\Delta^0_{jk} - \hat{\xi}_1 \Delta^1_{jk-1}} \Delta^1_{jk-1} \right\rangle , \tag{6.39}$$

which can be rewritten as:

$$\hat{\xi}'_2 = \frac{\left\langle \frac{\Delta^1_{jk-1}}{\Delta^0_{jk} - \hat{\xi}_1 \Delta^1_{jk-1}} \right\rangle}{\left\langle \frac{(\Delta^1_{jk-1})^2}{\Delta^0_{jk} - \hat{\xi}_1 \Delta^1_{jk-1}} \right\rangle - \frac{\langle \Delta^1_{jk-1} \rangle^2}{\langle \Delta^0_{jk} \rangle - \hat{\xi}_1 \langle \Delta^1_{jk-1} \rangle}} , \tag{6.40}$$

allowing for an explicit estimation of ξ'_2 as a function of ξ_1 and

$$0 = \left\langle \Delta^1_{jk-1} \ln \left(\frac{\Delta^0_{jk} - \hat{\xi}_1 \Delta^1_{jk-1}}{\hat{\theta}} \right) - \Delta^1_{jk-1} \psi(\hat{\xi}'_2 \Delta^1_{jk-1}) \right\rangle , \tag{6.41}$$

where $\psi(\cdot)$ denotes the digamma function. Equation (6.41), together with Equation (6.38), imply that:

$$\left\langle \Delta^1_{jk-1} \left[\psi \left(\hat{\xi}'_2 \Delta^1_{jk-1} \right) - \ln \left(\hat{\xi}'_2 \langle \Delta^1_{jk-1} \rangle \right) \right] \right\rangle = \left\langle \Delta^1_{jk-1} \ln \left(\frac{\Delta^0_{jk} - \hat{\xi}_1 \Delta^1_{jk-1}}{\langle \Delta^0_{jk} \rangle - \hat{\xi}_1 \langle \Delta^1_{jk-1} \rangle} \right) \right\rangle . \tag{6.42}$$

Equation (6.42) is a generalization of the framework of the classical maximum likelihood estimator of the shape parameter for a gamma distribution without a location parameter. In the standard case, with $\xi_1 = 0$, the condition to be solved is:

$$\psi(\hat{\xi}'_2) - \ln \hat{\xi}'_2 = \left\langle \ln \Delta^0_{jk} \right\rangle - \ln \left\langle \Delta^0_{jk} \right\rangle , \tag{6.43}$$

which is known to be numerically well-behaved.

Ideally, the inference problem of estimating ξ_1, ξ'_2, and θ can be solved by using the nonlinear system of Equations (6.40) and (6.42). As shown above, these equations are obtained by imposing $\overline{\nabla} \ln \mathcal{L} = 0$ to maximize the log-likelihood.

Hence, it is useful to check the numerical behavior of the gradient components evaluated at the "true" parameters $(\xi_1^\star, \xi_2^{\prime\star}, \theta^\star)$ to which estimators $(\hat{\xi}_1, \hat{\xi}_2, \hat{\theta})$ must converge. This can be achieved by simulating a significant number of scenarios for an invoice dataset $\{\Delta_{jk}^0, \Delta_{jk}^1\}$ generated as per chosen parameter set $(\xi_1^\star, \xi_2^{\prime\star}, \theta^\star)$. The same parameter values can be substituted in Equations (6.37), (6.39), and (6.41), together with the dataset generated in each scenario. The ideal distribution of each gradient component should have a small variance and a mean value as close as possible to zero, where $(\xi_1^\star, \xi_2^{\prime\star}, \theta^\star)$ is the local maximum of the likelihood function by construction.

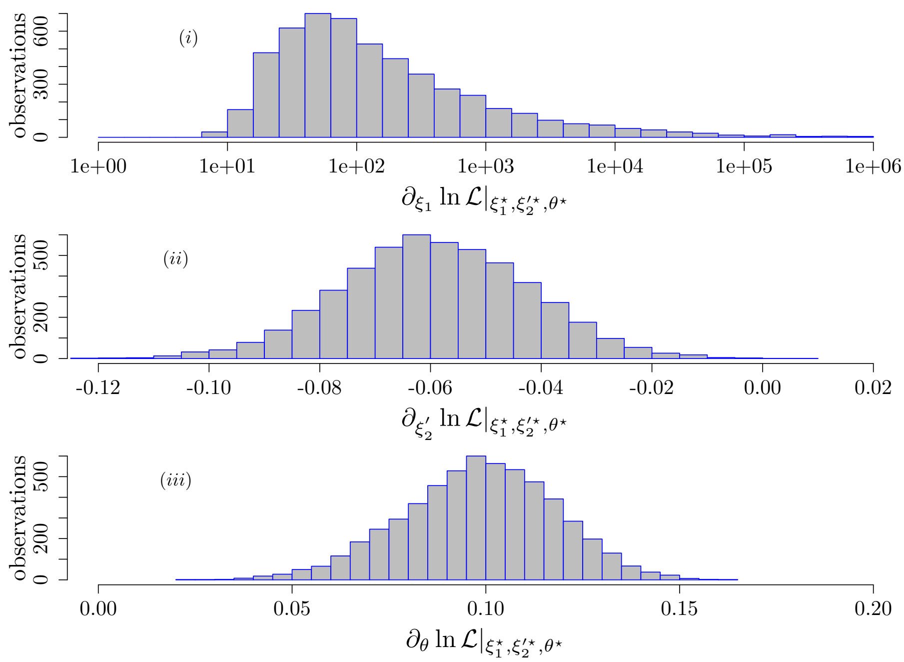

Figure 6.3. Numerical density distribution of the $\overline{\nabla} \ln \mathcal{L}$ components. A set of 5.000 Monte Carlo scenarios are considered (100 buyers $\times$ 5 years periods for each scenario). Scenarios are simulated using the parameterization "c" in Table 6.1.

An example of the numerical distributions obtained is shown in Figure 6.3, considering the parameterization "c" in Table 6.1.

Generally, $\partial_{\xi_1} \ln \mathcal{L}$ (panel i) is highly unstable, which is related to stochastic terms $(\Delta_{jk}^0 - \xi_1 \Delta_{jk-1}^1)^{-1}$. In fact, among considered cases a–d, this issue is less important only in case d because $\theta^\star \ll 1$ increases the relative weight of the other contributions to $\partial_{\xi_1} \ln \mathcal{L}|_{\theta^\star}$. However, the $\partial_{\xi_2'}$ and ∂_θ components (panels ii and iii, respectively) are stable across the considered scenarios, despite not being completely free of bias. This situation also holds true for Equation (6.42). Therefore, it can be used for the $\hat{\xi}_2'$ quantification with an acceptable error.

Equation (6.41) implies that $\Delta_{jk}^0 - \xi_1 \Delta_{jk-1}^1 > 0$ for each observed pair $(\Delta_{jk}^0, \Delta_{jk}^1)$. Furthermore, because $\Delta_{jk}^0 \geq 0$, we have $\xi_1 > 0$. These two conditions provide a range for ξ_1, defining a unidimensional optimization problem to be solved numerically to compute $(\hat{\xi}_1, \hat{\xi}'_2, \hat{\theta})$:

$$\hat{\xi}_1 = \underset{\xi_1}{\text{argmax}} \mathcal{L}\left(\xi_1, \hat{\xi}'_2, \hat{\theta}\right) \tag{6.44}$$

$$s.t. \begin{cases} 0 \leq \xi_1 \leq \min_{jk} \left\{ \frac{\Delta^0_{jk}}{\Delta^1_{jk-1}} \right\} \\ \hat{\xi}'_2 = \Xi\left(\xi_1\right) \\ \hat{\theta} = \Theta\left(\xi_1, \hat{\xi}'_2\right) \end{cases},$$

where functions $\Xi(\xi_1)$ and $\Theta(\xi_1, \xi'_2)$ are defined in Equations (6.42) and (6.38), respectively.

Alternatively, $\Xi(\cdot)$ and $\Theta(\cdot)$ can be defined through a moment-matching approach. Assumptions 6.4 and 6.5 provide a closed-form expression for each of the first two centered moments $\mathbf{E}_{k-1}\left[\Delta^0_{jk}\right]$ and $\sigma^2_{k-1}\left[\Delta^0_{jk}\right]$, implying that

$$\hat{\xi}_2 = \frac{\langle \Delta^0_{jk} \rangle}{\langle \Delta^1_{jk-1} \rangle} - \hat{\xi}_1, \tag{6.45}$$

$$\hat{\theta} = \frac{\langle (\Delta^0_{jk})^2 \rangle - \langle \Delta^0_{jk} \rangle^2}{\langle \Delta^1_{jk-1} \rangle} \frac{1}{\hat{\xi}_2}. \tag{6.46}$$

Equation (6.46) is an alternate choice for $\Theta(\cdot)$ with respect to equation (6.38). Further, a different $\Xi(\cdot)$ form can be derived by considering equations (6.45) and (6.46), instead of equation (6.42). As discussed in §6.3.2, the two alternate specifications for the couple $\Theta(\cdot), \Xi(\cdot)$ return estimates whose errors are comparable to each other.

Problem (6.44) can be approached numerically using a gradient-free method to explore the $\left[0, \min_{jk} \left\{\Delta^0_{jk}/\Delta^1_{jk-1}\right\}\right]$ interval.

Another possible way to estimate (ξ_1, ξ_2, θ) is a "pure" moment-matching approach instead of maximizing the likelihood function $\mathcal{L}(\xi_1, \xi'_2, \theta)$. In this case, $(\hat{\xi}_1, \hat{\xi}_2, \hat{\theta})$ can be fully specified by introducing the estimator

$$\hat{\xi}_1 := \min_{jk} \left\{ \frac{\Delta^0_{jk}}{\Delta^1_{jk-1}} \right\} \tag{6.47}$$

and applying equations (6.45) and (6.46) directly.

Assumption 6.2 allows for calibrating $\underline{w} = (w_1, \ldots, w_{N_\Delta})$ separately, which is required in Assumption 6.4. Considering the normalization condition imposed on $\underline{w}$, we can write the log-likelihood as:

$$\ln \mathcal{L}\left(\underline{w}\right) = m_1 \ln \left(1 - \sum_{n=2}^{N_\Delta} w_n\right) + \sum_{n=2}^{N_\Delta} m_n \ln w_n, \tag{6.48}$$

where m_n is the number of observed invoices whose credit term is equal to nd^1. It holds that $\sum_n m_n = \sum_j K_j$. The resulting maximum likelihood estimator is:

$$\underline{\hat{w}} = \left(\frac{m_1}{\sum_j K_j} \cdots \frac{m_{N_\Delta}}{\sum_j K_j} \right). \tag{6.49}$$

Equations (6.44) and (6.49) allow calibrating the model defined by Assumptions 6.1-6.5. Calibration returns $\underline{\hat{\pi}}$ enable the estimation of $\tilde{D}_{t,t'}$ as per Theorem 6.3 and the subsequent approximation provided in Equation (6.32).

6.3 Numerical application

This section applies the model presented in Section 6.2 to a realistic context. As a result, it demonstrates that the calibrated model can infer a latent time series D_t of default events from series $\hat{D}_t$ observed by the CIC. Therefore, we employ the notation:

$$D_t := D_{t,t+\delta_t}, \tag{6.50}$$

given that $t \in \{t\} := \{t_0, t_1 = t_0 + \delta_t, \ldots, t_i = t_0 + i\delta_t, \ldots\}$ and a dynamic set $\{b\}_t$ of the risky buyer, such that:

$$\{b\}_t := \{b_j : [t_j, T_j] \supseteq (t, t + \delta_t]\}. \tag{6.51}$$

Latent time series D_t is "realistic"; thus, it is Monte Carlo simulated using publicly available Italian quarterly bad loans rates [108, 112], measured by the Bank of Italy. A corresponding time series $\hat{D}_{t'}$ available to the CIC is generated, assuming that each policyholder can be classified into one of the four behavior profiles (a–d) introduced in §6.2.1.2.

Each $t \in \{t\}$ has a corresponding element in $t' \in \{t'\}$, such that $t = t' - \frac{1}{2}\underline{w} \cdot \underline{u}$. The presence of a positive credit term implies the existence of a stochastic lag between the actual default event, which contributes to D_t, and the observed due date of the expired invoice, which contributes to $\hat{D}_{t'}$.

The model is, thus, applied, leading to:

$$\tilde{D}_t = \frac{\hat{D}_{t'}}{\langle \tilde{P} \rangle_{t'}} \simeq \frac{\hat{D}_{t'}}{\frac{1}{|\{b\}_{t'}|} \sum_{j \in \{b\}_{t'}} \tilde{P}_{\inf}(\underline{\pi}_j)}, \tag{6.52}$$

where $\underline{\pi}_j$ is the parameter set associated with the behavior-type (a–d) that copes with the business relationship between each b_j and the corresponding ISE. This model verifies that the *a posteriori* estimate $\tilde{D}_t$ reproduces the latent *a priori* true series D_t, even when the variation of $\{b\}_{t'}$ over time produces an observable $\hat{D}_{t'}$ that may be uncorrelated or anticorrelated with D_t.

§6.3.1 describes the numerical setup adopted in the simulations. §6.3.2 reports a comparison among the calibration tecniques introduced in §6.2.3, based on the setup presented in §6.3.1. In §6.3.3, $\tilde{D}_t$, $\hat{D}_{t'}$, and $\tilde{D}_t$ are compared to simulate a "static" $\{b\}_{t'}$, where $\underline{\pi}_j = \underline{\pi}$ for each j. In §6.3.4, the same comparison is performed in the presence of a dynamic $\{b\}_{t'}$, allowing for the variation of $\langle \tilde{P} \rangle_{t'}$ over time. Finally, §6.3.5 provides an example where the investigated technique comes in handy in a real-life context.

6.3.1 Simulation of claims from historical time series

Latent time series D_t is generated using Italian quarterly bad loan rates, r_t, calculated based on the number of borrowers [112]. The considered data cube is publicly available from the Bank of Italy. The data were filtered to consider nonfinancial enterprises over the period between January 2007 and December 2017. r_t is evaluated by applying Equation (6.6) on a quarterly basis. Default events are generated by the corresponding deseasonalized daily hazard rates:

$$h_{t'} = \mathcal{S}_3 \left(\{t\}, \left\{ -\sum_{s=t}^{t+3} \ln(1 - r_s) \right\}, t' \right) \tag{6.53}$$

where $\{t\}$ is the quarterly schedule over the considered period, and $\mathcal{S}_3$ is a cubic spline interpolator employed to make the instantaneous hazard rate available with a daily sampling frequency. Therefore, default times τ_j are generated from cumulative distribution function $P(t \geq \tau | \mathcal{F}_{t'}) = 1 - \exp(-\frac{1}{365} \sum_{s=t'}^{t} h_s)$, and time series D_t is produced for each Monte Carlo scenario.

Further, to simulate $\hat{D}_t$, we generate a set of invoices for each buyer $\times$ scenario, as per the distribution defined in Assumptions 6.4 and 6.5 and one of the parameters sets discussed in §6.2.1.2. A stochastic uniform shift is applied to each scenario $(t^0_{jk}, \Delta^1_{jk})_{jkt}$, considering that the CIC does not know the timing of the first invoice issued after t_j. The simulated invoices, together with Assumption 6.1, allow for the selection of default events $\tau_j \in \{\tau : \exists k \text{ s.t. } \tau \in [t^0_{jk}, t^1_{jk}]\}$ to be observable to the CIC as claims. Transformation $\tau_j \mapsto \min\{t^1_{jk} : t^1_{jk} \geq \tau_j\}$ is applied to the selected events to compute the claim timing. Hence, $\hat{D}_t$ is generated.

The results discussed in §6.3.3 and §6.3.4 are obtained by considering 1000 scenarios and 5000 buyers. The latter is "small" relative to the typical size of a realistic credit insurance portfolio, suggesting that the proposed estimation method remains stable even after partitioning a given portfolio into homogenous risk clusters.

However, the next subsection 6.3.2 highlights the robustness of the calibration process under far more severe conditions (*i.e.*, 50 buyers and "wrong" distributional assumptions), leading to a fairly low estimation error even in these cases.

6.3.2 Calibration of model's parameters from historical time series

§6.2.3 introduces three possible approaches to the problem of estimating the parameters set (ξ_1, ξ_2, θ), briefly summarized in the following. The three approaches are labelled as "MLML", "MLMM", and "MMMM" for the sake of brevity.

MLML: a "pure" maximum likelihood approach, defined by Problem (6.44), and equations (6.42) and (6.38). Namely, auxiliary functions $\Xi(\xi_1)$ and $\Theta(\xi_1, \xi_2')$ are defined by nullifying the likelihood's gradient.

MLMM: a "mixed" maximum likelihood and moment-matching approach, where Problem (6.44) is specified through equations (6.45) and (6.46). Indeed, while parameters are still found by maximizing the likelihood value, auxiliary functions $\Xi(\xi_1, \theta)$ and $\Theta(\xi_2)$ are chosen by following a moment-matching criterion.

MMMM: a "pure" moment-matching approach, where equations (6.45) and (6.46) are completed by equation (6.47), instead of being applied to Problem (6.44). No numerical optimization is required in this case, implying (almost) zero computational cost.

The three techniques are used to process a simulated historical data set, generated by applying real data and methods described in §6.3.1. As anticipated in §6.3.1, calibration has been performed in two stressed setups to verify and compare the robustness of these three methods. Further details on the calibration outcomes are available in appendix B.3.

6.3.2.1 Calibration in presence of small populations

Table 6.2 shows the results obtained by processing only 50 business relationships in calibration. Each technique is applied to the four behavioral types *a*-*d* introduced in §6.2.1.2. MLML, MLMM, and MMMM replicate the true values of (ξ_1, ξ_2, θ) with a reasonable degree of approximation, although the population considered is extremely small in comparison with a realistic case. Further, the results displayed for parameters' estimation are almost equivalent among the considered techniques. Table 6.2 does not display $\underline{\hat{w}}$ for brevity, as their errors are negligible. On the other hand, $\tilde{P}$ is shown, as it is the quantity of interest to practical applications.

	Value	**Estimate** (expected value ± standard error)		
	a	MLML	MLMM	MMMM
ξ_1	$1.00 \cdot 10^{0}$	$(1.00 \pm 0.00) \cdot 10^{0}$	$(1.00 \pm 0.00) \cdot 10^{0}$	$(1.00 \pm 0.00) \cdot 10^{0}$
ξ_2	$1.00 \cdot 10^{-3}$	$(8.52 \pm 9.81) \cdot 10^{-4}$	$(8.52 \pm 9.81) \cdot 10^{-4}$	$(8.52 \pm 9.81) \cdot 10^{-4}$
θ	$5.00 \cdot 10^{-1}$	$(2.44 \pm 3.07) \cdot 10^{-1}$	$(2.44 \pm 3.07) \cdot 10^{-1}$	$(2.44 \pm 3.07) \cdot 10^{-1}$
$\tilde{P}$	$9.99 \cdot 10^{-1}$	$(9.99 \pm 0.01) \cdot 10^{-1}$	$(9.99 \pm 0.01) \cdot 10^{-1}$	$(9.99 \pm 0.01) \cdot 10^{-1}$
	b	MLML	MLMM	MMMM
ξ_1	$5.00 \cdot 10^{0}$	$(4.35 \pm 1.69) \cdot 10^{0}$	$(5.00 \pm 0.00) \cdot 10^{0}$	$(5.00 \pm 0.00) \cdot 10^{0}$
ξ_2	$1.00 \cdot 10^{-3}$	$(6.51 \pm 16.89) \cdot 10^{-1}$	$(1.27 \pm 3.45) \cdot 10^{-3}$	$(1.27 \pm 3.45) \cdot 10^{-3}$
θ	$5.00 \cdot 10^{-1}$	$(9.68 \pm 28.86) \cdot 10^{-2}$	$(9.68 \pm 28.86) \cdot 10^{-2}$	$(9.68 \pm 28.86) \cdot 10^{-2}$
$\tilde{P}$	$2.00 \cdot 10^{-1}$	$(2.00 \pm 0.00) \cdot 10^{-1}$	$(2.00 \pm 0.00) \cdot 10^{-1}$	$(2.00 \pm 0.00) \cdot 10^{-1}$
	c	MLML	MLMM	MMMM
ξ_1	$0.00 \cdot 10^{0}$	$(3.08 \pm 5.22) \cdot 10^{-5}$	$(3.08 \pm 5.22) \cdot 10^{-5}$	$(3.15 \pm 5.20) \cdot 10^{-5}$
ξ_2	$1.71 \cdot 10^{0}$	$(1.58 \pm 0.06) \cdot 10^{0}$	$(1.58 \pm 0.06) \cdot 10^{0}$	$(1.58 \pm 0.06) \cdot 10^{0}$
θ	$8.00 \cdot 10^{-1}$	$(7.31 \pm 0.67) \cdot 10^{-1}$	$(7.36 \pm 0.53) \cdot 10^{-1}$	$(7.36 \pm 0.53) \cdot 10^{-1}$
$\tilde{P}$	$4.16 \cdot 10^{-1}$	$(4.13 \pm 0.16) \cdot 10^{-1}$	$(4.12 \pm 0.13) \cdot 10^{-1}$	$(4.12 \pm 0.13) \cdot 10^{-1}$
	d	MLML	MLMM	MMMM
ξ_1	$0.00 \cdot 10^{0}$	$(1.76 \pm 1.20) \cdot 10^{-2}$	$(1.76 \pm 1.20) \cdot 10^{-2}$	$(1.99 \pm 1.07) \cdot 10^{-2}$
ξ_2	$1.00 \cdot 10^{0}$	$(9.78 \pm 0.14) \cdot 10^{-1}$	$(9.78 \pm 0.14) \cdot 10^{-1}$	$(9.75 \pm 0.14) \cdot 10^{-1}$
θ	$5.00 \cdot 10^{-2}$	$(8.96 \pm 1.55) \cdot 10^{-2}$	$(9.50 \pm 0.32) \cdot 10^{-2}$	$(9.52 \pm 0.32) \cdot 10^{-2}$
$\tilde{P}$	$8.37 \cdot 10^{-1}$	$(7.79 \pm 0.25) \cdot 10^{-1}$	$(7.72 \pm 0.09) \cdot 10^{-1}$	$(7.70 \pm 0.08) \cdot 10^{-1}$

Table 6.2. Parameters' estimates for the behavioral types *a*-*d* defined in §6.2.1.2, given historical data from 50 buyers observed over 10 years. Data are simulated as per the setup described in §6.3.1, allowing for a comparison between true and estimated parameters.

6.3.2.2 Calibration under a violation of the distributional assumptions

As discussed in §6.2.1.1, assumptions 6.4 and 6.5 are "natural" choices in our framework.

The waiting time between two subsequent observations is exponentially distributed if the occurrence rate is constant through time. The chosen location-scale Gamma distribution (*i.e.*, assumption 6.5) is a flexible generalization, that allows to consider any reasonable non-uniform occurrence rate in a realistic setup.

Further, the categorical distribution considered in assumption 6.4 is the most flexible choice to describe a realistic business context, where time is usually discretized in weeks, months, or quarters.

The categorical distribution can approximate any "actual" credit term distribution, provided that d^1 is chosen small enough. On the other hand, it is worth verifying that assumption 6.5 is robust against distributional misspecifications, as the Gamma distribution could not be able to fit well any possible density profile. It holds $\Delta^0_{jk} \in \mathbb{R}_+$ by definition, excluding any distribution defined outside this domain. However, it is perfectly possible choosing RV's defined in $\mathbb{R}_+$ which are poorly fitted by a Gamma distribution. Multimodal density profiles do not cope with a Gamma probability density function, especially when considering a low number of well-spaced modes. The opposite case of many near modes can be approximated as a whole by widening the volatility of a given unimodal distribution.

Under these premises and without claim to completeness, we suppose that Δ^0_{jk} is distributed according to a bimodal Gamma mixture. Thus, the model is unable to infer the actual density associated with Δ^0_{jk}, since assumption 6.5 excludes multimodal densities. However, the calibration may still lead to a practical result, depending on the quality of the $\tilde{P}$ estimate. Namely, although $\underline{\pi}$ in equation (6.27) does not suffice to describe the considered mixture, it can still return an estimate of $\tilde{P}$ that is adequate to practical purposes.

We consider the behavioral type c, whose location parameter is nullified through the choice $\xi_1 = 0$ (see table 6.1). The following mixture is defined based on type c parameterization:

$$\rho^0_{jk|k-1}\left(v|\underline{\pi}_c, \delta\xi_1\right) \coloneqq \frac{1}{2}\left[\rho^0_{jk|k-1}\left(v\,\middle|0, \xi_2^{(c)}, \theta^{(c)}\right) + \rho^0_{jk|k-1}\left(v\,\middle|\delta\xi_1, \xi_2^{(c)}, \theta^{(c)}\right)\right] \quad (6.54)$$

where $\rho^0_{jk|k-1}(v)$ is defined in equation (6.17). Namely, equation (6.54) introduces an equally weighted mixture between the probability density associated with Δ^0_{jk} in case c and the density of the shifted RV $\Delta^0_{jk} + \delta\xi_1\Delta^1_{jk-1}$. Hence, the resulting distribution is bimodal, where the average distance between the two modes is $\delta\xi_1\langle\Delta^1_{jk-1}\rangle$. In the limit $\delta\xi_1 \to 0$, the mixture degenerates into the location-scale distribution introduced in assumption 6.5, while the two densities in equations (6.17) and (6.54) cannot approximate each other at increasing $\delta\xi_1$.

c	$\tilde{P}$ **Estimate** (expected value ± standard error)			
$\delta\xi_1$	NP	MLML	MLMM	MMMM
0.0	0.42 ± 0.01	0.41 ± 0.01	0.41 ± 0.01	0.41 ± 0.01
0.1	0.42 ± 0.01	0.41 ± 0.01	0.41 ± 0.01	0.41 ± 0.01
0.2	0.42 ± 0.01	0.41 ± 0.01	0.41 ± 0.01	0.41 ± 0.01
0.5	0.42 ± 0.01	0.40 ± 0.01	0.40 ± 0.01	0.40 ± 0.01
1.0	0.40 ± 0.01	0.38 ± 0.01	0.38 ± 0.01	0.38 ± 0.01
2.0	0.32 ± 0.01	0.33 ± 0.01	0.34 ± 0.01	0.34 ± 0.01
5.0	0.21 ± 0.00	0.23 ± 0.01	0.24 ± 0.01	0.24 ± 0.01

Table 6.3. $\tilde{P}$ estimates for the behavioral type c defined in §6.2.1.2, given historical data from 100 buyers observed over 10 years. Data are simulated as per the setup described in §6.3.1, allowing for a comparison between true and estimated parameters. "NP" stands for the nonparametric method introduced in Theorem 6.2.

Results in table 6.3 cope with the intuition. In fact, a greater distance between the two modes of the mixture implies a lesser frequency of issued invoices. Thus, $\tilde{P}$ is expected to decrease at increasing $\delta\xi_1$ values. The nonparametric estimation "NP" of $\tilde{P}$ is considered as a reference, since no distributional assumption is required in Theorem 6.2. Table 6.3 shows that parametric methods are robust against a violation of assumption 6.5, returning results comparable with the ones obtained by the nonparametric estimator.

6.3.3 Inference of the latent default time series in the stationary portfolio case

Let us consider a homogenous, static portfolio that needs a set of parameters $\underline{\pi}$ to be represented across the entire considered time horizon. In this case, Equation (6.52) implies that D_t can be simply inferred from $\hat{D}_t$ by a scale transformation and a time shift.

Figures 6.4 and 6.5 show that the inferred variable $\tilde{D}_t$ fits the latent variable D_t adequately for the behavioral types (a-d) introduced in §6.2.1.2. Furthermore, the approximation $\tilde{P}_{\sup}(\underline{\pi}_j) - \tilde{P}_{\inf}(\underline{\pi}_j) \simeq 0$ (discussed in §6.2.2.2) results are yet to be numerically verified. The distance between D_t and $\hat{D}_t$ is intuitively larger when considering a lower average density of the issued invoices (i.e., compare Figures 6.2 and 6.4).

Figure 6.5 also suggests that the stability of $\tilde{D}_t$ can be improved by adopting a short-term moving average.

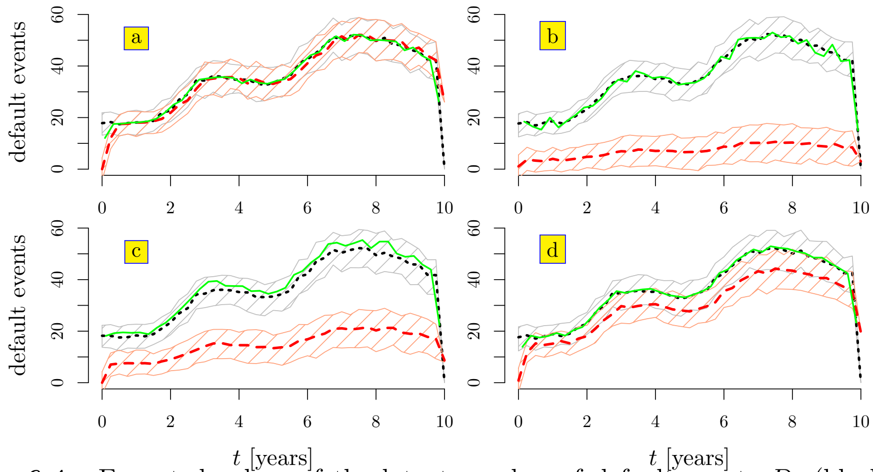

Figure 6.4. Expected values of the latent number of default events D_t (black, dotted line); number of observed claims $\hat{D}_t$ (red, dashed line); and inferred number of default events $\tilde{D}_t$ (green, solid line). Panels *a-d* correspond to the examples in Table 6.1 and Figure 6.2. The gray and red dashed areas represent the standard errors around the expectations for D_t and $\hat{D}_t$, respectively.

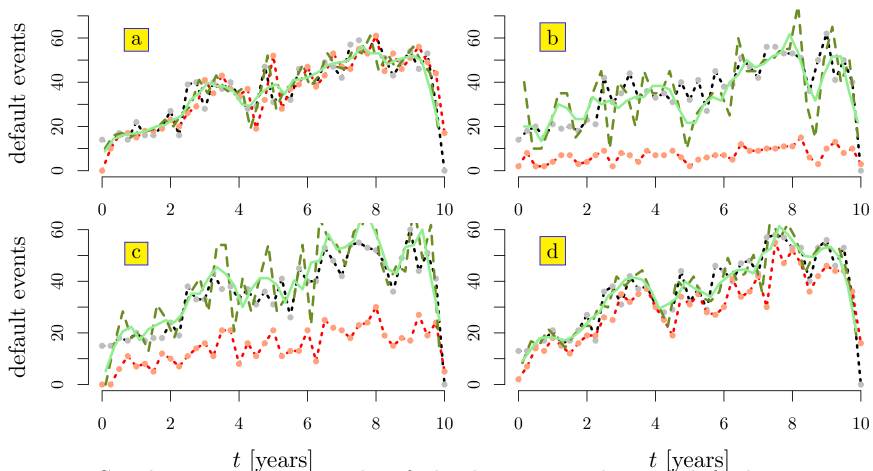

Figure 6.5. Single scenario example of the latent number of default events D_t (black, dotted line); number of observed claims $\hat{D}_t$ (red, dashed line); and inferred number of default events $\tilde{D}_t$ (green lines). The dashed green line represents the simple application of the estimator introduced in Equation (6.52), while the solid green line plots the moving average of the same quantity over three subsequent observations. Panels *a–d* correspond to the examples in Table 6.1 and Figure 6.2.

6.3.4 Inference of the latent default time series in the dynamic portfolio case

The dynamic composition of the CIC liabilities is considered, together with the D_t dynamics already considered in §6.3.3. Let us assume that the CIC has already calibrated the model with different $\hat{\underline{\pi}}$ values that depend on a given classification of possible commercial relationship types. This clustering operation should be performed based on the categorical variables available to the CIC when granting a new credit limit (e.g., economic sectors of both the seller and the buyer). Hence, the CIC associates each new credit limit (i.e., each new couple seller/buyer) to the proper cluster c and, thus, to the proper parameter set $\hat{\underline{\pi}}_c$.

Given these premises, the CIC can update $\langle \tilde{P} \rangle_t$ in Equation (6.52) instantaneously, as the risk composition evolves. This update is relevant for practical purposes. A dynamic risk composition can lower the correlation between $\hat{D}_t$ and D_t. Hence, by applying a predictive macroeconomic model to $\hat{D}_t$, distorted forecasts may be obtained. Macroeconomic variables explanatory to D_t are not explanatory to $D_t - \hat{D}_t$, which is affected only by the policyholder behavior.

However, $\tilde{D}_t$ estimates D_t, thus eliminating the bias arising from latent processes $(t^0_{jk}, \Delta^1_{jk})_{jkt}$.

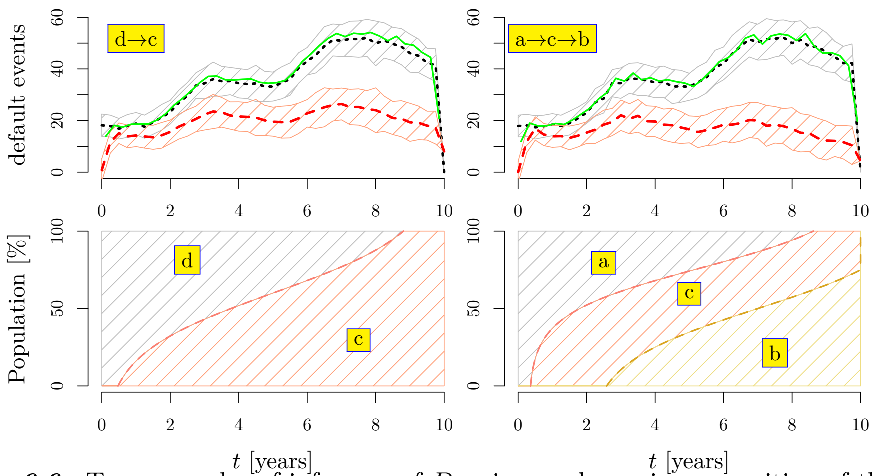

Figure 6.6. Two examples of inference of D_t, given a dynamic composition of the underlying portfolio. The top panels follow the conventions used in Figure 6.4. The bottom panels represent the temporal evolution of the portfolio composition in the two cases. Letters a-d are the types of business relationships introduced in §6.2.1.2.

Figure 6.6 displays two examples of how a time-dependent risk composition can induce a nonlinear map from D_t to $\hat{D}_t$. It is worth noting that the considered portfolios evolve smoothly regarding the composition of behavioral types. Hence, the presence of a strong and irregular variation among the behavioral types (a-d) is not necessary to observe a significant decrease in the correlation between D_t and $\hat{D}_t$. However, even when considering a dynamic portfolio, such as in the examples reported in this section, $\tilde{D}_t$ allows for estimating D_t with a high degree of accuracy.

6.3.5 A simplified case study

In the following, the utility of the presented technique is illustrated by a toy case study. Let us consider a CIC whose ISEs belong only to the economic sectors "H.49.41 - *Agricolture: growing of non-perennial crops*" and "A.01.10 - *Freight transport by road*", where we are following the ATECO sectors classification [111]. We assume that the two sectors may be associated to the behavioral types b (seasonal) and d (irregular/frequent), respectively. Further, let the covered buyers belong to the economic sectors "C.10.00 - *Manifacture of food products*" and "I.56.00 - *Food and beverage service activities*", *inter alia.*

As a working example, the CIC experiences the situation summarized in table 6.4.

Cluster	Economic sector		True values		Past experience			Present situation
ID	ISE	Buyer	p	$\tilde{P}$	N	$\mathbf{E}[\hat{D}]$	$\boldsymbol{\sigma}[\hat{D}]$	N' (requests)
i.	A.01.10	C.10.00	0.031	0.200	1,000	6.2	2.5	1,000
ii.	A.01.10	I.56.00	0.037	0.200	1,000	7.4	2.7	1,000
iii.	H.49.41	C.10.00	0.031	0.837	100	2.6	1.6	1,000

Table 6.4. Information available to the CIC. The default probabilities $p_{\mathrm{C.10.00}}$ and $p_{\mathrm{I.56.00}}$ are inferred from data publicly available in [113].

The CIC aims to reject the riskiest credit limit requests and must decide based on its own past experience. Cluster "iii" is the riskiest one. However, the observed number of claims $\hat{D}$ is distributed as Binomial$(N, p \cdot \tilde{P})$. In the example, the probability that CIC fails to identify the riskiest cluster based on the observed claim frequencies $\hat{f} = \hat{D}/N$ is approximately equal to 11%, that is,

$$\text{Prob}\left(\hat{f}_{\text{iii}} = \max\{\hat{f}_{\text{i}}; \hat{f}_{\text{ii}}; \hat{f}_{\text{iii}}\}\right) \approx 89\%. \tag{6.55}$$

Let us assume that 100 past business relationships are available both for ISEs belonging to A.01.10 and H.49.41 sectors. Thus, the CIC can apply the behavioral model to estimate the claim frequency associated to cluster iii as

$$\tilde{f}_{\text{iii}} := \frac{\tilde{P}_{\text{H.49.41}}}{\tilde{P}_{\text{A.01.10}}}\hat{f}_{\text{i}}. \tag{6.56}$$

Doing so, the probability of identifying the actual riskiest sector becomes

$$\text{Prob}\left(\tilde{f}_{\text{iii}} = \max\{\hat{f}_{\text{i}}; \hat{f}_{\text{ii}}, \tilde{f}_{\text{iii}}\}\right) \approx 97\%, \tag{6.57}$$

where the estimation errors of $\tilde{P}_{\text{A.01.10}}$ and $\tilde{P}_{\text{H.49.41}}$ are considered. Namely, the misjudgement probability is just equal to 3% by using the model, instead of 11%.

6.4 Summary

This chapter introduces a method to handle the lack of information that a CIC faces when estimating the default frequency of a homogenous set of risky buyers based on its own claims database. This method comprises a behavioral model and a parametric estimator.

The model presented in §6.2.1 describes the time distribution of the invoices issued by an ISE to a buyer. Since the insurer can observe a default event only through the expired invoices that generate a claim, the invoice distribution reduces the number of observed default events over a given period relative to the actual number of default events in the same period. A calibration technique to infer the model parameters is provided in §6.2.3 and tested in §6.3.2. The model is parsimonious, although it is possible to represent various behavioral patterns through different parameter choices.

The parametric estimator based on the model parameters produces an accurate estimation of the latent true number of claims, thereby allowing for a precise default frequency estimation. A closed-form expression for the parametric estimator is explicitly derived. Hence, the computational cost of this estimator is negligible. Furthermore, a nonparametric version of the estimator is also provided to suit distributional assumptions that may differ from that chosen to define the model. However, the parametric estimator shows to be robust against distributional misspecifications.

The estimator is applied to data simulated from a historical time series of default frequencies. The considered series has a nontrivial shape, which is transformed by stochastic censoring events generated by policyholder behavior. However, the estimator results are effective in inferring the latent time series of default events from the available time series of claims, even when the dynamics of the underwritten risk composition (in terms of policyholders behavior) induces a nonlinear transformation from the series of latent events to that of observable events. Therefore, the contributions of the business relationships and latent default rates can be separated, and the default rate dynamics can be further modeled and forecasted without distortion by the CIC.

Hence, the presented technique enables the precise estimation of *point-in-time* default probabilities and the corresponding forward-looking claim probabilities in trade credit insurance. Applications range from pricing to risk management under the Solvency 2 framework to financial reporting under the IFRS framework. There is also scope to develop a specific severity model for credit insurance based on the distributional assumptions introduced in this chapter to describe the ISE's behavior, implying the existence of a natural interdependence between the frequency and severity of claims. This problem can be addressed in future studies.

Chapter 7

Concluding Remarks

In the following, we summarize the structure of this thesis and the new results obtained.

The central topic of our study is the claim event in C&S insurance and its occurrence probability. C&S is a peculiar insurance line of business, resembling more specific banking products than the majority of the other non-life insurance lines. The remarkable similarities between C&S products and the banking sector have incentivized a certain lack of attention to this type of insurance by the scientific community.

On the one hand, this deficiency in the literature may be justified by the fact that part of the classic credit risk theory and tools developed for the banking and financial sector can be effectively applied to manage credit insurance and suretyship products to some extent. On the other hand, the similarities among these two sectors have lessened the study of their differences, which are nonetheless worth being investigated and addressed with dedicated tools.

The differences between C&S insurance and the most similar banking products imply reduced information available to the insurer to infer the claims probabilities. Hence, the thesis proposes two new estimators to quantify C&S claims probabilities based on a poor or incomplete data set.

The first estimator allows calibrating the parameters underlying the CreditRisk$^+$ dependence structure with higher precision than the standard calibration techniques. This result is obtained by generalizing the CreditRisk$^+$ assumptions to a multi-period framework and is relevant to the investigated topic, as the CreditRisk$^+$ model is shown to describe well the joint probabilities of the claims arising from a set of C&S policies.

The second estimator handles the information asymmetry between the insurer and the insured in credit insurance. In fact, the classic frequency estimators are biased or not applicable to credit insurance claim frequency due to such an asymmetry. The lack of information is modeled as a collection of stochastic censoring events whose distribution and effects on the claim probability estimation are quantified by a behavioral model developed in this work.

A third research line developed in this thesis regards a particular class of claims in suretyship insurance. It is shown that a properly designed risk appetite framework can completely prevent this type of claim. Further, we numerically verified that the creditworthiness dynamics of the underlying risky subjects does not play a relevant role in the origination of those claims. This fact supports the choice of the CreditRisk$^+$ model as a tool to infer the joint probabilities of future C&S claim events.

The chapter is organized as follows. §7.1 summarizes the content of the first three chapters, where the investigated topic and the related literature are analyzed. §7.2 briefly recall the content of the second half of this thesis, where the research results are presented. Finally, §7.3 displays the lines of research that this work leaves open and that the candidate intends to develop in the next steps of his research activity.

7.1 Analysis of the topic and the related literature

The main features of credit insurance and suretyship products have been presented in chapter 1, where an essential background is also provided about the history and current market development of this insurance line. These products' features are then analyzed in chapter 2 to identify a feasible modeling approach, with particular reference to the estimation of the claim probabilities. The first step to shorten the list of techniques available from banking and finance models was to outline the claim event in this line of business.

As discussed in §2.1, a C&S claim is an absorbing event generated by the breach of an obligation (either financial or not). The worsening of the risky subject's "standing" (*i.e.*, creditworthiness, reputation, or performance capability in a broader sense) is almost never the direct cause of a claim, with the only possible exception of the conversion from a bid bond to a performance bond in the suretyship case. Nonetheless, the macroeconomic context where the risky subjects operate may affect the claim probabilities of comparable subjects in a similar way.

Hence, in §2.2.7, the CreditRisk$^+$ model is identified as a feasible choice to describe the joint distribution of future C&S claim events, both for credit insurance and suretyship products. The model allows generating joint absorbing events (*i.e.*, default events, according to the semantics of the original version). The claim intensity dynamics is not considered, as the model is defined in an uniperiodal framework, and the claims are independent. However, the claim probabilities associated with the underlying risky subjects are modeled jointly. Their dependence structure is defined through a set of latent market factors and a matrix of factor loadings that model the effect of the macroeconomic or market state on each of the risk sources. It is worth noticing that there is no claim to completeness in this choice. Although the CreditRisk$^+$ model appears to be an excellent choice to describe the joint distribution of claims in C&S, it is certainly not the only possible. For example, among the models described in §2.2, also the Vasicek model and the scoring systems appear to be feasible - the latter to some extent, depending on the information available.

In chapter 3, some classic results are introduced with regards to inferring the probability of an absorbing event. The case of incomplete information is also considered, recalling the inference theory developed in the oncology context to handle censoring events, such as the withdrawal. The tools introduced in §§3.2−3.4 are then considered for being applied to estimate marginal C&S claim probabilities. In §3.5

it is discussed how the proposed techniques are not sufficient to address credit insurance, where stochastic censoring events impair observability of claims. On the other hand, information to estimate probability can be considered complete in suretyship, despite being reasonable to expect a poor data set considering the generally low level of the claim probability.

7.2 Original results

Three research problems are posed in the analysis displayed in chapters 4−6.

The first one is related to the general scarcity of data that affects C&S insurance companies, unlike banks and other financial institutions. The CreditRisk$^+$ model appears to be a proper choice for our purposes, once calibrated. However, we have to keep in mind that the model is designed to be calibrated and applied considering a single time scale, but its natural applications both in banking and in insurance contexts require that the time scale needed is equal to one year. This is a "long" sampling period to calibrate a model with a non-parsimonious dependence structure, where many parameters have to be specified. A one-year-long sampling period may significantly reduce the number of observations available in a frequencies time series, implying greater estimation errors. The problem is addressed in chapter 4, where the model is generalized to a multi-period framework and weakly autocorrelated time series are admitted, resulting in the possibility to calibrate the model onto a time scale different from the one used in simulations. It is shown that the estimation error reduces as the calibration time scale shortens, with particular reference to the estimation of the dependence structure parameters. The original results reported in the chapter have been published in *Mathematics*[1] in July 2021 [103] and as a part of the book *Mathematical and Statistical Methods for Actuarial Sciences and Finance - eMAF 2020* [104] in December 2021.

The second problem regards the possibility that a bidder's creditworthiness dynamics plays a significant role in the claim probability of a bid bond. In chapter 5, the case of Italian public tenders is investigated, considering the actual PD volatility of Italian constructors, based on the non-performing loans time series publicly available from the Bank of Italy statistical database. The numerical analysis of the bidding mechanisms highlights how a poor starting price choice plays a major role in the failure of a tender process. This error is due to the procuring entity that promotes the public works underlying the tender and is utterly unrelated to the participants' creditworthiness. Moreover, it is shown how a relatively simple risk appetite framework lets the sureties prevent almost all the bid bond claims, except for those originated by the bidder's default. Hence, the assumption to neglect the creditworthiness dynamics appears to be fully justified when modeling the claim probability in C&S insurance. The original results reported in the chapter have been published in *Mathematics* in September 2021 [105]. The paper has been selected as "*feature paper*" by the Editor.

The third and last problem addressed in this thesis originates from the inadequacy of the classic frequency estimators in the presence of censoring events when consid-

[1] ANVUR recognises *Mathematics* among the "Area 1" - *Mathematics and Informatics* eligible journals.

ering them for credit insurance applications. The information asymmetry between the insured seller and the credit insurance company generates stochastic censoring events that are not tractable by using the methods presented in chapter 3. Hence, chapter 6 proposes a behavioral model designed to fit any sounding business relationship in terms of the temporal distribution of issued invoices. The behavioral model enables the introduction of a new frequency estimator both in parametric and nonparametric forms. Such an estimator can compensate for the bias caused by the stochastic censoring events, enabling a proper inference of claim probability in credit insurance. The original results reported in the chapter have been published in JORS[2] in February 2022 [106].

Appendix C reports information regarding the source code developed to produced the results mentioned above.

7.3 Further developements

C&S claim probability is broad, complex, and scarcely explored to date. Hence, this work has no claim to completeness, either on the research problems that have been investigated or on the methods that we chose or developed to address them. Many other issues related to the ones discussed are worth further investigation in the future. This section briefly summarizes the candidate's intentions with regards to the topics that he intends to address in the next steps of his research work.

- *Credit insurance.* The study reported in chapter 6 considers a key feature of credit insurance (*i.e.*, the information asymmetry existing between the insurer and the insured). However, it neglects at least another feature that affects the frequency of observed claims, namely, the insurer's right to nullify previously granted credit limits. This management action has a non-trivial effect on the claim generation process because already issued invoices may still lead to a claim. However, this is a relevant element of risk mitigation to credit insurers, especially combined with an effective scoring system, frequently updated. How do the scoring system, the credit limits cancelation, and the information asymmetry relate? The candidate intends to address this question in future works. The complexity of this topic can be even increased when considering the hindrance that the insurer withstands in the presence of a *grace period* (see §1.2.1.1) or similar contractual clauses that impair the management actions on credit limits. Finally, a further area worth being investigated is how to model the joint effect of all the elements mentioned above on the severy of a claim before the application of mitigation clauses, recovery actions, and reinsurance (*i.e.*, namely, the *exposure at default* probability distribution).

- *Suretyship.* The study reported in chapter 5 analyses part of the mechanisms that originate a claim from a bid bond policy. This analysis could be improved by simulating a complete Solvency 2 risk profile representation of the

[2]The *Journal of the Operational Research Society* is ANVUR A-rated for "Area 13" -*economics and statistics*. ANVUR also recognises JORS among the "Area 1" - *Mathematics and Informatics* eligible journals.

surety, including missing elements, such as market and reserve risks. Further, the stochastic duration of performance bonds - reported in §1.3 - could be considered in the framework, provided that a reliable data source is publicly available to address this specific topic. Finally, an area worthy of being investigated is the heterogeneity of loss event types guaranteed by the suretyship insurance and their different degree of dependence with the credit risk of the principal. This is implicitly considered by choice of a latent factors model, such as the CreditRisk^{+} model. Nonetheless, the candidate plans a detailed and systematic analysis of the existing relations among suretyship loss event types and credit risk in his future research activity.

- *Modeling generalizations.* The study reported in chapter 4 considers a specific model among the broader set of models that are feasible to describe the joint probability of claim occurrences in a portfolio of C&S policies. The model's assumptions are generalized to a multi-period framework, and it is shown that such an extension of the original model implies an improvement in calibrating the underlying dependence structure when a poor data set is available, as is often the case in C&S insurance. It is legitimate to ask whether one could formally identify the broadest class of models that cope with the features of a C&S claim, as listed in §2.1. Further, assuming this task to be well-posed and achievable and considering the subset of uniperiodal models only, it would be interesting to attempt a generalization of the results introduced in chapter 4. It is worth noticing that part of the theorems reported in §§4.2−4.3 need the closure property of the gamma distribution to be proved. Hence a strict generalization of the same results to other models would not be possible, as they assume different underlying distributions. However, the candidate is planning to investigate to what extent a "weak" generalization of these results is feasible.

- *Different applications.* The estimator developed in chapter 6 is an abstract result, both in its parametric and nonparametric forms. The underlying assumptions are developed based on a description of what happens among the credit insurer, the insured seller, and the underlying buyer when a credit limit is granted on the buyer. Nonetheless, the same assumptions could fit similar problems well (*i.e.*, estimating the probability of observing a specific absorbing event given the presence of stochastic censoring events) in different contexts, such as oncology and epidemiology. As discussed in chapter 3, the exchange of techniques between credit risk and medical statistics is quite common, given the similarities between death and default in a modeling perspective. The candidate hopes that theorems 6.2 and 6.3 may also prove useful for some applications to medical sciences and intends to investigate this possibility deeply.

Finally, the software implementations of the techniques proposed in this thesis are subject to future improvements both in terms of efficiency and algorithms - the latter with particular reference to the cases where the numerical search of a (sub)optimal solution is required to implement the method.

Appendix A

Proofs and details supporting results in chapter 4

This appendix reports some calculations regarding the analyitical results stated in chapter 4.

A.1 Proofs of propositions and theorems

This section presents the proofs of theorems and propositions discussed in chapter 4.

A.1.1 Proof of Theorem 4.1

Proof. Firstly, the statement is proven considering Assumptions 4.3 and 4.1.

Assumption 4.3 implies by construction that $\{Y_i^{(j)}\}_{j=1\dots m}$ is a set of Poisson r.v.'s, which are mutually independent, conditionally on the realization of $\{\mathbf{\Gamma}^{(j)}\}_{j=1\dots m}$. Poisson distribution is closed with respect to addition. Hence

$$\sum_{j=1}^{m} Y_i^{(j)}|\mathbf{\Gamma}^{(j)} \sim \text{Poisson}(p_{i\Sigma}), \tag{A.1}$$

where the distribution parameter is

$$p_{i\Sigma} = \sum_{j=1}^{m} \underbrace{q_i \tfrac{t_j - t_{j-1}}{T-t}}_{q_i^{(j)}} \left(\omega_{i0} + \sum_{k=1}^{K} \omega_{ik} \Gamma_k^{(j)} \right). \tag{A.2}$$

Eq. (4.20) in Assumption 4.3, the choice $\xi_{kj} = 1$ and the scaling property of Gamma distribution imply that

$$\tfrac{t_j - t_{j-1}}{T-t} \Gamma_k^{(j)} \sim \text{Gamma}\left(\sigma_k^{-2} \tfrac{t_j - t_{j-1}}{T-t}, \sigma_k^2 \right) \tag{A.3}$$

Furthermore, Assumption 4.5 and the fact that independent Gamma r.v.'s with the same scale parameter are closed with respect to addition imply that

$$\sum_{j=1}^{m} \tfrac{t_j - t_{j-1}}{T-t} \Gamma_k^{(j)} \sim \text{Gamma}\left(\sigma_k^{-2}, \sigma_k^2 \right). \tag{A.4}$$

Hence $\sum_{j=1}^{m} \frac{t_j - t_{j-1}}{T-t} \Gamma_k^{(j)} \equiv \Gamma_k$ and so $\sum_{j=1}^{m} Y_i^{(j)} \equiv Y_i$. This implies that $(t, T]$ verifies Assumption 4.1.

The proof above can be extended to the exponential case - *i.e.* when considering Assumptions 4.4 and 4.2 instead of Assumptions 4.3 and 4.1. The form of parameter $p_{i\Sigma}$ in (A.2) can be obtained also can be obtained also from Assumption 4.4. In fact, the substitution $Y_i^{(j)} \mapsto \widetilde{Y}_i^{(j)}$ implies that $\widetilde{Y}_i \sim$ Bernoulli$(\tilde{p}_i)$ where

$$\ln(1 - \tilde{p}_i) = \ln \prod_{j=1}^{m} \left(1 - \tilde{p}_i^{(j)}\right) = \sum_{j=1}^{m} q_i^{\frac{t_j - t_{j-1}}{T-t}} \left(\omega_{i0} + \sum_{k=1}^{K} \omega_{ik} \Gamma_k^{(j)}\right). \tag{A.5}$$

Considering eq. (A.5) instead of (A.2), the proof presented above holds for the $\widetilde{Y}_i$ representation of risks, *ceteris paribus*, implying that $(t, T]$ verifies Assumption 4.2. ∎

A.1.2 Proof of Theorem 4.2

Proof. The same arguments that lead to (A.2) or to (A.5) in proof of Theorem 4.1 are still valid in this case. Hence, it suffices to prove that mean and variance of the latent variable

$$\Gamma'_k := \sum_{j=1}^{m} \frac{t_j - t_{j-1}}{T-t} \Gamma_k^{(j)} = \frac{1}{m} \sum_{j=1}^{m} \Gamma_k^{(j)}$$

remain consistent with CreditRisk$^+$ requirements, stated in Assumption 4.1. It holds $\mathbf{E}[\Gamma'_k] = 1$, since $\mathbf{E}[\Gamma_k^{(j)}] = 1$. Moreover, the coefficient ξ_{jk} compensates the bias introduced in $\mathbf{var}[\Gamma'_k]$ by the fact that $\Gamma_k^{(j)}$ $(j = 1 \ldots m)$ are autocorrelated according to the ACF ϱ_{xk}:

$$\begin{aligned}
\mathbf{var}\left[\sum_{j=1}^{m} \Gamma_k^{(j)}\right] &= \sum_{j=1}^{m} \mathbf{var}\left[\Gamma_k^{(j)}\right] + \sum_{j=1}^{m} \sum_{j' \neq j} \mathbf{cov}\left[\Gamma_k^{(j)}, \Gamma_k^{(j')}\right] \\
&= \mathbf{var}\left[\Gamma_k^{(1)}\right] \underbrace{\left(m + 2 \sum_{x=1}^{m-1} (m - x)\, \varrho_{xk}\right)}_{m\xi_{kj}^{-2}}
\end{aligned}$$

which implies $\mathbf{var}[\Gamma'_k] = \sigma_k^2$ directly.

The fact that Γ'_k is Gamma distributed is imposed in Assumption 4.6, implying that $\Gamma'_k \equiv \Gamma_k$ and so that Assumption 4.1 is satisfied. ∎

A.1.3 Proof of Proposition 4.2

Proof. Given a time interval $(t, T] \subseteq (t_a, t_b]$ and a uniform partition $(j = 1 \ldots m)$ over $(t, T]$, Assumptions 4.3 and 4.5 imply that $(t, T]$ verifies Assumption 4.1 by Theorem 4.1. Assumption 4.7 guarantees the convergence of F_h to $\mathbf{E}[F_h|\mathbf{\Gamma}]$ and of $F_h^{(j)}$ to $\mathbf{E}[F_h^{(j)}|\mathbf{\Gamma}^{(j)}]$, where we recall that $F_h = F_h(t, T)$.

For any interval $(t, T] \subseteq (t_a, t_b]$ and any pair of clusters $c_h, c_{h'}$, definitions (4.34), (4.35) and Assumption 4.5 imply that the covariance between F_{mh} and $F_{mh'}$ is given by

$$\mathbf{cov}\left(F_{mh}, F_{mh'}\right) = \prod_{j=1}^{m} \left[\mathbf{cov}\left(F_h^{(j)} F_{h'}^{(j)}\right) + s_h^{(j)} s_{h'}^{(j)}\right] - s_h s_{h'} \tag{A.6}$$

Since all the considered subintervals $(t_{j-1}, t_j]$ have the same length $\delta_m = t_j - t_{j-1}$, the frequencies $F_h^{(j)}$ are *i.i.d.*, so that the above expression simplifies to:

$$\mathbf{cov}\left(F_{mh}, F_{mh'}\right) + s_h s_{h'} = \left[\mathbf{cov}\left(F_h^{(j)}, F_{h'}^{(j)}\right) + s_h^{(j)} s_{h'}^{(j)}\right]^m \tag{A.7}$$

for any $j = 1, \dots, m$.

Each cluster c_h is supposed to be homogenous by definition: *i.e.* $\boldsymbol{\omega}^{(i)} = \boldsymbol{\omega}^{(h)}$ for each risk $Y_i \in c_h$. Hence, distributional Assumptions 4.1 and 4.3 imply that both $F_h \xrightarrow{n_h(t)\to\infty} \mathbf{E}[F_h|\boldsymbol{\Gamma}]$ and $F_h^{(j)} \xrightarrow{n_h(t_j)\to\infty} \mathbf{E}[F_h^{(j)}|\boldsymbol{\Gamma}^{(j)}]$ are sample estimators of the parameters $p_h(\boldsymbol{\Gamma}) \coloneqq q_h(\omega_{h0} + \sum_k \omega_{hk}\Gamma_k)$ and $p_h^{(j)}(\boldsymbol{\Gamma}^{(j)})$ respectively, leading to the equivalence relation

$$F_{mh} = F_h = \hat{q}_h\left(\omega_{h0} + \sum_{k=1}^{K} \omega_{hk}\Gamma_k\right), \tag{A.8}$$

therefore both $F_{mh}(t, T)$ and $F_h(t, T)$ are estimators of the default frequency for the $(t, T]$ interval. Thus, eq. (A.7) can be rewritten as:

$$\mathbf{cov}\left(F_h, F_{h'}\right) + s_h s_{h'} = \left[\mathbf{cov}\left(F_h^{(j)}, F_{h'}^{(j)}\right) + s_h^{(j)} s_{h'}^{(j)}\right]^m \tag{A.9}$$

and, since $m = (T - t)/\delta_m$,

$$\left[\mathbf{cov}\left(F_h, F_{h'}\right) + s_h s_{h'}\right]^{1/(T-t)} = \left[\mathbf{cov}\left(F_h^{(j)}, F_{h'}^{(j)}\right) + s_h^{(j)} s_{h'}^{(j)}\right]^{1/\delta_m}. \tag{A.10}$$

To complete the proof, let $(t, T]$ and $(t', T']$ be two sub-intervals of $(t_a, t_b]$, such that $(T - t)/(T' - t') \in \mathbb{Q}$. Hence, $\mathrm{GCD}\{T - t; T' - t'\} \eqqcolon \overline{\delta} \in \mathbb{R}_+$ exists. $\overline{\delta}$ can be used as the mesh to define two uniform partitions over the two considered intervals.

Given these partitions, (A.10) can be applied both to $T - t$ and to $T' - t'$ leading to

$$\begin{aligned}&\left[\mathbf{cov}\left(F_h(t, T), F_{h'}(t, T)\right) + s_h(t, T) s_{h'}(t, T)\right]^{1/(T-t)} = \\ &\left[\mathbf{cov}\left(F_h(t', T'), F_{h'}(t', T')\right) + s_h(t', T') s_{h'}(t', T')\right]^{1/(T'-t')}\end{aligned}$$

and completing the proof. The requirement $(T - t)/(T' - t') \in \mathbb{Q}$ can be easily weakened by the convergence of finite continued fractions with an increasing number of terms, until the desired degree of precision is reached. ∎

A.1.4 Proof of Proposition 4.3

Proof. Given a time interval $(t, T] \subseteq (t_a, t_b]$ and a uniform partition $(j = 1 \dots m)$ over $(t, T]$, Assumptions 4.4 and 4.5 imply that $(t, T]$ verifies Assumption 4.2 by Theorem 4.1.

Assumption 4.7 guarantees the convergence of L_h to $\mathbf{E}[L_h|\mathbf{\Gamma}]$, where we recall that $L_h = L_h(t,T)$. Furthermore, it holds by definition that $\mathbf{E}[L_h|\mathbf{\Gamma}] = p_h(\mathbf{\Gamma})$, where the notation p_h has been introduced in the proof of Proposition 4.2.

The same apply to $L_h^{(j)}$ ($j = 1 \dots m$) for each uniform partition of $(t,T]$ considered: indeed, Assumption 4.7 implies $L_h^{(j)} \to \mathbf{E}[L_h^{(j)}|\mathbf{\Gamma}^{(j)}] = p_h^{(j)}(\mathbf{\Gamma})$.

Since $p_h(\mathbf{\Gamma}) = \sum_{j=1}^m p_h^{(j)}(\mathbf{\Gamma}^{(j)})$ and given that the partition is uniform, it holds $L_h = mL_h^{(j)}$ for each $j = 1 \dots m$. Since $m \coloneqq (T-t)/\delta_m$, we have

$$\frac{1}{T-t} L_h = \frac{1}{\delta_m} L_h^{(j)} \tag{A.11}$$

Assumption 4.5 and eq. (A.11) imply that

$$\frac{1}{T-t} \mathbf{cov}[L_h, L_{h'}] = \frac{1}{\delta_m} \mathbf{cov}[L_h^{(j)}, L_{h'}^{(j)}] \tag{A.12}$$

for each considered pair of clusters $c_h, c_{h'}$. The proof is completed by the same argument used in proof of Proposition 4.2, after eq. (A.10). ∎

A.1.5 Proof of Theorem 4.3

Proof. Assumptions 4.3, 4.5 and $\xi_{kj} = 1$ imply Assumption 4.1 by Theorem 4.1. The same theorem implies Assumption 4.2 in case Assumption 4.4 is considered instead of Assumption 4.3, *ceteris paribus*. Furthermore, Assumptions 4.3, 4.5, 4.7 and $\xi_{kj} = 1$ imply that

$$\hat{A}_{hh'}^{(L,m)} = \frac{1}{q_h q_{h'}} \left[\left(\hat{c}_{hh'}^{(Lm)} \right)^m - s_h s_{h'} - \delta_{hh'} \frac{q_h}{n_h} \right] \tag{A.13}$$

by Proposition 4.2, for any $j = 1 \dots m$ and $h, h' = 1 \dots H$. Analogously, considering Assumption 4.4 instead of Assumption 4.3, it holds

$$\hat{A}_{hh'}^{(E,m)} = \frac{m}{q_h q_{h'}} \hat{c}_{hh'}^{(Em)} \tag{A.14}$$

by Proposition 4.3, for any $j = 1 \dots m$ and $h, h' = 1 \dots H$.

The next step of the proof is showing that $\Gamma_k \dot{\sim} \mathcal{N}(1, \beta_k)$ in the limit $\sigma_k \to 0^+$. In fact, both Assumptions 4.3 and 4.4 state that

$$\Gamma_k^{(j)} \sim \Gamma\left(\tfrac{1}{m\beta_k}, m\beta_k\right), \quad \mathbf{E}\left[\Gamma_k^{(j)}\right] = 1, \quad \mathbf{var}\left[\Gamma_k^{(j)}\right] = m\beta_k, \quad j = 1, \dots, m.$$

Hence their probability densities $dF_k(x)$ satisfy the following:

$$dF_k(x) \propto x^{(m\beta_k)^{-1}-1} \exp\left(-(m\beta_k)^{-1} x\right) dx \tag{A.15}$$

Since it holds $(m\beta_k)^{-1} - 1 \xrightarrow{\sigma_k \to 0^+} (m\beta_k)^{-1}$, we have

$$\lim_{\sigma_k \to 0^+} dF_k(x) \propto \exp\left(\frac{\ln x - x}{m\beta_k}\right) dx. \tag{A.16}$$

By introducing the auxiliary variable $x' := x - 1$ and replacing $\ln(1 + x')$ with the first three terms of its Maclaurin series, relation (A.16) can be equivalently written as

$$\lim_{\sigma_k \to 0^+} dF_k\left(x(x')\right) \propto \exp\left(-\frac{x'^2}{2m\beta_k}\right) dx' \tag{A.17}$$

In the limit $\sigma_k = \beta_k \to 0^+$, eq. (A.17) implies that

$$\Gamma_k^{(j)} \sim \mathcal{N}\left(\mu = 1, \sigma^2 = m\beta_k\right). \tag{A.18}$$

Hence it holds that each $F_h^{(j)}$ is normally distributed, with variance $m\sigma_h^2 := m \sum_k \omega_{hk}\beta_k$ - when considering the linear case (*i.e.* Assumptions 4.1 and 4.3). Analogously, also each $L_h^{(j)}$ is normally distributed in the exponential case (*i.e.* Assumptions 4.2 and 4.4).

Considering the market factors - as well as the historical observations of default frequency - as normal random variables is relevant to prove the theorem, since it implies that the covariance matrix estimators $\hat{c}^{(Lm)}$ and $\hat{c}^{(Em)}$ are Wishart distributed. Hence the variance associated to a given matrix element is

$$\mathbf{var}\left[\hat{c}_{hh'}^{(m)}\right] = \frac{m^2}{m \cdot n - 1}\left(\rho_{hh'}^2 + 1\right)\sigma_h^2\sigma_{h'}^2 \tag{A.19}$$

in both linear and exponential cases. In the exponential case eq. (A.19) is equivalent to the following

$$\mathbf{var}\left[\hat{c}_{hh'}^{(Em)}\right] = \frac{1}{m \cdot n - 1}\left[\left(c_{hh'}^{(Em)}\right)^2 + c_{hh}^{(Em)} c_{h'h'}^{(Em)}\right] \tag{A.20}$$

while the same is not true in the linear case. Given equation (A.19), it is possible to prove equation (4.54) separately in the two cases.

Proof in the linear case. Proposition 4.2 implies

$$\begin{aligned}\mathbf{var}\left[\hat{A}_{hh'}^{(L,m)}\right] &= \frac{1}{(q_h q_{h'})^2}\mathbf{var}\left[\left(\hat{c}_{hh'}^{(Lm)}\right)^m\right] = \frac{1}{(q_h q_{h'})^2}\left[\left(\mathbf{E}\left[\left(\hat{c}_{hh'}^{(Lm)}\right)^2\right]\right)^m - \left(\mathbf{E}\left[\hat{c}_{hh'}^{(Lm)}\right]\right)^{2m}\right] \\ &= \frac{1}{(q_h q_{h'})^2}\left[\left(\mathbf{var}\left[\hat{c}_{hh'}^{(Lm)}\right] + \left(\mathbf{E}\left[\hat{c}_{hh'}^{(Lm)}\right]\right)^2\right)^m - \left(\mathbf{E}\left[\hat{c}_{hh'}^{(Lm)}\right]\right)^{2m}\right] \end{aligned} \tag{A.21}$$

In the limit $\boldsymbol{\sigma} \to \mathbf{0}^+$ the binomial above can be replaced with its leading term. Hence

$$\mathbf{var}\left[\hat{A}_{hh'}^{(L,m)}\right] = \frac{1}{(q_h q_{h'})^2}\mathbf{var}\left[\hat{c}_{hh'}^{(Lm)}\right]\left(\mathbf{E}\left[\hat{c}_{hh'}^{(Lm)}\right]\right)^{2(m-1)} \tag{A.22}$$

By applying equation (A.19) we have

$$\begin{aligned}\mathbf{var}\left[\hat{A}_{hh'}^{(L,m)}\right] &= \frac{1}{(q_h q_{h'})^2}\frac{m^2}{m \cdot n - 1}\left(\rho_{hh'}^2 + 1\right)\sigma_h^2\sigma_{h'}^2\left(c_{hh'}^{(Lm)}\right)^{2(m-1)} \\ &= \frac{1}{(q_h q_{h'})^2}\frac{1}{m \cdot n - 1}\frac{\rho_{hh'}^2 + 1}{\rho_{hh'}^2}\left(c_{hh'}^{(Lm)} - s_h^{(j)} s_{h'}^{(j)}\right)^2\left(c_{hh'}^{(Lm)}\right)^{2(m-1)}\end{aligned} \tag{A.23}$$

Applying Proposition 4.2 once again we have $\left(c_{hh'}^{(Lm)}\right)^{2m} = \left(c_{hh'}^{(L1)}\right)^2$. Furthermore, we have $s_h^{(j)} s_{h'}^{(j)} \left(c_{hh'}^{(Lm)}\right)^{2(m-1)} \xrightarrow{\boldsymbol{\sigma} \to \mathbf{0}^+} \left(s_h^{(j)} s_{h'}^{(j)}\right)^{2m} = s_h^2 s_{h'}^2$. Hence it holds

$$\mathbf{var}\left[\hat{A}_{hh'}^{(L,m)}\right] = \frac{1}{(q_h q_{h'})^2} \frac{1}{m \cdot n - 1} \frac{\rho_{hh'}^2 + 1}{\rho_{hh'}^2} \left[\left(c_{hh'}^{(L1)}\right)^2 - s_h^2 s_{h'}^2\right] \tag{A.24}$$

and thus the ratio $\mathbf{var}\left[\hat{A}_{hh'}^{(L,m)}\right] / \mathbf{var}\left[\hat{A}_{hh'}^{(L,1)}\right]$ verifies equation (4.54), completing the proof for the linear case.

Proof in the exponential case. Equation (A.20) and Proposition 4.3 imply

$$\begin{aligned} \mathbf{var}\left[\hat{A}_{hh'}^{(E,m)}\right] &= \frac{m^2}{(q_h q_{h'})^2} \mathbf{var}\left[\hat{c}_{hh'}^{(Em)}\right] = \frac{m^2}{(q_h q_{h'})^2} \frac{1}{m \cdot n - 1} \left[\left(c_{hh'}^{(Em)}\right)^2 + c_{hh}^{(Em)} c_{h'h'}^{(Em)}\right] \\ &= \frac{1}{(q_h q_{h'})^2} \frac{1}{m \cdot n - 1} \left[\left(c_{hh'}^{(E1)}\right)^2 + c_{hh}^{(E1)} c_{h'h'}^{(E1)}\right] \end{aligned} \tag{A.25}$$

The latter implies that in case $m = 1$ we have

$$\mathbf{var}\left[\hat{A}_{hh'}^{(E,1)}\right] = \frac{1}{(q_h q_{h'})^2} \frac{1}{n - 1} \left[\left(c_{hh'}^{(E1)}\right)^2 + c_{hh}^{(E1)} c_{h'h'}^{(E1)}\right] \tag{A.26}$$

Hence, the ratio $\mathbf{var}\left[\hat{A}_{hh'}^{(E,m)}\right] / \mathbf{var}\left[\hat{A}_{hh'}^{(E,1)}\right]$ verifies equation (4.54), completing the proof for the exponential case. ∎

A.2 Covariance estimation error in presence of autocorrelation

In this section a generalization of equation (4.54) is provided, considering the presence of autocorrelation. Only the exponential case is discussed, because a closed form for $\hat{A}_{hh'}^{(E,m)}$ is still available when autocorrelation has to be considered - while only a second order approximation has been computed for the linear case $\hat{A}_{hh'}^{(L,m)}$. A comparison between equations (4.41) and (4.49) allows us to generalize Proposition 4.3.

$$c_{hh'}^{(E1)} = m c_{hh'}^{(Em)} + 2 \sum_{x=1}^{m-1} (m - x)_x c_{hh'}^{(Em)} \tag{A.27}$$

where

$${}_x c_{hh'}^{(Em)} := \mathbf{cov}\left[L_h^{(j)}, L_{h'}^{(j+x)}\right] \tag{A.28}$$

It holds by definition

$$\begin{aligned} c_{hh'}^{(Em)} &= \frac{q_h q_{h'}}{m^2} \sum_{k=1}^{K} \omega_{hk} \omega_{h'k} \mathbf{var}\left[\Gamma_k^{(j)}\right] \\ {}_x c_{hh'}^{(Em)} &= \frac{q_h q_{h'}}{m^2} \sum_{k=1}^{K} \omega_{hk} \omega_{h'k} \mathbf{cov}\left[\Gamma_k^{(j)}, \Gamma_k^{(j+x)}\right] \end{aligned}$$

Hence, Assumption 4.6 implies

$$\mathbf{E}\left[{}_x\hat{c}_{hh'}^{(Em)}\right] = \mathbf{E}\left[\frac{q_h q_{h'}}{m^2}\sum_{k=1}^{K}\omega_{hk}\omega_{h'k}\varrho_{xk}\mathbf{var}\left[\Gamma_k^{(j)}\right]\right] = {}_x\tilde{\varrho}_{hh'} \tag{A.29}$$

where

$${}_x\tilde{\varrho}_{hh'} := \frac{\sum_{k=1}^{K}\tilde{w}_{khh'}\varrho_{xk}}{\sum_{k=1}^{K}\tilde{w}_{khh'}}; \quad \tilde{w}_{khh'} := \omega_{hk}\omega_{h'k}m\xi_k^2\sigma_k^2 \tag{A.30}$$

Furthermore, applying equation (A.19), it follows that

$$\begin{aligned}\mathbf{var}\left[{}_x\hat{c}_{hh'}^{(Em)}\right] &= \frac{1}{m\cdot n-1}\left(\mathbf{E}\left[{}_x\hat{c}_{hh'}^{(Em)}\right]^2 + \mathbf{E}\left[\mathbf{c\hat{o}v}\left[L_h^{(j)}, L_h^{(j)}\right]\right]\mathbf{E}\left[\mathbf{c\hat{o}v}\left[L_{h'}^{(j+x)}, L_{h'}^{(j+x)}\right]\right]\right)\\ &= \frac{1}{m\cdot n-1}\left({}_x\tilde{\varrho}_{hh'}^2\left(c_{hh'}^{(Em)}\right)^2 + c_{hh}^{(Em)}c_{h'h'}^{(Em)}\right)\\ &= \mathbf{var}\left[\hat{c}_{hh'}^{(Em)}\right] - \frac{1-{}_x\tilde{\varrho}_{hh'}^2}{m\cdot n-1}\left(c_{hh'}^{(Em)}\right)^2 \end{aligned}\tag{A.31}$$

Equation (A.29) leads to another version of equation (A.27)

$$c_{hh'}^{(Em)} = \tfrac{1}{m}\left(1 + 2\sum_{x=1}^{m-1}(1-\tfrac{x}{m})_x\tilde{\varrho}_{hh'}\right)^{-1} c_{hh'}^{(E1)} \tag{A.32}$$

From eq. (4.49) we have

$$\mathbf{var}\left[\hat{A}_{hh'}^{(E,m)}\right] = \frac{m^2}{(q_h q_{h'})^2}\mathbf{var}\left[\hat{c}_{hh'}^{(Em)} + 2\sum_{x=1}^{m-1}(1-\tfrac{x}{m})_x\hat{c}_{hh'}^{(Em)}\right] \tag{A.33}$$

Equation (A.33) implies that $\mathbf{var}\left[\hat{A}_{hh'}^{(E,m)}\right]$ depends on the correlation matrix $\varrho_{xx'}^{(\hat{c})}$ among the considered covariance estimators ${}_x\hat{c}_{hh'}^{(Em)}$ $(x = 0, 1, \dots)$, as shown below by choosing an equivalent expression for the RHS:

$$\mathbf{var}\left[\hat{A}_{hh'}^{(E,m)}\right] = \frac{m^2}{(q_h q_{h'})^2}\sum_{x,x'=0}^{m-1}\varrho_{xx'}^{(\hat{c})}\ {}_x s_{hh'}\ {}_{x'}s_{hh'} \tag{A.34}$$

where

$${}_x s_{hh'} := (2-\delta_{0x})(1-\tfrac{x}{m})(\mathbf{var}\left[{}_x\hat{c}_{hh'}^{(Em)}\right])^{\frac{1}{2}} \tag{A.35}$$

In case the covariance estimators are independent from each other (i.e. $\varrho_{xx'}^{(\hat{c})} = \delta_{xx'}$), an inferior limit to the considered variance is obtained

$$\mathbf{var}\left[\hat{A}_{hh'}^{(E,m)}\right] \geq \tfrac{m^2}{(q_h q_{h'})^2}\left(\mathbf{var}\left[\hat{c}_{hh'}^{(Em)}\right] + 4\sum_{x=1}^{m-1}(1-\tfrac{x}{m})^2\mathbf{var}\left[{}_x\hat{c}_{hh'}^{(Em)}\right]\right) \tag{A.36}$$

Equation (A.31) can be substituted into eq. (A.34). Hence, RHS of inequality (A.36) becomes

$$\mathbf{var}\left[\hat{c}_{hh'}^{(Em)}\right]\tfrac{m^2}{(q_h q_{h'})^2}\left[1 + 4\sum_{x=1}^{m-1}(1-\tfrac{x}{m})^2\left(1 - \tfrac{1-{}_x\tilde{\varrho}_{hh'}^2}{m\cdot n-1}\mathbf{cv}^{-2}\left[\hat{c}_{hh'}^{(Em)}\right]\right)\right]$$

where the notation **cv** $[\cdot]$ stands for the coefficient of variation.
$(c_{hh'}^{(Em)})^2$ and **var** $\left[\hat{c}_{hh'}^{(Em)}\right]$ can be expressed by using equation (A.32). Hence eq. (A.36) can be used to estimate an inferior limit to $\varepsilon\left[\hat{A}_{hh'}^{(E,m)}\right]$ in the gaussian regime. A superior limit for the same quantity can be computed as well, imposing $\varrho_{xx'}^{(\hat{c})} = 1$ for each considered x, x'.

Remark A.1. *Equation* (A.34) *does not converge to* (A.25) *in the limit* $\varrho_{xk} \to 0 \Rightarrow {}_x\tilde{\varrho}_{hh'} \to 0$. *This copes with the fact that assuming* $\varrho_{xk} = 0$ *in eq.* (A.25) *implies a lesser error than measuring it.*

Appendix B

Calibration of the behavioral model proposed in chapter 6

This appendix reports and comments samples of the R source codes developed in this work, with specific reference to the behavioral model introduced in chapter 6 and its calibration. §B.1 provides an insight on the simulation of the insured sellers' behavior. This part of the engine is briefly described in §6.3.1, and enables the testing activity reported in §6.3.2 on the calibration techniques proposed in §6.2.3. The next §B.2 discloses the numerical investigation done to achieve the result displayed in figure 6.3. This is relevant step in the developement of the proposed approach to the model calibration, as it highlights the numerical instability of equation (6.40). Finally, §B.3 reports the source code actually used in calibrating the behavioral model.

B.1 Simulating the business relationships

The R code below reproduces results shown in figure 6.2. Namely, the code generates a single scenario in a Monte Carlo simulation of a given business relationship.

```
param = data.frame(xi1 = c(1.000, 5.000, 0.000, 0.000),
                   xi2   = c(0.001, 0.001, 1.714, 1.000),
                   theta = c(0.500, 0.500, 0.800, 0.050),
                   w1    = c(0.000, 0.000, 0.000, 0.250),
                   w2    = c(0.000, 1.000, 0.000, 0.000),
                   w3    = c(0.000, 0.000, 0.500, 0.000),
                   w4    = c(1.000, 0.000, 0.500, 0.750),
                   d1    = array(1/12,dim=4))

CASE_ID = NA # 1-Periodic; 2-Seasonal, 3-Irregular; 4-Intermediate
cw = NA
xi2_prime = NA
# distribution of delta1
rdelta1 <- function(d1_ = param$d1[CASE_ID], cw_ = cw) {
  return( d1_*which.max(runif(1)<=cw) )
}
```

```
# (conditioned) distribution of t0 and delta0
rt0delta0 <- function(t0_previous,
                      delta1_previous,
                      xi1_      = param$xi1[CASE_ID],
                      xi2_prime_ = xi2_prime,
                      theta_    = param$theta[CASE_ID]) {
  delta0 = xi1_*delta1_previous +
          rgamma(1,
                 shape = xi2_prime_*delta1_previous,
                 scale = theta_)
  t0    = t0_previous + delta0
  return(c(t0,delta0))
}
# generate pseudo-historical invoices
HORIZON = 2 # observation period (years)
EXAMPLES = length(param$xi1) # number of examples
invoice = data.frame(t0   = double(),
                     t1     = double(),
                     delta0 = double(),
                     delta1 = double())
invoices = rep(list(invoice), EXAMPLES)
set.seed(1)
for(j in 1:EXAMPLES) {
  CASE_ID  = j
  cw       = cumsum(param[CASE_ID,4:7])
  xi2_prime = param$xi2[CASE_ID]/param$theta[CASE_ID]
  aus_invoice <- runif(1)*0.1 #initial delta0 and t0
  aus_invoice <- c(aus_invoice, aus_invoice, rdelta1())
  k=1
  while(aus_invoice[1] < HORIZON) {
    invoices[[j]] <- rbind(invoices[[j]], data.frame(t0 = aus_invoice[1],
    t1     = aus_invoice[1]+aus_invoice[3],
    delta0 = aus_invoice[2],
    delta1 = aus_invoice[3]))
    aus_invoice  <- rt0delta0(invoices[[j]]$t0[k],
    invoices[[j]]$delta1[k])
    aus_invoice  <- c(aus_invoice, rdelta1())
    k=k+1
  }
}
```

B.2 Stability of $\overline{\nabla} \ln \mathcal{L}$

This section displays the source code used to generate the results in figure 6.3.

First, a set of 5.000 scenarios is generated by using a code similar to the one shown in paragraph B.1, choosing CASE_ID= 3 and HORIZON= 5.

Then, the following measures out_consistency are taken across the scenarios (SETTING= 5000 and NBUYERS= 100).

```
for(s in 1:SETTING) {
```

```
Delta0 = c()
Delta1 = c()
for(b in 1:NBUYERS) {
Delta0 <- c( Delta0, invoices_settings[[s]][[b]]$delta0[-1] )
Delta1 <- c( Delta1,
    invoices_settings[[s]][[b]]$delta1[-length(invoices_settings[[2]][[b]]$delta1)]
    )
}
Delta0_m <- mean(Delta0)
Delta1_m <- mean(Delta1)
# panel (i)
out_consistency$test_dxi1[s] <- grad_d_xi1(param$theta[CASE_ID],
                                          param$xi1[CASE_ID],
                                          xi2_prime,
                                          Delta0, Delta1,
                                          Delta0_m, Delta1_m)
# panel (ii)
out_consistency$test_dx2prime[s] <- grad_d_xi2prime(param$theta[CASE_ID],
                                                  param$xi1[CASE_ID],
                                                  xi2_prime,
                                                  Delta0, Delta1,
                                                  Delta0_m, Delta1_m)
# panel (iii)
out_consistency$test_dtheta[s] <- grad_d_theta(param$theta[CASE_ID],
                                              param$xi1[CASE_ID],
                                              xi2_prime,
                                              Delta0, Delta1,
                                              Delta0_m, Delta1_m)
}
```

where the functions used above are defined as follows

```
grad_d_xi1 <- function(theta, xi1, xi2prime,
                      Delta0, Delta1,
                      Delta0_m, Delta1_m) {
  return(
    Delta1_m/theta -
    mean( Delta1 * (xi2prime*Delta1-1) / (Delta0-xi1*Delta1))
  )
}
grad_d_xi2prime <- function(theta, xi1, xi2prime,
                           Delta0, Delta1,
                           Delta0_m, Delta1_m) {
  return(
    mean(Delta1 *
    ( log(((Delta0-xi1*Delta1))/theta)
     -digamma(xi2prime*Delta1))
    )
  )
}
grad_d_theta <- function(theta, xi1, xi2prime,
                        Delta0, Delta1,
                        Delta0_m, Delta1_m) {
```

```
  return(
    mean(xi2prime*Delta1-(Delta0-xi1*Delta1)/theta)
  )
}
```

B.3 Results of numerical optimization

This section displays the numerical outcomes of the three calibration methods introduced in §§6.2.3 and summarized in §6.3.2. The following graphical evidence confirms the results shown in table 6.2 (*i.e.*, the 50 buyers case). Indeed, the phenomenologies exhibited by the three methods MLML, MLMM, and MMMM across the scenarios are comparable to each other.

However, a further information emerges from the plots. In certain cases, a contribution to the error and the bias of the resulting estimates is due to the existence of alternative solutions of $(\xi_1^\star, \xi_2^\star, \theta^\star)$ that still leads to reproduce $\tilde{P}$ with a adequate level of approximation and is able to "explain" a specific scenario data set in terms of likelihood or moment-matching.

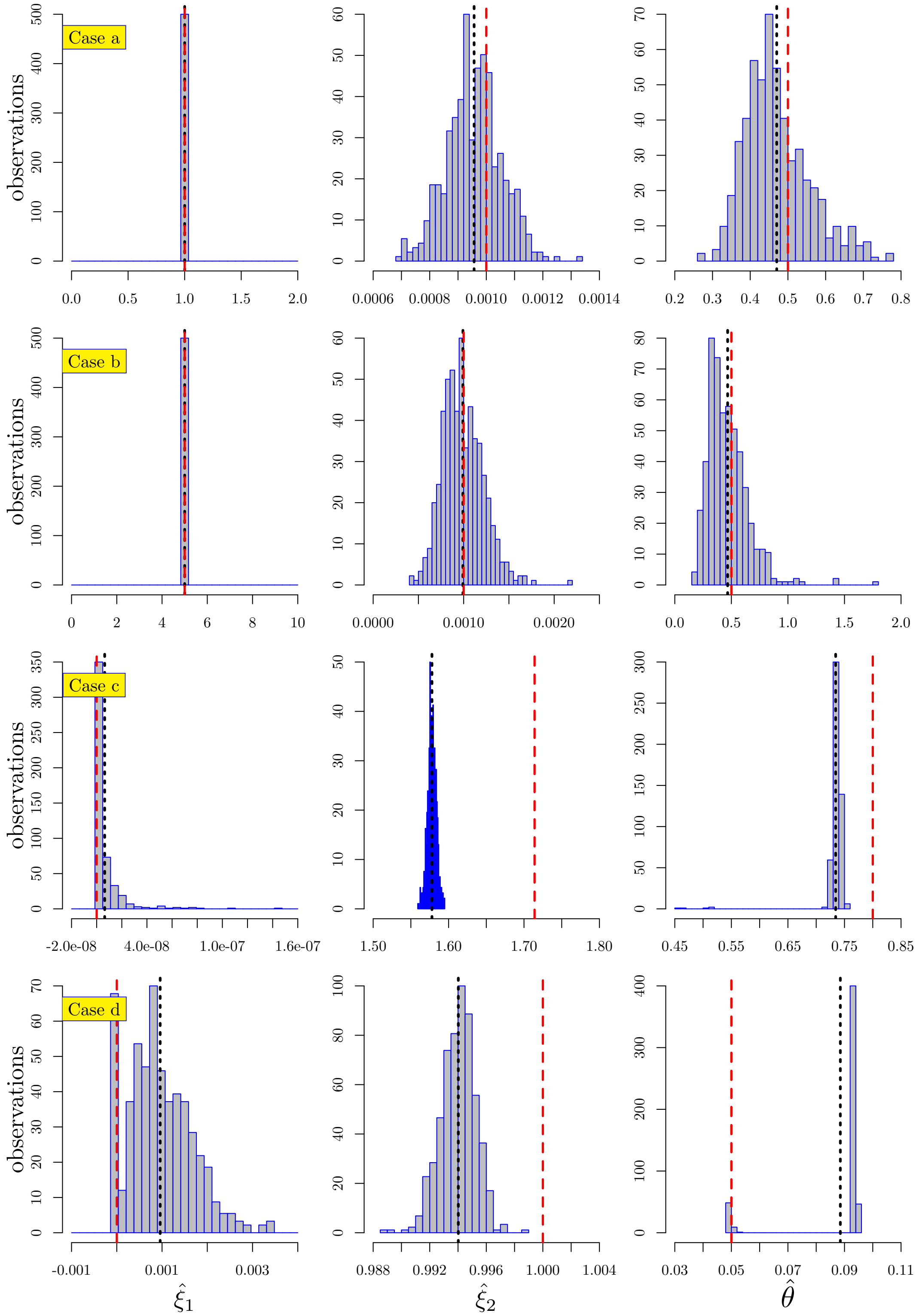

Figure B.1. Calibration outcomes for the behavioral types $a - d$ introduced in §6.2.1.2, considering the MLML technique over 500 scenarios. The red vertical line represents the true parameter's value, while the blue line is the average of the simulated estimations.

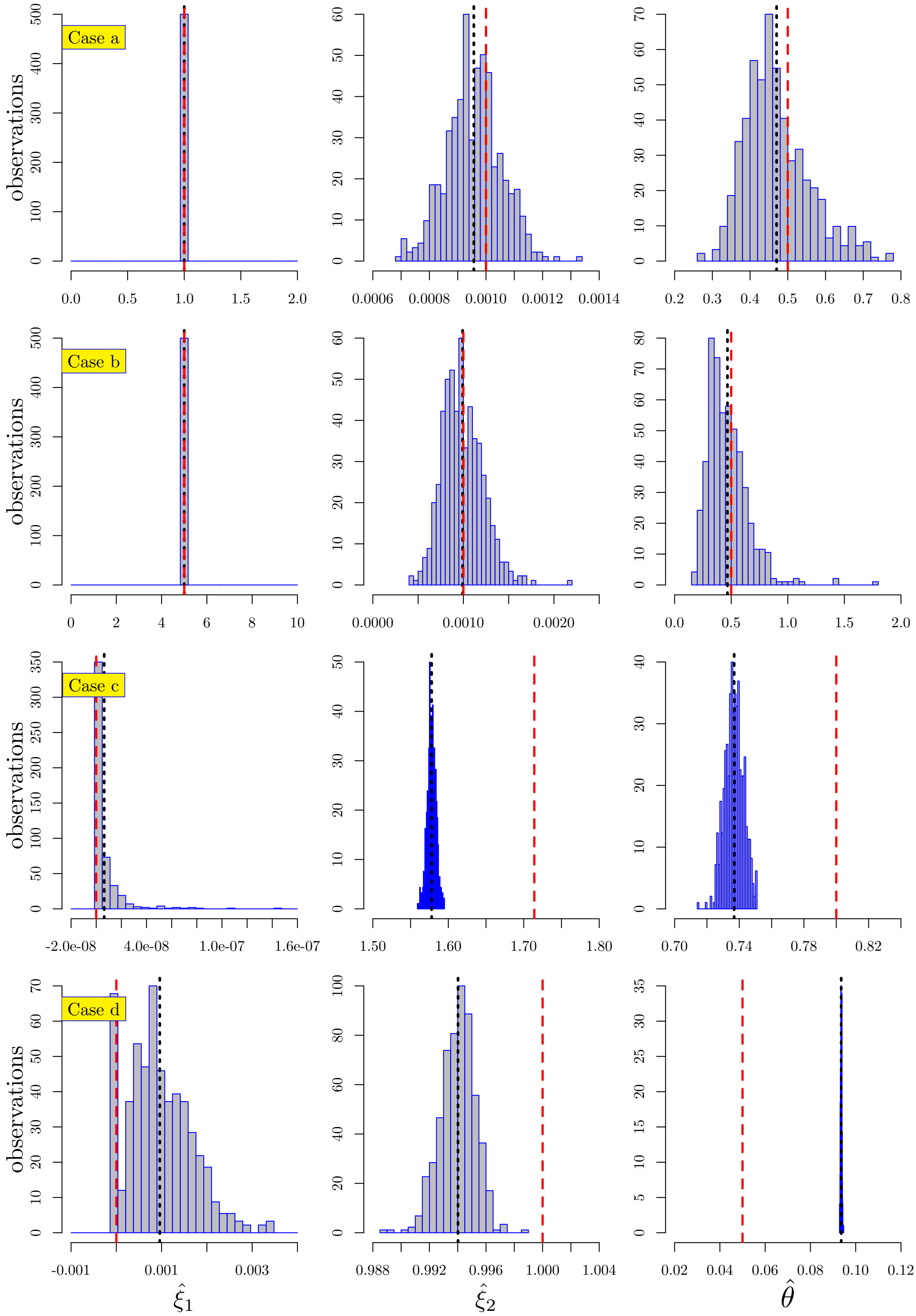

Figure B.2. Calibration outcomes for the behavioral types $a - d$ introduced in §6.2.1.2, considering the MLMM technique over 500 scenarios. The red vertical line represents the true parameter's value, while the blue line is the average of the simulated estimations.

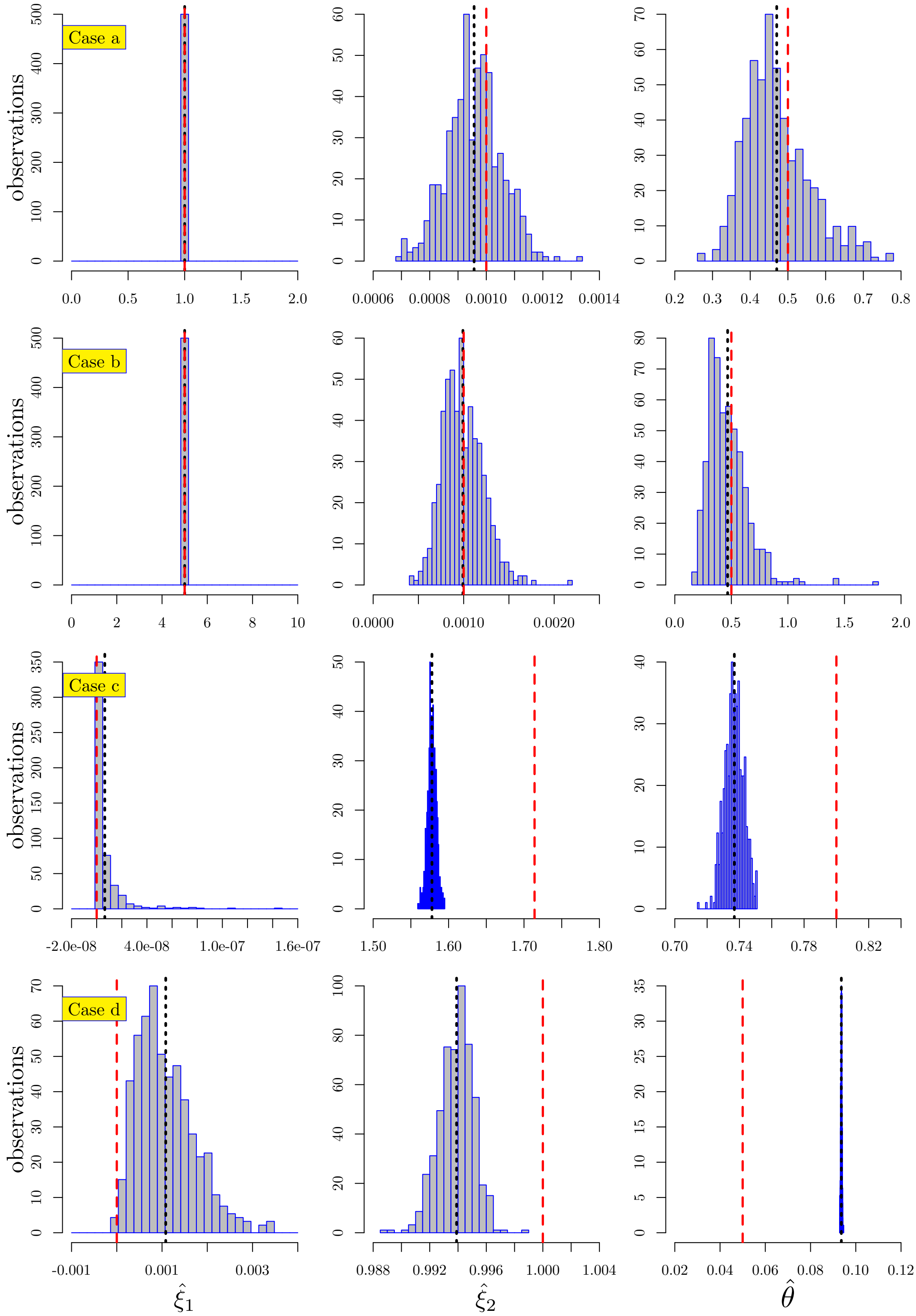

Figure B.3. Calibration outcomes for the behavioral types $a - d$ introduced in §6.2.1.2, considering the MMMM technique over 500 scenarios. The red vertical line represents the true parameter's value, while the blue line is the average of the simulated estimations.

Appendix C

Generic comments on simulations

Each of chapters 4, 5, and 6 addresses a specific research problem related to the central topic investigated in this work. Chapters 4 and 6 report both analytical and numerical results, the latter being needed to test the presented techniques and assess their fitness in increasingly realistic cases, where the assumptions made to obtain the analytical results are weakened. Further, the main results in chapter 5 are obtained only by numerical investigation.

Estimation and risk management techniques proposed in this work are tested against simulated data sets to check their capability of reconstructing and handling a priori known information. Hence, each numerical experiment performed in the previous chapters comprises two phases: data generation and processing. The generation phase has a relevant computational cost. Indeed, an actual database typically contains granular information: each elementary "object" (e.g., buyer, contractor, policy, bid, invoice, claim) is simulated and recorded separately, together with the data set needed to characterize it. A different granular data set must be produced for each Monte Carlo scenario, requiring a large amount of computational time. One can achieve a partial time saving if the scenarios are recorded instead of being simulated multiple times to be applied in testing and comparing different techniques. However, the drawbacks of this possibility (i.e., extensive memory usage and time required for writing/reading operations) limit its application.

Hence, the most effective optimization of computational time and memory can be achieved by choosing the proper number of scenarios N. The choice of N must consider the stability and significance of the results - which grow with N itself - and the computational cost. The optimal N value depends (mainly) on the simulated model, the measured quantity, and the maximum acceptable level of numerical error. Thus, it is necessary to perform an empirical analysis to choose a suboptimal N for each simulation.

In this work, the number of scenarios N per simulation has been selected as follows

i. given a fixed seed, N is increased until the results stabilize;

ii. following the single-seed stabilization, alternative seeds are tested with the same N;

iii. if the results obtained in step ii. remain stable up to the desired precision, N is confirmed. Otherwise, it is increased further.

Three to five alternative seeds have been considered for each simulation. The chosen N ranges from 10^5, when computing $A_{12}^{(m)}$ in §§4.4.2-4.4.3, to 10^3, when assessing the effectiveness of the proposed risk appetite framework in §§5.2.2-5.2.4. Namely, Monte Carlo simulations exhibit the slowest convergence when computing the covariance estimator and its error among the considered quantities. This fact is not surprising, as the estimation of covariance is generally affected by an error greater than the one that affects the estimation of a marginal probability.

The code used in this work has been developed entirely by the candidate in R language. Although R offers many packages dedicated to numerical simulation and statistical analysis purposes, the code developed in this work relies mainly on the basic functions to maximize the user's control over the computation. Exceptions to this choice are the packages used to handle the output and the package "pracma", which implements the incomplete Gamma function needed in chapter 6.

Simulations have been performed on three different computers to distribute the computational effort and as a mitigation factor against the reproducibility risks that may arise when a specific computer is the only source of a numerical result. All the computers employed in the numerical investigation were based on Intel 64-bit architecture.

Bibliography

[1] Morgan W.D., *History and Economics of Suretyship*, 12 Cornell L. Rev. 153 (1927)

[2] Sokolovska O., *Trade credit insurance and asymmetric information problem*, Scientific Annals of Economics and Business 64 (1), 2017, 123-137

[3] Russell J.S., *Surety Bonds for Construction Contracts*, American Society of Civil Engineers - ASCE Press, first edition (2000)

[4] L.W. King, 1915 translation of *Code of Hammurabi* - publicly available. We followed the edition of P.J.S. Pereira, licensed under the *Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.*

[5] "*Heard Act*", August 13, 1894. 28 Stat. 278, ch. 280

[6] Riestra A., *Credit Insurance in Europe. Impact, Measurement and Policy Recommendations.*, CEPS research report in finance and banking 31, Centre for European Policy Studies (2003)

[7] Jones P.M., *Trade credit insurance primer series on insurance.*, Global capital markets development department, The World Bank (2010)

[8] Cushman E.H., *Surety Bonds on Federal Construction Contracts: Current Decisions Reviewed*, 25 Fordham L. Rev. 241 (1956)

[9] "*Miller Act*", August 24, 1935. 49 stat. 793-794, ch. 642, Sec. 1-3

[10] The *Miller Act* (1935) is codified as amended in the *Code of Laws of the United States of America*, Title 40 *Public buildings, Property, and Works*

[11] People's Republic of China - Ministry of Commerce *Bids Law of the People's Republic of China*, order No.21 of the President of People's Republic of China on 30 August 1999.

[12] Federal Law No. 223-FZ *On Procurement of Goods, Works and Services by Certain Types of Legal Entities*, signed into law by the Russian president on 18 July 2011.

[13] Bezer, D. *The Inadequacy Of Surety Bid Bonds In Public Construction Contracting.* Public Contract Law Journal **2010**, 40(1), 87–146; http://www.jstor.org/stable/25755803

[14] Hassan, A.A.; Adnan, H. *The problems and abuse of performance bond in the construction industry.* IOP Conference Series: Earth and Environmental Science **2018**, 117, 012044; https://doi.org/10.1088/1755-1315/117/1/012044

[15] Helw, A.A.; Ezeldin, A.S. *Toward Better Applicability of Public Procurement Law: Delay Claims by the Contractor and Limit of Compensation under the Performance Guarantee.* Journal of Legal Affairs and Dispute Resolution in Engineering and Construction **2021**, 13(4), 04521023; https://doi.org/10.1061/(ASCE)LA.1943-4170.0000490

[16] R. Cantor, F. Packer *The credit rating industry*, Federal Reserve Bank of New York Quarterly Review, 1–26 (1994) ISSN 0147-6580.

[17] The fraction of worldwide credit insurance market held by ICISA members, as well as other information on the Association and its members, was available on the ICISA website, accessed 1 september 2020: https://www.icisa.org/.

[18] ICISA, *Statutes of the international Credit Insurance and Surety Association*, adopted by the 78th general meeting held on 10 June 2020.

[19] Information about Berne Union members, mission and activities was available on the Berne Union website, accessed 1 September 2020: https://www.berneunion.org/.

[20] EIOPA, *EIOPA's second set of advice to the European Commission on specific items in the Solvency II Delegated Regulation*, EIOPA-BoS-18/075, 28 February 2018.

[21] EIOPA, *European Insurance Overview 2019 - Solo Undertakings - Year End 2018*, Luxenburg Publications Office of the European Union (2019)

[22] *Surety industry tops $ 9 trillion in protection*, 16 april 2018, the the Surety & Fidelity Association of America website, accessed 1 September 2020. https://www.surety.org/news/396741/Surety-Industry-Tops-9-Trillion-in-Protection.htm

[23] Yesilyaprak, M.; Erkök, B. *A new instrument in Turkish financial market: surety bonds.* Finansal Araştırmalar ve Çalışmalar Dergisi **2021**, 13(25), 904–917; https://doi.org/10.14784/marufacd.976563

[24] Choudhry, R.M.; Iqbal, K. *Identification of Risk Management System in Construction Industry in Pakistan.* Journal of Management in Engineering **2013**, 29(1), 42–49; https://doi.org/10.1061/(ASCE)ME.1943-5479.0000122

[25] Liu, J.; Li, B.; Lin, B.; Nguyen, V. *Key issues and challenges of risk management and insurance in China's construction industry: An empirical study.* Industrial Management & Data Systems **2007**, 107(3), 382–396; https://doi.org/10.1108/02635570710734280

[26] Calveras, A.; Ganuza, J.J.; Hauk, E. *Wild bids. Gambling for resurrection in procurement contracts.* Journal of Regulatory Economics **2004**, 26(1), 41–68; https://doi.org/10.1023/B:REGE.0000028013.76488.44

[27] Wambach, A.; Engel, A.R. *Surety Bonds with Fair and Unfair Pricing.* The Geneva Risk and Insurance Review **2011**, 36(1), 36–50; https://link.springer.com/article/10.1057/grir.2010.8

[28] Sandstrom, A., *Handbook of Solvency for Actuaries and Risk Managers: Theory and Practice*, Chapman&Hall/CRC Finance, Taylor&Francis Group, 2011. ISBN-13 978-1439821305

[29] Olivieri, A.; Pitacco, E., *Introduction to Insurance Mathematics: Technical and Financial Features of Risk Transfers*, Springer: Berlin/Heidelberg, Germany, 2011. ISBN 978-3-642-16028-8

[30] IVASS Regulation No. 38 of 3 July 2018; Article 5, Paragraph 2, Letter E. https://www.ivass.it/normativa/nazionale/secondaria-ivass/regolamenti/2018/n38/

[31] Paulusch, J. *The Solvency II Standard Formula, Linear Geometry, and Diversification.* J. Risk Financial Manag. **2017**, 10, 11. https://doi.org/10.3390/jrfm10020011

[32] Baione, F.; De Angelis, P.; Granito, I. *On a capital allocation principle coherent with the Solvency 2 Standard Formula.* IVASS Conference on Insurance Research, 13 July 2017. https://www.ivass.it/pubblicazioni-e-statistiche/pubblicazioni/att-sem-conv/2017/conf-131407/

[33] Renault, B.Y.; Agumba, J.N. *Risk management in the construction industry: a new literature review.* MATEC Web of Conferences **2016**, 66, 00008; https://doi.org/10.1051/matecconf/20166600008

[34] EIOPA, *Supervisory statement on the Solvency II recognition of schemes based on reinsurance with regard to COVID-19 and credit insurance*, EIOPA-BoS-20/448, 20 July 2020.

[35] *Garanzia SACE in favore delle assicurazioni sui crediti commerciali*, D.Lgs. 19 May 2020, N. 34. Art 35.

[36] Directive 2009/138/EC of the European Parliament and of the Council of 25 november 2009, https://eur-lex.europa.eu/eli/dir/2009/138/oj

[37] Commission Delegated Regulation (EU) 2015/35 of 10 october 2014 supplementing Directive 2009/138/EC of the European Parliament and of the Council on the taking-up and pursuit of the business of Insurance and Reinsurance, https://eur-lex.europa.eu/eli/reg_del/2015/35/oj

[38] Commission Delegated Regulation (EU) 2019/981 of 8 march 2019 amending Delegated Regulation (EU) 2015/35 supplementing Directive 2009/138/EC of the European Parliament and of the Council on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II) https://eur-lex.europa.eu/eli/reg_del/2019/981/oj

[39] International Credit Insurance & Surety Association, *A guide to trade credit insurance*, Anthem Press, first edition (2015)

[40] International Credit Insurance & Surety Association, *ICISA catalogue of credit insurance terminology - english edition*, http://www.icisa.org/catalogue-of-terminology/1560/

[41] Circ. Ministero dell'Industria e del Commercio - Ispettorato delle Assicurazioni Private 145/1960 *Modalità di esercizio dei rami "credito e cauzione"*, extended by Circ. Ministero dell'Industria, del Commercio e dell'Artigianato 433/1979 *Aggiornamento dei criteri di gestione dei rami "credito" e "cauzione"*, confirmed by Circ. ISVAP 162/1991 *Disciplina dei rami "credito" e "cauzione"* and subsequent updates.

[42] OECD – Organization for Economic Co-operation and Developement, *The Export Credit Arrangements 1978-2018* (2018), https://www.oecd.org/tad/xcred/40594872.pdf

[43] *Trade credit insurance & surety: taking stock after the financial crisis*, Swiss Re - Economic Research & Consulting, october 2014

[44] Glen Boswall R., *Construction Bonds Guide*, 2010, Clark Wilson LLP, https://www.cwilson.com/app/uploads/2010/09/construction-bonding-guide.pdf

[45] *Codice dei contratti pubblici*, D. Lgs. April 18, 2016, N. 50. Gazzetta Ufficiale n. 91 April 19, 2016, s.o. n.10. Art. 93, subsection 1.

[46] *Codice dei contratti pubblici*, D. Lgs. April 18, 2016, N. 50. Gazzetta Ufficiale n. 91 April 19, 2016, s.o. n.10. Art. 103, subsections 1-3.

[47] *Codice dei contratti pubblici*, D. Lgs. April 18, 2016, N. 50. Gazzetta Ufficiale n. 91 April 19, 2016, s.o. n.10. Art. 103, subsections 5.

[48] *Codice civile* (Regio Decreto 16 marzo 1942, n. 262), Art. 1916

[49] *Adgelion K. Hall v. Windsor Savings Bank* Supreme Court of Vermont, february Term, 1923.

[50] Downes J., GoodmanJ.E., *Dictionary of Finance & Investment Terms*, Baron's Financial Guides, 2003.

[51] Euler Hermes BeLux, *What is the difference between a guaranty and a surety?* Euler Hermes website, accessed on 1 October 2021. https://www.eulerhermes.com/en_BE/news/latest-news/bndg-what-is-the-difference-between-a-guaranty-and-a-surety.html

[52] Credit Suisse, *Guarantees and surety bonds.* Credit Suisse website, accessed on 1 October 2021. https://www.credit-suisse.com/media/assets/private-banking/docs/ch/unternehmen/kmugrossunternehmen/garantien-und-burgschaften-en.pdf

[53] D'Aurizio L., Mattei P., Mosco V., *L'attività assicurativa nel comparto property e nel ramo r.c. generale (2013 – 2018)*, IVASS - Bollettino Statistico, year VII, N. 3, march 2020

[54] *TRI30605: Annual flow of adjusted non-performing loans/stock of performing loans at the previous year - by customer geographical area, sector and economic activity.*
Bank of Italy statistical database is avaliable at
https://infostat.bancaditalia.it/inquiry/.

[55] Merton R.C., *On the Pricing of Corporate Debt: The Risk Structure of Interest Rates*, Journal of Finance 29 (1974), pp 449–470. doi:10.1111/j.1540-6261.1974.tb03058.x

[56] https://www.moodysanalytics.com/about-us/history/kmv-history Available on the Moody's website as accessed on 30 october 2020.

[57] Nazeran P., Dwyer D., *Credit Risk Modeling of Public Firms: EDF9*, Moody's Analytics Quantitative Research Group, june 2015 https://www.moodysanalytics.com/-/media/whitepaper/2015/2012-28-06-public-edf-methodology.pdf Available on the Moody's website as accessed on 30 october 2020.

[58] E. Stanghellini, *Introduzione ai metodi statistici per il credit scoring*, Springer Verlag, first edition (2009)

[59] P. Gurný, M. Gurný, *Comparison of credit scoring models on probability of defaults estimation for US banks*, Prague Economic Papers **2013**, 22(2), 163-181

[60] P.M. Konrad, *The Calibration of Rating Models. Estimation of the Probability of Default based on Advanced Pattern Classification Methods*, Tectum Verlag Marburg, first edition (2012)

[61] E. Fankenstein, A. Boral, L.V. Carty, *RiskCalc for private companies: Moody's default model*, Moody's Investor Service Global Credit Research, may 2000

[62] Basel Committee on Banking Supervision (BSBC), *The Internal Ratings-Based Approach*, Bank for International Settlements, january 2001

[63] N. Nehrebecka, *Probability-of-default curve calibration and validation of internal rating systems*, 8^{th} IFC Conference on "Statistical implications of the new financial landscape", 8-9 sep. 2016

[64] T. Fawcett, *An introduction to ROC analysis*, Pattern Recognition Letters **2006**, 27, 861–874

[65] Tasche D., *The art of probability-of-default curve calibration*, Journal of Credit Risk **2013**, 9(4), 63–103

[66] Durović A., *Macroeconomic Approach to Point in Time Probability of Default Modeling – IFRS 9 Challenges*, Journal of Central Banking Theory and Practice **2019**, 1, 209-223

[67] IASB, (2014) , *International Financial Reporting Standard 9 Financial instruments*, International Accounting Standards Board.

[68] Wilson T., *Portfolio Credit Risk - Part I*, Risk Magazine **1997**, 10(9), 111-117.

[69] Wilson T., *Portfolio Credit Risk - Part II*, Risk Magazine **1997**, 10(10), 56-62.

[70] Jarrow R.A., Lando D., Turnbull S.M., *A Markov Model for the Term Structure of Credit Risk Spreads*, The Review of Financial Studies **1997**, 10(2), 481-523.

[71] Jarrow R.A., Turnbull S.M., *Pricing Derivatives on Financial Secureties Subject to Credit Risk*, Journal of Finance **1995**, 50, 53-86.

[72] Cox D.R., Miller H.D., *The Theory of Stochastic Processes*, Chapman and Hall, 1972.

[73] J.P. Morgan & Co., Inc. *Introduction to CreditMetrics*^{TM}, april 1997.

[74] RiskMetrics Group, Inc. *CreditMetrics*^{TM} *– Technical Document*, 2007.

[75] Vasicek O.A., *Probability of Loss on Loan Portfolio*, Document Number 999-0000-056, KMV Corporation, february 1987. (*Since 2011, the document has been disclosed among the Moody's Analytics white papers*).

[76] Vasicek O.A., *Limiting Loan Loss Probability Distribution*, Document Number 999-0000-046, KMV Corporation, august 1991. (*Since 2011, the document has been disclosed among the Moody's Analytics white papers*).

[77] Vasicek O.A., *The Distribution of Loan Portfolio Value*, Risk **2002**, 15(12), 160-162.

[78] Basel Committee on Banking Supervision (BSBC), *An Explanatory Note on the Basel II IRB Risk Weight Functions*, Bank for International Settlements, july 2005

[79] Basel Committee on Banking Supervision (BSBC), *Basel II: International Convergence of Capital Measurement and Capital Standards: a revised framework – comprehensive version*, Bank for International Settlements, june 2006

[80] Chatterjee S., *Modelling Credit Risk*, Bank of England, 2015.

[81] EIOPA, *NLCS 2020 Logfile ˘ Survey A* - Version 1.1 of 19 July 2021. Available online: `https://www.eiopa.europa.eu/document-library/feedback-request/non-life-underwriting-risk-comparative-study-internal-models_en` (accessed on 1 september 2021).

[82] ANIA (Associazione Nazionale Imprese Assicuratrici), *Appendice statistica alla relazione annuale 2019-2020*, august 2020

[83] Committee of European Insurance and Occupational Pensions Supervisors (CEIOPS), *Solvency II Calibration Paper*, CEIOPS-SEC-40-10, 15 april 2010.

[84] SACE BT, *Bilancio d'esercizio e consolidato 2019*, p. 23

[85] Passalacqua L., *A pricing model for credit insurance*, Giornale dell'Istituto Italiano degli attuari **2006**, Vol. LXIX

[86] Passalacqua L., *Measuring effects of excess-of-loss reinsurance on credit insurance risk capital*, Giornale dell'Istituto Italiano degli attuari **2007**, Vol. LXX, 81-102

[87] Jus M., *Credit Insurance*, Elsevier Academic Press , first edition (2013)

[88] Basel Committee on Banking Supervision (BSBC), *Basel III: A global regulatory framework for more resilient banks and banking systems.*, Bank for International Settlements, june 2011

[89] W.B. Hickman *Corporate Bonds: Quality and Investment Performance* Study by National Bureau of Economic Research, Princeton University Press, Princeton (1958)

[90] M. Greenwood, *A Report on the Natural Duration of Cancer*, Reports of Public Health and Related Subjects, Vol. 33, His Majesty's Stationery Office, London (1926).

[91] E.L. Kaplan, P. Meier *Nonparametric estimation from incomplete observations*, J. Amer. Statist. Assoc. **1958**, 53(282), 457–481

[92] S.J. Cutler, F. Ederer *Maximum utilization of the life table method in analyzing survival* J. Chronic. Dis. **1958**, 8(6), 669-712

[93] Wilmoth J.R. *et al.*, *Methods Protocol for the Human Mortality Database*, https://www.mortality.org/Public/Docs/MethodsProtocol.pdf, Last revised: 5 Oct. 2019 (Version 6)

[94] Armitage P., Berry G., Matthews J.N.S., *Statistical Methods in Medical Research*, Blackwell Science Ltd, fourth edition (2002)

[95] Hakulinen, T. (1982). Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics* **38** (4), 933-942.

[96] Hakulinen, T., Seppä, K. & Lambert, P.C. (2011). Choosing the relative survival method for cancer survival estimation. *Eur. J. Cancer* **46** (14), 2202-2210.

[97] Kondre, A.R. & Perme, M.P. (2013). Informative censoring in relative survival. *Stat. in Med.* **32** (27), 4791-4802.

[98] E.I. Altman, *Measuring Corporate Bond Mortality and Performance*, Journal of Finance **1989**, 44, 909-922

[99] International Credit Insurance & Surety Association, Comments Template on *Consultation Paper on EIOPA's second set of advice to the European Commission on specific items in the Solvency II Delegated Regulation* - feedback from ICISA to EIOPA, Jan. 2018.

[100] D.T. Hamilton, R. Cantor, *Measuring Corporate Default Rates*, Special Comments, Moody's Investors Service − Global Credit Research, Nov. 2006

[101] Klein J.P., Moeschberger M.L. *Survival Analysis: Techniques for Censored and Truncated Data*, Springer, second edition (2003)

[102] Vandendorpe A., Ho N., Vanduffel S., Van Dooren P., *On the parameterization of the CreditRisk$^+$ model for estimating credit portfolio risk*, Insurance: Mathematics and Economics **2008**, 42, 736-745

[103] Giacomelli, J.; Passalacqua, L. *Calibrating the CreditRisk$^+$ Model at Different Time Scales and in Presence of Temporal Autocorrelation.* Mathematics **2021**, 9(14), 1679; https://doi.org/10.3390/math9141679.

[104] Giacomelli, J.; Passalacqua, L. *Improved precision in calibrating CreditRisk$^+$ model for Credit Insurance applications.* In: Corazza M., Gilli M., Perna C., Pizzi C., Sibillo M. (eds) *Mathematical and Statistical Methods for Actuarial Sciences and Finance - eMAF 2020*, Springer, Cham., first edition (2021). https://doi.org/10.1007/978-3-030-78965-7_35.

[105] Giacomelli, J.; Passalacqua, L. *Unsustainability Risk of Bid Bonds in Public Tenders.* Mathematics **2021**, 9(19), 2385; https://doi.org/10.3390/math9192385.

[106] Giacomelli, J. *Parametric estimation of latent default frequency in credit insurance.* Journal of The Operational Research Society, **2022**. https://doi.org/10.1080/01605682.2022.2039567

[107] Klugman S.A., Panjer H.H., Willmot G.E., *Loss Models: From Data to Decisions*, Wiley, fourth edition (2012).

[108] Definition of "quarterly bad loan rates, calculated on the number of borrowers", *Banks and Financial Institutions: Credit Conditions and Risk by Sector and Geographical Area*, Methods and Sources: Methodological Notes, p. 18. Bank of Italy, 29 November 2019.

[109] A.M. Mathai, P.G. Moschopoulos *A form of multivariate Gamma distribution* Ann. Inst Statist. Math. **1992**, 44(1), 97-106

[110] Hogg R.V., Craig A.T., *Introduction to Mathematical Statistics*, Macmillian Publishing, fourth edition (1978)

[111] Istat, (2021). ATECO 2007 - Classification of Economic Activities. Available from:
https://www.istat.it/it/archivio/17888

[112] Bank of Italy statistical database is avaliable at https://infostat.bancaditalia.it/inquiry/.
The bad loan rate series is labelled as *TRI30496_35120163*. The count of performing borrowers at the initial period series is labelled as *TRI30496_351122141*.

[113] Bank of Italy, (2020c). The data used to estimate the default probability per economic sector are openly available from the Infostat Bank of Italy statistical database at infostat.bancaditalia.it/inquiry/. The "*Quarterly flow of new adjusted bad loans by costumer economic activity: number of borrowers*" series is labeled as *TRI30529_351121441*. The "*Stock performing loans at the initial period: number of borrowers*" series is labeled as *TRI30529_351122141*.

[114] Crouhy M., Galai D, Mark R., *A comparative analysis of current credit risk models*, Journal of Banking & Finance **2000**, 24, 59-117.

[115] Murphy D., *Unravelling the Credit Crunch*, CRC Press, 2009.

[116] Schönbucher P. J., *Credit Derivatives Pricing Models: Model, Pricing and Implementation*, Wiley, 2003.

[117] Fréchet M., *Sur les tableaux de corrélation dont les marges sont donnés*, Annales de l'Université de Lyon, Science **1951**, 4, 13-84.

[118] Sklar A., *Fonctions de Répartition à n Dimensions et Leurs Marges*, Institut Statistique de l'Université de Paris, Paris, **1951**, 8, 229-231.

[119] Joe H., *Multivariate Models and Dependence Concepts*, Chapman and Hall/CRC, 1997.

[120] Nelsen R. B., *Introduction to Copulas* (first edition), Springer, 1999.

[121] Li D. X., *On Default Correlation: A Copula Function Approach*, Journal of Fixed Income, **2000**, 9(4), 43-54.

[122] McNeil A., Frey R., Embrechts P., *Quantitative Risk Management - Concepts, Techniques and Tools*, Princeton University Press, 2015.

[123] Kotz S., Adams J. W., *Distribution of Sum of Identically Distributed Exponentially Correlated Gamma-Variables*, The Annals of Mathematical Statistics, **1964**, 35 (1), 277-283.

[124] Florent C., Borgnat P., Tourneret J., Abry P., *Parameter estimation for sums of correlated gamma random variables. Application to anomaly detection in Internet Traffic*, IEEE Int. Conf. on Acoust., Speech and Signal Proc. ICASSP-08, IEEE **2008**, ensl-00290724.

[125] Feng Y., Wen M., Zhang J., Ji F., Ning G. *Sum of arbitrarily correlated Gamma random variables with unequal parameters and its application in wireless communications*, IEEE 2016 International Conference on Computing, Networking and Communications (ICNC), **2016**, DOI 10.1109/ICNC.2016.7440693

[126] *Non-performing loans (NPLs) in Italy's banking system*, **2017**. https://www.bancaditalia.it/media/views/2017/npl/

[127] Higham N., *Computing the nearest correlation matrix - a problem from finance*, IMA Journal of Numerical Analysis, **2002**, 22, 329-343.

[128] Mai J-F., Scherer M., *Simulating Copulas: Stochastic Models, Sampling Algorithms, and Applications* (second edition), World Scientific Publishing Company, 2017.

[129] Cox D.R., *Some Statistical Methods Connected with Series of Events*, Journal of the Royal Statistical Society, **1955**, 17(2), 129–164.

[130] Duffie D., Singleton K.J., *Modeling Term Structures of Defaultable Bonds*, The Review of Financial Studies, **1999**, 2(4), 687-720

[131] Duffie D., Singleton K.J., *Credit Risk. Pricing, measurement, and management*, Princeton University Press, 2003

[132] Spadaro, A., *Calibrazione di un modello alla Duffie & Singleton mediante l'applicazione del Particle Filter*, Technical Report 2-2019, Sapienza University of Rome.

[133] Wilde T., *CreditRisk$^+$*, in Encyclopedia of Quantitative Finance, John Wiley & Sons, 2010.

[134] Credit Suisse First Boston, *CreditRisk$^+$, A Credit Risk Management Framework*, 1998.

[135] Gundlach M., Lehrbass F. (eds.), *CreditRisk$^+$ in the Banking Industry*, Springer, 2004.

[136] Glasserman P., Li J., *Importance Sampling for Portfolio Credit Risk*, Management Science, **2005**, 51 (11), 1643-1656.

[137] Luque González, A.; Coronado Martín, J.Á.; Vaca-Tapia, A.C.; Rivas, F. *How Sustainability Is Defined: An Analysis of 100 Theoretical Approximations.* Mathematics **2021**, *9*, 1308. DOI: https://doi.org/10.3390/math9111308

[138] Directive 2014/24/EU of the European Parliament and of the Council of 26 February 2014 on public procurement and repealing Directive 2004/18/EC.

[139] Krishna, V., *Auction Theory*, San Diego, USA: Academic Press, 2002. ISBN 978-0-12-426297-3

[140] Milgrom, P., *Putting Auction Theory to Work*, Cambridge, United Kingdom: Cambridge University Press, 2004. ISBN 978-0-521-55184-7

[141] Fisher, R.A.; Tippett, L.H.C. (1928), "Limiting forms of the frequency distribution of the largest and smallest member of a sample", Proc. Camb. Phil. Soc., 24 (2): 180–190, https://doi.org/10.1017/s0305004100015681.

[142] Gnedenko, B.V. (1943), "Sur la distribution limite du terme maximum d'une serie aleatoire", Annals of Mathematics, 44 (3): 423–453, https://doi.org/10.2307/1968974.

[143] Suzuki, A. *Investigating Pure Bundling in Japan's Electricity Procurement Auctions*. Mathematics **2021**, 9, 1622. https://doi.org/10.3390/math9141622