Three Essays in Finance and Actuarial Science

Regis Luca

Università di Torino

THREE ESSAYS IN FINANCE AND ACTUARIAL SCIENCE

Luca Regis

Scuola di Dottorato in Economia "V.Pareto",
indirizzo in Statistica e Matematica Applicata

Università di Torino
Dipartimento di Statistica e Matematica Applicata
"Diego De Castro"

THREE ESSAYS IN FINANCE AND ACTUARIAL SCIENCE

Luca Regis

Scuola di Dottorato in Economia "V.Pareto",
indirizzo Statistica e Matematica Applicata
Ciclo XXII

Supervisore: Prof. ssa Elisa Luciano
Coordinatore: Prof. Paolo Ghirardato

Acknowledgements

The first and most important thank goes to my supervisor Elisa Luciano, for her constant and invaluable support throughout all my Ph.D. years, fruitful discussions and suggestions. I am deeply indebted to Fabrizio Restione, Alberto Fasano and the Risk Management Department at Fondiaria-Sai s.p.A. for providing me with the data I use in the first Part of my dissertation. I would also like to thank Carmelo Genovese for helpful discussions on the first Chapter.
The work benefited of comments from participants to the IV MAF Conference (Ravello, 7-9 April 2010), the IRMC 2010 (Firenze, 4-6 June 2010) and the XII Workshop on Quantitative Finance (Padova, 27-28 January 2011). I kindly thank the Fondazione Collegio Carlo Alberto and the Dipartimento di Statistica e Matematica Applicata "D. De Castro" and their staffs for facilities and technical assistance. Finally, a sincere thank goes to my family and Marta, for supporting me during these years.

Contents

Part I

Bayesian models for stochastic claims reserving

The first part of my Ph.D. dissertation develops a Bayesian stochastic model for computing the reserves of a non-life insurance company.
The first chapter is the product of my research experience as an intern at the Risk Management Department of Fondiaria-Sai S.p.A.. I present a short review of the deterministic and stochastic claims reserving methods currently applied in practice and I develop a (standard) Over-Dispersed Poisson (ODP) Bayesian model for the estimation of the Outstanding Loss Liabilities (OLLs) of a line of business (LoB). I present the model, I illustrate the theoretical foundations of the MCMC (Markov Chain Monte Carlo) method and the Metropolis-Hastings algorithm used in order to generate the non-standard posterior distributions. I apply the model to the Motor Third Party Liability LoB of Fondiaria-Sai S.p.A..
The Risk Management Department of the company was already developing a Bayesian model for stochastic claims reserving when I began my intern. My contribution to the project consisted in a re-parametrization of the model, that allowed to adopt unrestricted distributional assumptions for the priors. The practical implementation of this model, which I describe in Chapter 1, was my own work. I am grateful to the Head of the Technical Risk Management Department Fabrizio Restione and to Carmelo Genovese for giving me the opportunity to work on such an interesting subject and for their helpful suggestions.
This chapter is also an introduction to the next one, in which I explore the problem of computing the prudential reserve level of a multi-line non-life insurance company. In the second chapter, then, I present a full Bayesian model for assessing the reserve requirement of multiline Non-Life insurance companies. The model combines the Bayesian approach for the estimation of marginal distribution for the single Lines of Business and a Bayesian copula procedure for their aggregation. First, I consider standard copula aggregation for different copula choices. Second, I present the Bayesian copula technique. Up to my knowledge, this approach is totally new to stochastic claims reserving. The model allows to "mix" own-assessments of dependence between LoBs at a company level and market wide estimates. I present an application to an Italian multi-line insurance company and compare the results obtained aggregating using standard copulas and a Bayesian Gaussian copula. I am again grateful to Carmelo Genovese who suggested the use of a time series of loss ratios to estimate standard copula parameters.

Chapter 1

A Bayesian stochastic reserving model

1.1 Introduction

Reserve risk, together with premium risk, constitutes the main source of potential losses for an insurance non-life company. The Italian insurance companies' monitoring agency, ISVAP, stated in 2006 that, based on a research carried out through the most important actors in the domestic markets, actuaries did use mainly deterministic methods for computing claims reserves. The European (and also the Italian) regulator has strongly encouraged the use of more sophisticated stochastic models for claims reserving and a deeper interaction between actuarial and risk management departments within insurance companies. The aim of my research experience in the risk management department of Fondiaria-Sai Spa, one of the largest Italian insurance companies, consisted in developing an internal model (in the light of the Solvency II proposal directive) with the purpose of this "Use Test" principle of enhancing the integration between different departments of the company. Bayesian models for claims reserving have in this sense been pointed out as a tool capable of answering the need risk managers have of deriving a full distribution of the outstanding loss liabilities faced by the company, while permitting the inclusion of actuarial experience (expert judgement) in the calculation of technical reserves. The inclusion of experts' knowledge about qualitative factors that can influence the estimates of ultimate costs and the development pattern is of utmost importance. Indeed, standard deterministic and stochastic models usually have as inputs the triangle of payments (and in some cases the triangle of the number of claims paid) only. Hence, they can not account for important factors such as assessor's speed and accuracy, expected future changes in legislation, etc. which can indeed be analyzed by experts (the actuarial department) and included in the model through Bayesian techniques. We review the deterministic and stochastic models which are more frequently used in practice in sections 1.2 and 1.3 respectively. In section 1.4 we present bayesian models for claims reserving. Since these models have analytical solutions for very particular distributional assumptions, Markov Chain Monte Carlo simulation methods are often used. Section 1.4 offers a quick theoretical treatment of the MCMC simulation methods. Section 1.5 presents the ODP Bayesian model I developed during my internship at Fondiaria-Sai s.p.A., while the following section 1.6 shows an application of the model to the MTPL LoB of the company. Section 1.7 concludes and hints at possible further developments of the model. Basic references for this Chapter are Merz and Wuthrich (2008), England and Verrall (2002), Cowles and Carlin (1996), Gilks et al. (1996) and Robert (2007).

1.2 Deterministic models for claims reserving

Data about past claim payments are usually collected in triangles. It is common practice to organize them by accident (or origin) year (a.y.)- i.e. the year in which the claim originated - on the different rows and development year (d.y.) - i.e. the year in which the payment effectively takes place - on the columns, as in the following example:

	Development Year (d.y.)			
Accident Year (a.y.)	*a.y.+0*	*a.y.+1*	*a.y.+2*	*a.y.+3*
2005	100	90	80	65
2006	130	115	85	?
2007	125	98	?	?
2008	160	?	?	?

In the "upper" part of the triangle - above the main diagonal - lies the set of payments already reported. In the "lower" triangle - below the diagonal - we find data relative to the future. The question marks in the example symbolize that the payments are yet to be realized and hence their amount is still unknown. Triangles can be constructed to contain incremental (reporting in each cell the payment relative to the combination a.y/d.y.) or cumulative data (in each cell is reported the sum of the payments relative to the a.y. of the corresponding row up to the d.y. identified by the column). In order to prudentially evaluate the reserves the company has to keep in order to face future payments already originated, one has to estimate the unrealized figures of the lower triangle. The main assumption of almost every deterministic and stochastic models that the cumulative claims development pattern is the same across the different accident years. Another implicit and rather strong assumption of the model is that, since reserves are computed taking the last accident year as the last development year for the first a.y., triangles must be long enough to ensure that what is intentionally (without taking into account late payments) left out of it, i.e. claims relative to development years not considered are irrelevant.

Deterministic approaches, based on purely statistical techniques, are used in most actuarial departments of insurance companies. The reason is that they are very intuitive, ready-to-use and flexible. Their main downsides are that - at least in their original formulations - they only provide point estimates of claims reserves rather then distributions of OLLs and that they usually rely on a very limited set of observations, and are thus subject to high estimation errors, especially when they are applied to small LoBs. This is due to the use of aggregate data, that entails great variability in the triangles, in particular when they are obtained from a small number of individual claims.

Two are the most important deterministic methods used in the actuarial practice: the Fisher-Lange method and the Chain Ladder. Another technique, the Bornhuetter-Ferguson one, has gained increased importance since its introduction.

In the next sections, we briefly review these methods.

1.2.1 Fisher Lange method

The way practitioners apply the Fisher Lange method, which is probably the most used in the Italian market, but is almost unknown in foreign countries differs broadly from one user to another. The basic feature of this method is that it is based on the principle of the mean cost and thus requires data about the claim payments' amount and the number of claims. Different methods are applied to project these data and obtain estimates of future number of claims and mean costs for each combination of accident and development year. In the evaluation mean settlement delay and development pattern features such as late settlements, reopened claims and claims closed without settlement rates are taken into account.

1.2.2 Chain Ladder method

Chain Ladder is the most popular technique for claims reserving in the actuarial world. It has also attracted the scholars' attentions in the last fifteen years, being the first method to be extended stochastically Mack (1993). The technique uses cumulative data on claim payments to obtain so called "development factors" or " link ratios". Define by X_{ij} the incremental payment in d.y. j for claims originated in year i and by C_{ij} the corresponding cumulative claim ($C_{ij} = \sum_{k=1}^{j} X_{ik}$). The Chain Ladder estimates *development factors* as

$$\widehat{\lambda}_j = \frac{\sum_{i=1}^{n-j+1} C_{i,j}}{\sum_{i=1}^{n-j+1} C_{i,j-1}} \tag{1.1}$$

These quantities are directly linked to the development pattern:

$$\widehat{\beta}_j^{CL} = \prod_{k=j}^{J-1} \frac{1}{\widehat{\lambda}_k} \text{ , } k = 1...J-1 \tag{1.2}$$

$\widehat{\beta}_j^{CL}$ represents the proportion of the ultimate cost to be sustained - "developed" - in d.y. j. Forecasts of the lower triangle of cumulative claims,$\widehat{C}_{i,j}$, are obtained from these factors:

$$\widehat{C}_{i,n-i+2} = C_{i,n-i+1}\widehat{\lambda}_{n-i+2} \tag{1.3}$$
$$\widehat{C}_{i,k} = \widehat{C}_{i,k-1}\widehat{\lambda}_k \quad k = n-i+3, n-i+4, ..., n \tag{1.4}$$

Obviously, once can obtain the triangle of incremental claims simply differencing the cumulative claims' one. Such an easy model can be extended, assuming that the C_{ij} 's are stochastic, in order to obtain a statistical model and a prediction error. Many distributional assumptions have been proposed to describe claims' amount behavior, the over-dispersed Poisson being the most successful in literature[1].

1.2.3 Towards a Bayesian perspective: the Bornhuetter-Ferguson method

The first actuarial technique developed to provide a model-based interaction between deterministic data based methods and expert judgement is due to Bornhuetter (1972). The idea consists in using an external estimate of the

[1]Mack's (1993) DFCL model was the first stochastic extension of the chain ladder model. It assumed that payments of different a.y. are independent and proposed to use $\lambda_j C_{i,j-1}$ as the conditional mean and $\sigma_j^2 C_{i,j-1}$ as the variance of the distribution of cumulative payments. Other possible distributional choices used in literature include the negative binomial distribution for cumulative claims, the log-normal for incremental claims, gamma distributions. All models, for estimation purposes, are usually reparametrized to obtain linear forms that belong to the GLM class.

ultimate claims together with the development factors derived through the chain ladder method to estimate outstanding claims. Denoting by $x_i^{(CL)}$ the ultimate claim for accident year i obtained using the chain ladder technique, the estimated OLLs using the chain ladder method can be written as

$$\sum_{i=1}^{n} x_i^{(CL)} \frac{1}{\lambda_{n-i+2}...\lambda_n} \left(\lambda_{n-i+2}...\lambda_n - 1\right) \tag{1.5}$$

Bornhuetter and Ferguson suggested to use an external evaluation of expected ultimate claims, $x^{(BF)}$. Following the procedure we described above, reserves can then be obtained as

$$\sum_{i=1}^{n} x_i^{(BF)} \frac{1}{\lambda_{n-i+2}...\lambda_n} \left(\lambda_{n-i+2}...\lambda_n - 1\right) \tag{1.6}$$

Bayesian models, that we detail in section (1.4), are based on this principle of using information derived by expert judgement in order to obtain reserves estimates. This information, however, is included in the model as a parameter subject to uncertainty.

1.3 Stochastic models for claims reserving

While for the purpose of computing statutory reserves one needs to indicate just a point estimate of reserves, risk assessments can be done only in the presence of a full distribution of the underlying risk factor which, in claims reserving, is constituted by the OLLs. Stochastic methods are then necessary to derive this distribution.

In the QIS 4(Quantitative Impact Studies) that CEIOPS proposed to insurance companies in 2008, the following definitions were given:

- Best estimate: is the probability-weighted average of future cash-flows;
- Risk Margin: "should be such as to ensure that the value of technical provisions is equivalent to the amount that (re)insurance undertakings would be expected to require to take over and meet the (re)insurance obligations".
- Required reserve: is the sum of the Best Estimate and the prudential risk margin
- Risk capital: it is given by the difference between a worst case scenario value (computed as the 99.5 percentile of OLLs distribution) and the required reserve.

A high degree of freedom in the choice of the methodology used for computing non-life technical reserves is left to the companies through the possibility of adopting internal models under the Solvency II framework, given that, in the case of reserve risk, they match the current "actuarial best practice". This section reviews the most important stochastic techniques currently used in the claims' reserving practice.

1.3.1 ODP model and GLM theory

One of the most widely used models in stochastic claims reserving is the Over-dispersed Poisson model for incremental claims.

The model states that:

$$\mathbb{E}\left[X_{ij}\right] = m_{ij} = x_i y_j \qquad Var[X_{ij}] = \phi x_i y_j, \tag{1.7}$$

$$\text{with} \sum_{k=1}^{n} y_k = 1,\ y_k, x_k > 0\ \forall k,\ \phi > 0 \tag{1.8}$$

where x_i is the expected ultimate claim for accident year i and y_j is the proportion of ultimate claims to emerge in each development year and ϕ is the over-dispersion parameter. This multiplicative structure, while extremely meaningful from an economic point of view, brings the problem of being a non-linear form, difficult to handle for estimation purposes. It is then often simpler to re-parametrize the model in order to put it in linear form. Using the language of Generalized Linear Models, we defining a *log link* function:

$$\log(m_{ij}) = c + \alpha_i + \beta_j \tag{1.9}$$

Dealing with a GLM model, parameters are easy to estimate with any standard econometric software, but parameter transformations are then necessary in order to get back to the economic quantities of interest. Constraints have to be applied to the sets of parameters in order to specify the model. The most used ones are the corner constraints ($\alpha_1 = \beta_1 = 0$). Once parameter estimates are obtained, the lower triangle is predicted plugging parameters' estimations into equation (1.9) and exponentiating. A possible shortcoming of the model consists in the fact that column sums of incremental claims can not be negative. This is almost always true, but there are cases, for example in the motor hull insurance, where insurance companies at certain late development years receive more money through subrogation and interest deductions than it spends in payments.

1.3.2 Obtaining full predictive distributions by simulation

Bootstrap methods

Bootstrapping allows to obtain and use information form a single sample of data. Assume to have a triangle of past claim payments. The bootstrap technique is the simplest way of simulating the full predictive distribution since it entails using past information to creating a set of pseudo-data. Starting from the triangle of past incremental or cumulative payments, one obtains a triangle of adjusted residuals and resamples it without replacement to obtain pseudo-triangles created from the initial distribution of past settlements.

The procedure to apply bootstrap for claims reserving estimation is the following (England and Verrall (2002)):

1. Obtain standard chain-ladder development factors.

2. Obtain fitted values by backwards recursion using chain ladder estimates.

3. Obtain incremental fitted values $\widehat{m}_{ij}$ for the past triangle obtained in step 2 by differencing.

4. Calculate the Pearson residuals:

$$r_{ij=}\frac{X_{ij}-\widehat{m}_{ij}}{\sqrt{\widehat{m}_{ij}}} \tag{1.10}$$

5. Adjust Pearson's residuals by the degrees of freedom of the model.

6. Iteratively create the set of pseudo-data, resampling without replacement the adjusted Pearson's residuals and obtain incremental claims as

$$X_{ij}=r_{ij}\sqrt{\widehat{m}_{ij}}+\widehat{m}_{ij} \tag{1.11}$$

7. Create the set of pseudo-cumulative data, and use chain ladder to fill the lower triangle, obtaining for each cell the mean of the future payment to use when simulating the payment from the process distribution.

Figure 1.1 represents the (rescaled) OLL distribution of the MTPL LoB of Fondiaria-Sai Spa obtained by Bootstrapping from an ODP model.

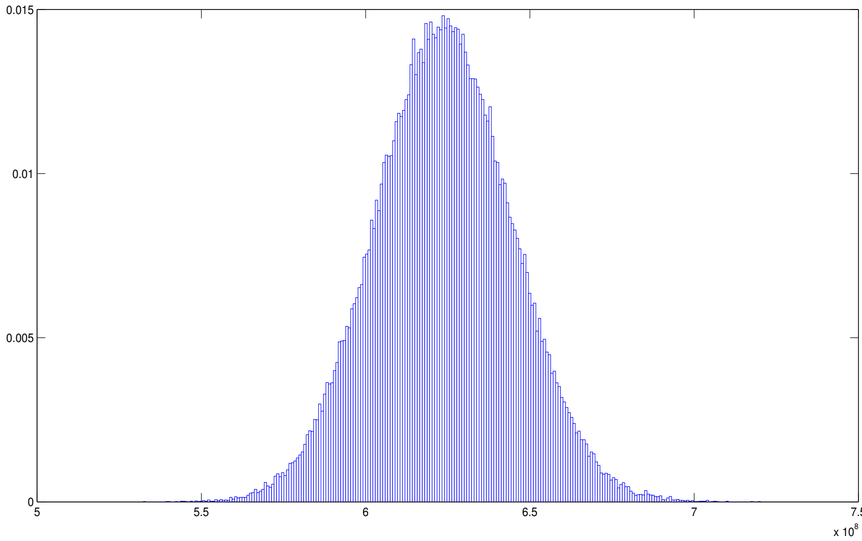

Figure 1.1: OLL distribution obtained by bootstrapping. The underlying model is an ODP for incremental claims.

Simulation from parameters

Another possibility consists in simulating the predictive distribution of outstanding claims from the joint distribution of the parameters of the model. Anyway, either the definition and the estimation of a joint distribution for the parameters of the distribution of claim settlements is usually a difficult (or at least imprecise) task. The usual procedure consists in making a distributional assumption on the linear predictor of a GLM model with Normal errors, estimating the parameters and their variance/covariance matrix and then sampling. The parameters are usually drawn from a multivariate normal distribution and lead to a full distribution of each future payment in the run-off triangle. It is then easy to sum them up to form the required reserve estimation. However, when we specify a ODP model for incremental claims, this procedure can not be applied, since the GLM model we are dealing with does not have Normal errors. Moreover, a Normal approximation is not a good choice for any triangle.

1.4 Bayesian models and Markov Chain Monte Carlo methods

The Bayesian paradigm can partly solve the identification problem we addressed in the previous section. In the Bayesian world, in fact, parameters are treated in the same way as the observable quantities of interest. It is thus necessary to make distributional assumptions about the prior (at least about the marginals, if they are supposed independent) distributions of parameters in order to obtain a joint *posterior* distribution. Markov Chain Monte Carlo is the simulation tool that allows us to treat analytically intractable joint distributions and permits us to merge prior assumptions and model uncertainty in a single framework. We denote by D the set of observed data and by θ the set of model parameters and missing data. Their joint distribution is given by

$$P(D,\theta) = P(D|\theta)P(\theta) \tag{1.12}$$

Applying Bayes' theorem it is easy to obtain the conditional distribution of θ given D, called *posterior distribution*:

$$P(\theta|D) = \frac{P(\theta)P(D|\theta)}{\int P(\theta)P(D|\theta)d\theta} \tag{1.13}$$

All the relevant quantities, such as moments and quantiles of the posterior distribution can be expressed in terms of posterior expectations of functions of θ.

For example,

$$\mathbb{E}[f(\theta)|D] = \frac{\int f(\theta)P(\theta)P(D|\theta)d\theta}{\int P(\theta)P(D|\theta)d\theta} \tag{1.14}$$

The task of evaluating these expectations in high dimensions is hardly an analytically tractable object.

Markov Chain Monte Carlo allows us to evaluate expectations of the form

$$\mathbb{E}[f(X)] = \frac{\int f(x)\pi(x)dx}{\int \pi(x)dx} \tag{1.15}$$

Since $\pi(\cdot)$ is usually a non-standard distribution, drawing samples of $\{X_t\}$ independently from π is not possible. Here comes the Markov Chain approach.

A Markov Chain is a sequence of variables $\{X_0, X_1, X_2, ...\}$, whose evolution is defined by a transition kernel $P(\cdot|\cdot)$. X_{t+1} is thus sampled at each t from $P(X_{t+1}|X_t)$ that depends on the current state of the chain only.

The chain, under some regularity condition, will converge to a *stationary distribution*, which is independent of the initial state X_0 or of time t and, moreover, it is unique.

Following Roberts (1996), we define $\tau_{ii} = \min\{t > 0 \ : X_t = i | X_0 = i\}$. The following definitions apply:

Definition 1.4.1 ***Irreducible***, *a chain such that for all i, j $\exists$ $t > 0$ s.t. $P_{ij}(t) > 0$.*

Definition 1.4.2 ***Recurrent***, *an irreducible chain such that $P[\tau_{ii} < \infty] = 1$ for some (and hence for all i)[2].*

Definition 1.4.3 ***Positive recurrent*** *an irreducible chain X for which $E[\tau_{ii}] < \infty$ for some (and hence for all i), or, equivalently, if there exists a stationary probability distribution, i.e. there exists $\pi(\cdot)$ such that*

$$\sum_i \pi(i) P_{ij}(t) = \pi(j) \tag{1.16}$$

for all j and $t \geq 0$.

Definition 1.4.4 ***Aperiodic***, *an irreducible chain such that for some (and hence for all i), the greatest common divider $\{t > 0 : P_{ii}(t) > 0\} = 1$.*

The following theorem holds true:

Theorem 1.4.5 *If a Markov chain X is positive recurrent and aperiodic, then its stationary distribution $\pi(\cdot)$ is the unique probability distribution satisfying (1.16).*

Then, we say that X is ergodic and the following consequences hold:

1. *$P_{ij}(t) \to \pi(j)$ as $t \to \infty$ for all i, j.*
2. *(Ergodic theorem) If $E_\pi[|f(X)|] < \infty$, then $P[\overline{f}_N \underrightarrow{N \to \infty} E_\pi[f(X)] = 1$,*

   where N is the size of the sample, $\overline{f}_N = \dfrac{1}{N} \sum_{i=1}^N f(X_i)$ is the sample mean, $E_\pi[f(X)] = \sum_i f(i)\pi(i)$ is the expectation of $f(X)$ with respect to $\pi(\cdot)$.

---

[2]Equivalently, $\sum_i P_{ij}(t) = \infty$ for all i, j.

The second part of this theorem is fundamental for applications, since it regards convergence of the quantities we want to evaluate. Anyway, since convergence is not ruled by a central limit theorem, we have in general no direct clue about how fast this convergence is reached [3]. In practical applications it is necessary to determine whether or not this convergence is achieved and how many iterations are necessary in order to 'forget' the effect of the initial state and reach the stationary distribution (those first n iterations are named *burn-in* and are discarded from the chain and from the calculation of the posterior mean). To these extents, some convergence diagnostic methods based on the output of the MCMC simulations are available.
There are some simple ways of monitoring convergence. A first and really naive one consists in looking at the means of the parameters in the chain. Once a mean is calculated taking a sufficient number of iterations, taking into account more iterations should not modify its estimate substantially. The Gelman-Rubin diagnostic (Gelman and Rubin (1992)) is again a simple but more robust approach - even though subject to some criticism. It consists in running several (n) parallel chains starting from points over dispersed with respect to the stationary distribution and monitoring some scalar quantities to assess convergence:

$$
\begin{aligned}
GR &= \frac{\widehat{V}(\theta)}{W}, && (1.17)\\
\widehat{V}(\theta) &= \left(1 - \frac{1}{n}\right) W + \frac{1}{n} B && (1.18)
\end{aligned}
$$

W is the sum of the empirical variance of each parallel chain, while B is the empirical variance of the long unique chain obtained by merging all the runs.

So, $\widehat{V}(\theta)$ is an estimated variance, that, when convergence is achieved, must be very close to W. Hence, a GR statistic far from 1 points out that convergence is not achieved.

The practical steps to apply the Markov Chain Monte Carlo technique consist in starting from a given joint prior distribution for the parameters and constructing Markov chains of the parameters using the Metropolis Hastings (Hastings (1970)) algorithm.

This algorithm is structured as follows:

1. Initialize X_0.

[3]However, as Roberts (1996) shows, for geometric ergodic chains we can apply central limit theorems for ergodic averages.

2. For each t (for the desired number of iterations) repeat the following steps:
   (a) Sample X^*_{t+1} from a proposal distribution $q(\cdot|X_t)$
   (b) Sample a uniform variate $U \sim Uniform(0,1)$
   (c) Compute the acceptance probability:
   $$\alpha(X^*_{t+1}, X_t) = \min\left(1, \frac{\pi(X^*_{t+1})q(X_t|X^*_{t+1})}{\pi(X_t)q(X^*_{t+1}|X_t)}\right)$$
   (d) If $U \leq \alpha(X^*_{t+1}, X_t)$ set $X_{t+1} = X^*_{t+1}$, else set $X_{t+1} = X_t$.

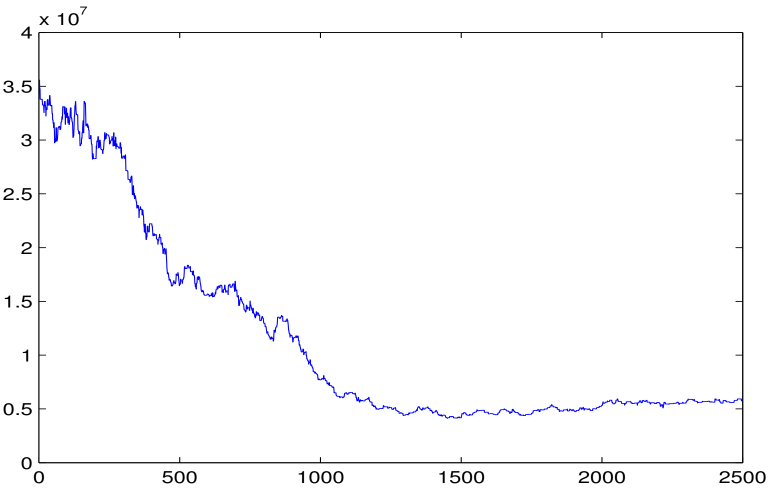

Figure 1.2: A chain for a parameter in the MCMC, with convergence to the stationary distribution.

Applied to our claims reserving case, while running the algorithm, it is straightforward to obtain reserves estimates by looking at the run-off triangles computed at each iteration. The samples can then be used for the estimation of quantiles and risk capital measures based on the VaR approach.

1.5 An ODP Bayesian model for claims reserving

The model presented in this section uses parameters with very intuitive economic meaning. It rests on an ODP assumption for the distribution of incremental claims, introducing uncertainty on the parameters describing both

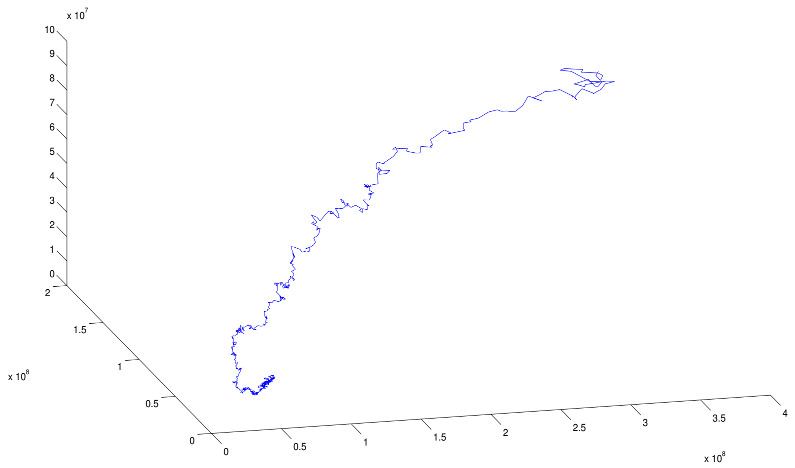

Figure 1.3: Convergent path shown from a 3-dimensional perspective. It is evident that, after being wandering for several iterations towards the stationary distribution, convergence is finally achieved, and the path concentrates around a smaller area.

the ultimate costs and the development factors in a Bayesian way. Two different distributional prior assumptions for the parameters are described.

The (over) dispersion parameter ϕ is simply estimated using Pearson's residuals. Even if one could argue that this prevents the model from being fully Bayesian, this choice is backed by two important considerations: first, ϕ has hardly an economic interpretation and, consequently, it will be hard to define a reasonable prior distribution to model it. Since the model is multiplicative, we do not have closed form for the reserves' posterior distribution. Hence, we resort to MCMC in order to obtain it. The model makes use of a Metropolis Hastings random walk algorithm for the implementation of the MCMC simulation technique, which proved to be very efficient, far more than the independence sampler. Since several studies found that the most efficient acceptance probability for the MH algorithm for d-dimensional target distributions with i.i.d. components is around 23.4% (Roberts and Rosenthal (2001)), we constructed a simple tool that automatically sets the proposal distribution's characteristics before running the algorithm, in order to ensure an acceptance rate very close to that value.

The model assumes that $\dfrac{X_{ij}}{\phi_i}$ are independently Poisson distributed with

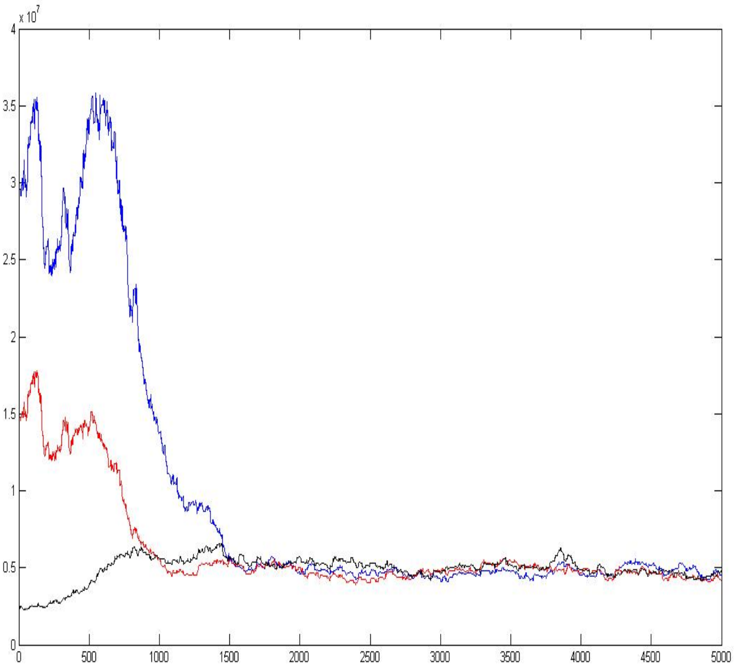

Figure 1.4: Convergence to the stationary distribution is achieved from different starting points after around 1500 iterations.

mean $\dfrac{\mu_i \gamma_j}{\phi_i}$.

$$\begin{aligned} \mathbb{E}\left[\frac{X_{ij}}{\phi_i}|\theta\right] &= \frac{\mu_i \gamma_j}{\phi_i} \qquad \mathrm{Var}\left[\frac{X_{ij}}{\phi_i}|\theta\right] = \frac{\mu_i \gamma_j}{\phi_i}, \\ \phi_i &> 0, \mu_k > 0 \;\forall\; k = 1, ..., I \;, \gamma_k > 0 \;\forall\; k = 1, ..., J, \end{aligned}$$

where $\theta = (\mu_1, ...\mu_I, \gamma_1, ..., \gamma_J, \phi_1, ..., \phi_I)$.

When ϕ_i 's are supposed deterministic (constant), as we will do, this is equivalent to assume that X_{ij} follows an overdispersed Poisson distribution with mean $\mu_i \gamma_j$ and variance $\mu_i \gamma_j \phi_i$. We choose multivariate normal or gamma distributions for the priors of μ and γ. We describe the modelling consequences of the two different prior choices in the following sections.

1.5.1 Link with GLM theory and CL parameters

Following Merz and Wuthrich (2008), we did not use directly the parameters used in the chain ladder. Such a parametrization limits the freedom in the choice of distributional assumptions, since it is necessary to require that the β_j' s - the proportion of the ultimate claim which is paid in d.y. j - sum up to 1. Hence, we defined parameters which are directly linked to the chain ladder estimates, but allow for almost unrestricted distributional assumptions: μ has the same meaning as the ultimate claim obtained with the chain ladder method, but in relative terms (taking the observed first year ultimate claim, x_1 as the benchmark), while γ is related to the development pattern. We characterize the model with the constraint $\mu_1 = 1$. The linearized version of the model, using GLM theory, follows:

$$\log \mu_i \gamma_j = c + \alpha_i + \beta_j, \tag{1.19}$$
$$\mu_1 = 1 \tag{1.20}$$

The use of this constraint, together with $\alpha_1 = 0$ immediately leads to the following transformations:

$$\gamma_j = e^{c+\beta_j} \tag{1.21}$$
$$\mu_i = e^{\alpha_i} \tag{1.22}$$

We can easily verify that the ML estimates of the parameters agree with the estimates obtained starting from chain ladder ones:

$$\widehat{\gamma}_1 = x_1 \beta_1^{CL} \tag{1.23}$$
$$\widehat{\gamma}_j = x_1 \left(\beta_j^{CL} - \beta_{j-1}^{CL} \right), \quad j = 2, ...n \tag{1.24}$$
$$\widehat{\mu}_i = \frac{x_i^{CL}}{x_1}, i = 2, ...n \tag{1.25}$$

1.5.2 Multivariate normal priors

We assume a multivariate normal distribution for the priors of $\theta = (\mu_2, ..., \mu_I, \gamma_1, ..., \gamma_J)$. We set, as normalization constraint - as we described in the previous section - $\mu_1 = 1$,and $\phi_i = \phi$constant and equal to the Pearson residuals' estimate from the triangle of incremental claims. We run a Metropolis Hastings algorithm with proposal distribution

$$q(\theta^* | \theta_t) \sim N(\theta_t, \Sigma^{prop})$$

Different ways of choosing the proposal distribution were tested: an algorithm in which the variance covariance matrix of the proposal distribution was updated at each step and an independence sampler proved to be less efficient than the random walk algorithm with fixed variance/covariance matrix we finally used.
The acceptance probability is computed at each step as the minimum between 1 and the ratio between the likelihoods evaluated at $\theta^* \sim q(\theta^*|\theta_t)$ and at θ_t. This ratio can be written as

$$\frac{f(X_{ij}, \theta^*) u_{\theta^*}\ q(\theta_t|\theta^*)}{f(X_{ij}, \theta_t) u_\theta\ q(\theta^*|\theta_t)},$$

where f(X$_{ij}$, ·)denotes the likelihood function of X$_{ij}$ evaluated at · and $u_\cdot$ is the prior density evaluated at ·.
We now write explicitly the likelihood ratio of the ODP model using multivariate normal proposals for both μ and γ and a random walk Metropolis Hastings algorithm:

$$LR = \frac{\prod_{i+j\le I}\left[\exp\left(-\frac{\mu_i^*\gamma_j^*}{\phi}\right)\frac{\left(\frac{\mu_i^*\gamma_j^*}{\phi}\right)^{\frac{X_{ij}}{\phi}}}{\left(\frac{X_{ij}}{\phi}\right)!}\right]* \frac{1}{(2\pi)^{N/2}|\Sigma_{prior}|}\exp\left(-\frac{1}{2}\left(\theta^*-\mu^{prior}\right)'\Sigma_{prior}^{-1}\left(\theta^*-\mu^{prior}\right)\right)* \frac{1}{(2\pi)^{N/2}|\Sigma_{prop}|}\exp\left(-\frac{1}{2}\left(\theta_t-\mu_{prop}^{(t)}\right)'\Sigma_{prop}^{-1}\left(\theta_t-\mu_{prop}^{(t)}\right)\right)}{\prod_{i+j\le I}\left[\exp\left(-\frac{\mu_i^{(t)}\gamma_j^{(t)}}{\phi}\right)\frac{\left(\frac{\mu_i^{(t)}\gamma_j^{(t)}}{\phi}\right)^{\frac{X_{ij}}{\phi}}}{\left(\frac{X_{ij}}{\phi}\right)!}\right]* \frac{1}{(2\pi)^{N/2}|\Sigma_{prior}|}\exp\left(-\frac{1}{2}\left(\theta_t-\mu^{prior}\right)'\Sigma_{prior}^{-1}\left(\theta_t-\mu^{prior}\right)\right)* \frac{1}{(2\pi)^{N/2}|\Sigma_{prop}|}\exp\left(-\frac{1}{2}\left(\theta^*-\mu_{prop}^*\right)'\Sigma_{prop}^{-1}\left(\theta^*-\mu_{prop}^*\right)\right)} \tag{1.26}$$

Thus, getting rid of some terms and passing to the log likelihood ratio we get:

$$\log LR = \sum_{i+j\leq I}\left[\frac{X_{ij}}{\phi}\log\left(\frac{\mu_i^*\gamma_j^*}{\mu_i^{(t)}\gamma_j^{(t)}}\right)-\frac{\mu_i^*\gamma_j^*}{\phi}+\frac{\mu_i^{(t)}\gamma_j^{(t)}}{\phi}\right]+ \quad (1.27)$$

$$+\frac{1}{2}\left[\begin{array}{c}\left(\theta_t-\mu^{prior}\right)'\Sigma_{prior}^{-1}\left(\theta_t-\mu^{prior}\right)+\\ -\left(\theta^*-\mu^{prior}\right)'\Sigma_{prior}^{-1}\left(\theta^*-\mu^{prior}\right)\end{array}\right]+ \quad (1.28)$$

$$+\frac{1}{2}\left[\begin{array}{c}\left(\theta^*-\mu_{prop}^*\right)'\Sigma_{prop}^{-1}\left(\theta^*-\mu_{prop}^*\right)+\\ -\left(\theta_t-\mu_{prop}^{(t)}\right)'\Sigma_{prop}^{-1}\left(\theta_t-\mu_{prop}^{(t)}\right)\end{array}\right] \quad (1.29)$$

in which we can recognize the parts of the likelihood ratio respectively due to the ODP model(1.27), to the prior(1.28) and to the proposal(1.29) distributions.

The logarithmic transformation we performed is made for computational convenience, in order to avoid overflows while running the algorithm. Since we made a monotonic transformation, our Metropolis Hastings algorithm is still proper, the acceptance probability being $min(0, \log LR)$.

The MH algorithm we implemented is a random walk one and works as follows:

1. Sample μ_i^*, γ_j^* for $i = 2, ...I,\ j = 1, ...J$ from the multivariate normal distribution $q\left(\theta^*|\theta_t\right) = N\left(\theta_t, \Sigma^{prop}\right)$.

2. Draw a random uniformly distributed number $U \sim Unif(0, 1)$

3. If $log\ U \leq \min\left(0, \log LR\left(\theta^{(*)}, \theta^{(t)}\right)\right)$,then set $\theta^{(t+1)} = \theta^*$; else set $\theta^{(t+1)} = \theta^{(t)}$.

1.5.3 Gamma priors

In this section we explore the implementation of the model when a multivariate Gamma is chosen as the prior distribution for parameters. The Gamma distribution is conjugate to the Over Dispersed Poisson we chose for the incremental claims. Thus, if one fixes one of the two parameters to a constant value - for example if we give to the $\gamma's$ their ML estimates - analytical expressions for the posterior distributions are available (Merz and Wuthrich (2008)). However, if we want to implement a Bayesian model with uncertainty on both the $\mu's$ and the $\gamma's$, we have again to apply simulation techniques. The parameters of the gamma distribution are updated at each step in order to set the mean of the distribution to the current values of $\mu^{(t)}$ and $\gamma^{(t)}$. The prior distribution is set to match the expert judgement on the cv and mean of each parameter. As before, we keep the overdispersion parameter deterministic and constant: $\phi_i = \phi$.

Hence, we have μ such that

$$\begin{aligned} \mu_1 &= 1, \quad (1.30) \\ \mu_i &\sim \Gamma(a, b_i) \;\; i = 2, ..., I \quad (1.31) \\ & \text{with } a = \frac{1}{cv(\mu)^2} \text{ and } b_i = \frac{\mu_i^{prior}}{a} \end{aligned}$$

and γ

$$\begin{aligned} \gamma_j &\sim \Gamma(c, d_j), \;\; j = 1, ..., J \quad (1.32) \\ & \text{with } c = \frac{1}{cv(\gamma)^2} \text{ and } d_j = \frac{\gamma_j^{prior}}{c} \end{aligned}$$

The proposal distributions for implementing the Metropolis Hastings random walk algorithm are also gamma distributions and have chosen coefficients of variation (cv) and fixed shape parameters. The scale parameters are updated at each step in order to let the mean coincide with the value of the chain at the previous step:

$$\begin{aligned} q(\mu_i^* | \mu_i^{(t)}) &\sim \Gamma(e, f_i) \quad (1.33) \\ & \text{with } e = \frac{1}{cvp(\mu)^2} \text{ and } f_i = \frac{\mu_i^{(t)}}{e} \end{aligned}$$

$$\begin{aligned} q(\gamma_j^* | \gamma_j^{(t)}) &\sim \Gamma(g, h_j) \quad (1.34) \\ & \text{with } g = \frac{1}{cvp(\gamma)^2} \text{ and } h_j = \frac{\gamma_j^{(t)}}{g} \end{aligned}$$

Then, the likelihood ratio follows:

$$LR = \frac{\prod_{i+j\leq I} \exp\left(-\frac{\mu_i^*\gamma_j^*}{\phi}\right) \frac{\left(\frac{\mu_i^*\gamma_j^*}{\phi}\right)^{\frac{X_{ij}}{\phi}}}{\left(\frac{X_{ij}}{\phi}\right)!} * \prod_{i\leq I} \mu_i^{*(a-1)} \frac{\exp\left(-\frac{\mu_i^*}{b_i}\right)}{\Gamma(a) b_i^{(a)}} \prod_{j\leq J} \gamma_j^{*(c-1)} \frac{\exp\left(-\frac{\gamma_j^*}{d_j}\right)}{\Gamma(c) d_j^{(c)}} * \prod_{i\leq I} \mu_i^{(t)\ (e-1)} \frac{\exp(-e)}{\Gamma(e)\left(\frac{\mu_i^{(t)}}{e}\right)^e} \prod_{j\leq J} \gamma_j^{(t)\ \ (g-1)} \frac{\exp(-g)}{\Gamma(g)\left(\frac{\gamma_j^{(t)}}{g}\right)^g}}{\prod_{i+j\leq I} \exp\left(-\frac{\mu_i^{(t)}\gamma_j^{(t)}}{\phi}\right) \frac{\left(\frac{\mu_i^{(t)}\gamma_j^{(t)}}{\phi}\right)^{\frac{X_{ij}}{\phi}}}{\left(\frac{X_{ij}}{\phi}\right)!} * \prod_{i\leq I} \mu_i^{(t)\ (a-1)} \frac{\exp\left(-\frac{\mu_i^{(t)}}{b_i}\right)}{\Gamma(a) b_i^{(a)}} \prod_{j\leq J} \gamma_j^{(t)\ \ (c-1)} \frac{\exp\left(-\frac{\gamma_j^{(t)}}{d_j}\right)}{\Gamma(c) d_j^{(c)}} * \prod_{i\leq I} \mu_i^{*(e-1)} \frac{\exp(-e)}{\Gamma(e)\left(\frac{\mu_i^*}{e}\right)^e} \prod_{j\leq J} \gamma_j^{*\ \ (g-1)} \frac{\exp(-g)}{\Gamma(g)\left(\frac{\gamma_j^*}{g}\right)^g}} \tag{1.35}$$

and the log-likelihood:

$$\begin{aligned}
\log LR \quad &= \sum_{i+j\leq I} \left[\frac{X_{ij}}{\phi} \log\left(\frac{\mu_i^*\gamma_j^*}{\mu_i^{(t)}\gamma_j^{(t)}}\right) - \frac{\mu_i^*\gamma_j^*}{\phi} + \frac{\mu_i^{(t)}\gamma_j^{(t)}}{\phi}\right] + && (1.36)\\
&+ \sum_{i\leq I} (a-1)\log\frac{\mu_i^*}{\mu_i^{(t)}} + \frac{1}{b_i}\left(\mu_i^{(t)} - \mu_i^*\right) && (1.37)\\
&+ \sum_{i\leq I} (e-1)\log\frac{\mu_i^{(t)}}{\mu_i^*} + e\ \log\frac{\mu_i^*}{\mu_i^{(t)}} && (1.38)\\
&+ \sum_{j\leq J} (c-1)\log\frac{\gamma_j^*}{\gamma_j^{(t)}} + \frac{1}{d_j}\left(\gamma_j^{(t)} - \gamma_j^*\right) && (1.39)\\
&+ \sum_{j\leq J} (g-1)\log\frac{\gamma_j^{(t)}}{\gamma_j^*} + g\ \log\frac{\gamma_j^*}{\gamma_j^{(t)}} && (1.40)
\end{aligned}$$

in which we can recognize respectively the contribution of the ODP model (1.36), of the prior on μ (1.37), of the proposal for μ (1.38), of the prior on γ (1.39) and of the proposal on γ (1.40).

Another way of implementing this same model consists in fixing the variance of the proposal distribution instead of the coefficient of variation.

1.6 Applying the model in practice

The model obviously needs prior distributions for the parameters and thus at least an estimate of their mean and of how precise this estimate is expected to be. Here comes the expert judgement to provide this information. The reserve evaluation - the output of the model - which has still its basis on statistical methods (namely the Chain Ladder approach) can incorporate input data which are not captured by the historical triangle of paid claims.
We apply our model to the undiscounted 1997-2008 triangle of the MTPL Line of Business of Fondiaria-Sai Spa, which is one of the largest Italian insurance companies. We chose to use statutory reserves as a part of the expert judgement process to obtain priors x_i^P on the ultimate claims. We choose the mean of the distribution for each a.y. i to be the sum of the statutory reserve (R_i^{STAT}) and the value on the diagonal of the triangle of cumulative paid claims, which represents the last observation on total claim payments relative to each accident year:

$$\mathbb{E}[x_i^P] = D_{i,n-i+1} + R_i^{STAT} \tag{1.41}$$

We use chain ladder estimates as prior means $\mathbb{E}[y_j^P]$ for the development pattern:

$$\mathbb{E}[y_j^P] = \beta_j^{CL} - \beta_{j-1}^{CL} \tag{1.42}$$

It is indeed worth noticing that prior reserves' mean, $\mathbb{E}[R_i^P]$, assuming independence between all the parameters, is different from statutory reserves:

$$\mathbb{E}[R_i^P] = \sum_{i=2}^{n} \sum_{j=n-i+2}^{n} \mathbb{E}[x_i^P] * \mathbb{E}[y_j^P] = \sum_{i=2}^{n} \sum_{j=n-i+2}^{n} \left(D_{i,n-i+1} + R_i^{STAT}\right) * y_j^P \tag{1.43}$$

1.6.1 Tests of convergence to the prior and to the chain ladder estimates

We carried out some preliminary tests in order to establish the convergence and stability properties and quality of the different MH algorithms we implemented. For the multivariate normal priors, we tested two different versions

of the simulation algorithm: one in which the parameters were updated all at once at each step of the Metropolis-Hastings algorithm, and one in which they were updated in blocks (the μ's first and then the γ's). For the Gamma priors, instead, as described in section 1.4, we used two different ways of describing the prior and proposal distributions (fixing the coefficient of variation (cv) - i.e. the ratio of the standard deviation and the mean of the distribution - or the variance of the proposal distribution as constants throughout all the algorithm).
We based are tests on the observation that, using precise priors, the reserves' "best estimate" should converge to the prior one, since the information given is supposed to be very precise and "drives" the result, while, using vague priors, it is supposed to converge to the chain ladder estimate, since the importance given to observed data is maximal.
We discarded a burn-in of 50000 iterations and analyzed the convergence on a long run of 250000 iterations. Table 1.6.1 summarize the results obtained on the triangle of the MTPL LoB. We used the triangle of dimension 12×12 collecting the claims amount paid for a.y. 1997 to 2008. While convergence to the prior is ensured by every model and method we used, convergence to the chain ladder while using vague priors is a bit more difficult and, looking at the tables, it is evidently faster and more precise for the Gamma prior choice implemented by fixing the coefficient of variation of the proposal distributions.
MCMC methods naturally lead, especially when the acceptance probability of the MH algorithm is set to be low, to high posterior autocorrelation in the chains. This can result in a slow exploration of the whole domain of the posterior distribution. To avoid this problem, one can consider the draws sampled every k iterations of the algorithm. A *thinning the chain* algorithm of this kind was implemented, in order to reduce autocorrelation in the chains. Anyway, the results and convergence properties for our model did not different significantly, while the efficiency was clearly reduced (because of the discarded iterations).

As Table 1.6.1 highlights, some convergence problem are found, for very vague priors, for the Normal prior choice model. In particular, the last element of the chain is the most subject to those problems, since it is obtained through the use of the lowest number of observed data.

Convergence to the chain ladder estimates				
Priors	Normal		Gamma	
	En bloc updating	Single component updating	fixed cv	fixed variance
$\frac{R_{TOT}^{MCMC} - R_{TOT}^{CL}}{R_{TOT}^{CL}}$	1.4%	1.0%	0.03%	1.1%
$max_i \left(\frac{R_i^{MCMC} - R_i^{CL}}{R_i^{CL}} \right)$	11.6% (year 1)	9.0% (year 1)	5% (year 1)	24.9%
$max_i \left(\frac{\mu_i^{MCMC} - \mu_i^{CL}}{\mu_i^{CL}} \right)$	0.7%	0.6%	0.4%	0.1%
$max_i \left(\frac{\gamma_i^{MCMC} - \gamma_i^{CL}}{\gamma_i^{CL}} \right)$	11.6%	8.5%	5.3%	24.6%
Convergence to the prior mean				
Priors	Normal		Gamma	
	En bloc updating	Single component updating	fixed cv	fixed variance
$\frac{R_{TOT}^{MCMC} - R_{TOT}^{PRIOR}}{R_{TOT}^{PRIOR}}$	0.2%	0.2%	0.2%	0.2%
$max_i \left(\frac{R_i^{MCMC} - R_i^{PRIOR}}{R_i^{PRIOR}} \right)$	0.9%	0.9%	0.9%	0.9%
$max_i \left(\frac{\mu_i^{MCMC} - \mu_i^{PRIOR}}{\mu_i^{CL}} \right)$	0.9%	0.9%	0.9%	1.0%
$max_i \left(\frac{\gamma_i^{MCMC} - \gamma_i^{PRIOR}}{\gamma_i^{CL}} \right)$	2.1%	2.1%	2.1%	2.1%

Table 1.1: The upper panel of the Table analyzes convergence to the chain ladder for different models with vague priors (cv=3.5), while the lower one explores convergence to the prior means, when the priors are set to be precise (cv=0.01). The subscript $MCMC$ refers to Bayesian posterior estimates, while the subscript CL refers to Chain Ladder estimates.

The Normal prior choice model, moreover, has an important theoretical drawback. In fact, it theoretically allows for negative values in the chains[4]. An accurate tuning of the model - and eventually adequate controls - can easily prevent this from happening. Anyway, when the priors are very vague and the proposal distributions have a high coefficient of variation, it is likely that throughout the algorithm the system considers or accepts negative values in the chains at some step. This problem is surely one of the causes of the worst performance of the model with Normal priors with respect to the Gamma one.

1.6.2 Results

There is an important trade-off in claims reserving: on one side reserves constitute a cost for the firm, since they are kept for prudential purposes and can't be used for firm's core activities. On the other side they are mandatory, in the sense that they have to be there when claim payments to insured entities are due. Thus, a too high value of reserve stored prevents the firm from expanding itself, while a too low one exposes the company to the risk of being unable to repay its obligations. The challenge of our Bayesian model is then to give the right importance (in terms of variability) to the expert judgement - and hence establishing its precision: very precise prior obviously shrink the distribution of OLL towards the prior mean, while very vague priors cancel the effect of prior knowledge. The first consideration from our analysis is that the prior mean of OLLs - computed from realized losses and statutory reserves - is clearly lower than the one predicted by the Chain Ladder method. The consequence will be obvious: when the priors on ultimate costs are judged to be precise, the distribution of OLLs obtained following our Bayesian model will be shifted to the left with respect to the bootstrapped one.

We are interested in computing the following quantities:

- Worst Case Scenario (W): it is computed as the 99.5 percentile of the distribution of OLLs.
- Best Estimate (BE): it's the expected value of the distribution of OLLs.
- Unanticipated Loss (U): the difference between the "worst case scenario" (RR) and the best estimate.

[4]This problem is avoided when choosing Gamma priors, since the Gamma distribution has support $[0, \infty]$.

	Prior mean	**ML estimate**
μ_2	1.055	1.036
μ_3	1.163	1.147
μ_4	1.247	1.235
μ_5	1.343	1.334
μ_6	1.410	1.410
μ_7	1.399	1.436
μ_8	1.337	1.394
μ_9	1.354	1.427
μ_{10}	1.435	1.541
μ_{11}	1.227	1.556
μ_{12}	1.292	1.898
γ_1	52912764.33	52912764.33
γ_2	60161364.85	60161364.85
γ_3	27342320.13	27342320.13
γ_4	14084776.78	14084776.78
γ_5	9031656.063	9031656.063
γ_6	7264483.908	7264483.908
γ_7	5932867.041	5932867.041
γ_8	4877265.307	4877265.307
γ_9	3508310.21	3508310.21
γ_{10}	2513839.843	2513839.843
γ_{11}	1785509.988	1785509.988
γ_{12}	1453146.824	1453146.824

Table 1.2: Prior means and ML estimates of model parameters

Table 1.3 presents required reserves, best estimates and unanticipated losses for the different models and different prior cv's for the choice of prior we describe in Table 1.2. In the Appendix we report the results of the estimation of the "original" GLM model.

	cv	**W(99.5%)**	**BE**	**U**
Precise priors (cv 0.01)				
ODP Bootstrap	3.33 %	6.79E+08	6.24E+08	5.47E+07
Normal en bloc	0.50 %	5.17E+08	5.10E+08	6.66E+06
Normal single component	0.52 %	5.17E+08	5.10E+08	6.92E+06
Gamma fixed cv	0.52 %	5.17E+08	5.10E+08	6.86E+06
Gamma fixed variance	0.52 %	5.17E+08	5.10E+08	7.13E+06
Intermediate priors (cv 0.08)				
ODP Bootstrap	3.33 %	6.79E+08	6.24E+08	5.47E+07
Normal en bloc	2.17 %	6.11E+08	5.78E+08	3.32E+07
Normal single component	2.18 %	6.12E+08	5.79E+08	3.37E+07
Gamma fixed cv	2.33 %	6.16E+08	5.82E+08	3.40E+07
Gamma fixed variance	2.34 %	6.20E+08	5.83E+08	3.69E+07
Vague priors (cv 0.2)				
ODP Bootstrap	3.33 %	6.79E+08	6.24E+08	5.47E+07
Normal en bloc	2.90 %	6.60E+08	6.14E+08	4.54E+07
Normal single component	2.90 %	6.61E+08	6.14E+08	4.76E+07
Gamma fixed cv	2.91 %	6.63E+08	6.15E+08	4.86E+07
Gamma fixed variance	3.03 %	6.64E+08	6.14E+08	5.00E+07
Very vague priors (cv 1)				
ODP Bootstrap	3.33 %	6.79E+08	6.24E+08	5.47E+07
Normal en bloc	3.29 %	6.85E+08	6.30E+08	5.51E+07
Normal single component	3.49 %	6.87E+08	6.29E+08	5.79E+07
Gamma fixed cv	3.18 %	6.78E+08	6.25E+08	5.30E+07
Gamma fixed variance	3.42 %	6.79E+08	6.23E+08	5.56E+07

Table 1.3: This Table reports the VaR 99.5% (W), the mean (BE) and the Unexpected Loss (U) of the distribution of OLLs for the MTPL LoB of Fondiaria-Sai SpA for different prior choices and algorithms.

Table 1.2 highlights the fact that, especially for more recent accident years, the predictions given using balance sheets indications are very distant from the ones obtained using standard methods.

Table 1.3 shows that the models seem to agree in their predictions. Anyway, different choices of the prior distribution and (with very vague priors) of the algorithm specification, bring different results. A goodness-of-fit test for the four models can be carried out by backtesting. It is important to notice that no convergence problems are found, both by using a traceplot approach and by controlling the results using the Gelman-Rubin statistics (described in section 1.4). A very conservative choice of the burn-in (50000 iterations) was anyway made, in order to avoid any possible troubles. The priors we used naturally lead to a very low best estimate and risk capital, if they are judged to be precise (see the upper panel of Table 1.3). The distribution is in fact concentrated on the left of the ODP one, and its coefficient of variation is very low. The next figures (which refer to the Gamma Gamma with fixed cv model) show how, giving precise priors, the distribution of OLLs is concentrated around the prior mean (cv=0.5%), while, as the coefficient of variation of the prior distribution increases - and external information is meant to be less reliable - the distribution widens and approaches the bootstrap one.

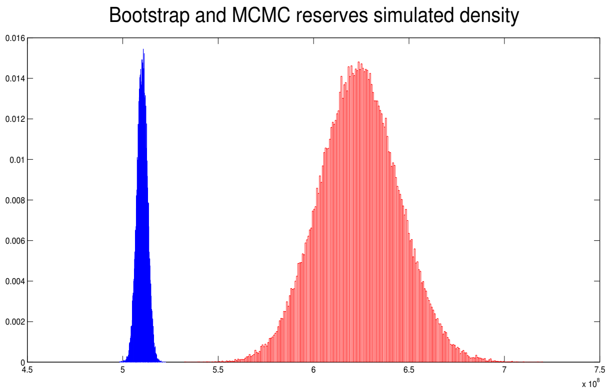

Figure 1.5: OLL distribution obtained from ODP bootstrap (red) and MCMC with precise priors (blue).

The coefficient of variation of the MCMC simulated distribution becomes higher than the ODP bootstrap distribution for very vague priors. Table 1.3 highlights that the OLL distribution obtained with the Gamma model presents is slightly more skewed to the right than the Normal one when priors are intermediate or vague. Also the Unexpected Loss predicted by the former model is higher. When priors are very vague, instead, the Normal model shifts to the right and presents both a higher mean (629 vs. 623 millions) and a higher 99.5 percentile (687 vs. 679 millions) thant the Gamma model. The same analysis was carried out for other LOBs' triangles. When dealing with lower dimension triangles (the Italian regulator mandates the use of 8x8 triangles for almost every line of business with the exception of the MTPL LoB when compiling survey modules), obviously convergence becomes more difficult because the number of observed data is lower and their variability is greater. Thus, results from the application of our Bayesian model are obviously less accurate.

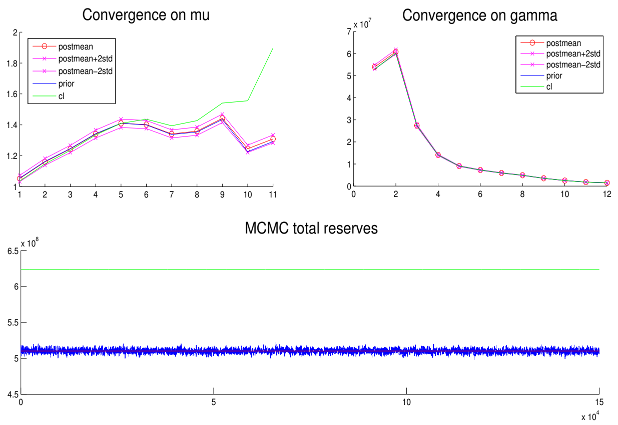

Figure 1.6: Convergence to prior parameters' means(μ, upper left panel, and γ, upper right panel) and chain of MCMC reserves. Green lines represent chain ladder figures (parameter estimates in the upper panels and reserve estimate in lower panel), while red lines represent prior figures. Notice the convergence towards these latter values.

1.7 Conclusions and further possible model extensions

There are two main directions of further development of the model: including inflation and aggregate OLLs across Lines of Business. The first issue can be naively addressed by rescaling the triangle of past settlements using some measure of inflation specific to the LoB (the issue of estimating specific inflation is indeed not at all simple) and then running the Bayesian model on the rescaled triangle. However, the whole procedure requires careful attention and is not easy to tackle. The most important problem lies in the estimation of LoB-specific inflation, which is very hard to perform. The second issue regards the aggregation of the reserve risk estimates at the company's level. In this chapter we estimated the risk capital requirements for each single line of business. The next step consists in finding a measure of the overall company and group risk capital requirements. Solvency II preliminary studies involved the use of a given correlation matrix to aggregate data. They thus implicitly assumed to be dealing with a Gaussian world.
In the next chapter, we specifically address this problem and we present a highly flexible model for aggregating outstanding loss liabilities across LoBs. The most important issue is defining and measuring the dependence struc-

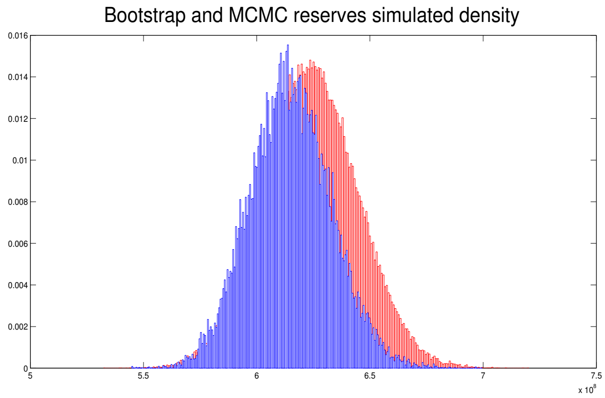

Figure 1.7: OLL distribution from ODP bootstrap (in red) and MCMC with vague priors from Table 1 and 2.

ture for the different line of business (and companies). We resort to the use of copulas and provide a theoretical framework for estimating the OLLs of a multiline non-life insurance company, extending the Bayesian approach we used in this chapter to a multi-dimensional setting.

1.8 Appendix

The following Table reports GLM parameter estimates for the triangle of incremental claim payments we used. It refers to the MTPL LoB of Fondiaria-Sai spa.

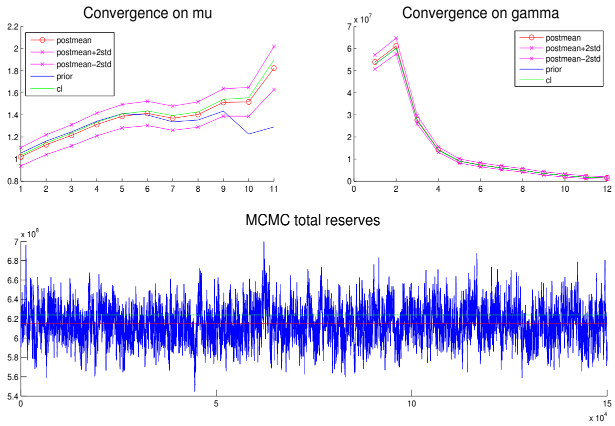

Figure 1.8: Convergence to the chain ladder parameters (μ, upper left panel, and γ, upper right panel) and chain of MCMC reserves. Green lines represent chain ladder figures (parameter estimates in the upper panels and reserve estimate in lower panel), while red lines represent prior figures. Notice that, with respect to Figure 4, parameters and reserves are shifting towards the chain ladder values.

Table 7: GLM parameter estimates

Parameter	Estimate
c	14.1892
α_1	0
α_2	0.0350
α_3	0.1371
α_4	0.2110
α_5	0.2884
α_6	0.3436
α_7	0.3619
α_8	0.3323
α_9	0.3558
α_{10}	0.4323
α_{11}	0.4419
α_{12}	0.6407
β_1	3.5949
β_2	3.7233
β_3	2.9347
β_4	2.2714
β_5	1.8270
β_6	1.6093
β_7	1.4068
β_8	1.2109
β_9	0.8814
β_{10}	0.5481
β_{11}	0.2060
β_{12}	0

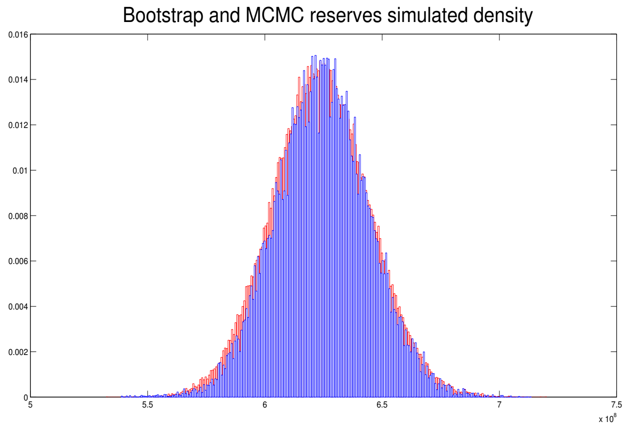

Figure 1.9: With very vague prior(cv=1), MCMC reserves' posterior distribution converge to the ODP bootstrap predictive distribution. With the Gamma/Gamma model with fixed cv, ODP reserves' distribution is larger with respect to the MCMC one, Best Estimates and W are very close to each other.

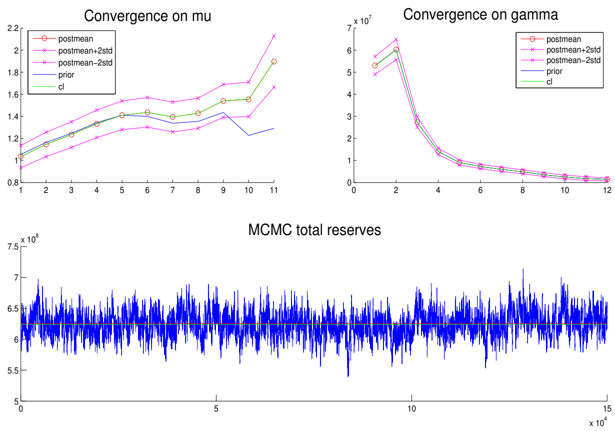

Figure 1.10: Posterior means plots are almost indistinguishable from chain ladder estimates ones, while total reserves fluctuate around the chain ladder estimate.

Chapter 2

A Bayesian copula model for stochastic claims reserving

2.1 Introduction

The estimation of Outstanding Loss Liabilities (OLLs) is crucial to reserve risk evaluation in risk management. Classical methods based on run-off triangles need a small amount of input data to be used. This fact determined their fortune, making them immediate to use, requiring the knowledge of triangle of annual paid claims amount only. However, this fact constitutes also an important shortcoming, since using a small sample of data to predict future outcomes may possibly lead to inaccurate estimates. Anyway, their widespread use in professional practice encourages further improvements to limit this problem.
Starting from the beginning of this century, bayesian methods in estimating run-off triangles gained increasing attention as a tool to include expert judgement in stochastic models[1] and enlarge the information set on which reserves are computed. The use of Bayesian methods in loss reserving started decades ago, but it was the possibility of using MCMC fast computer-running algorithms that gave high flexibility to the application of this methodology, allowing for almost unrestricted distributional assumptions. De Alba (2002), De Alba and Nieto-Barajas (2008) - who introduced correlation among different accident years - and Ntzoufras and Dellaportas (2002) offer examples of how Bayesian methods can be implemented in the estimation of outstanding claims for a line of business, introducing prior information on both future claim amount (ultimate costs) and frequency. Simultaneously, some works

[1]For a nice treatment on the use copulas to aggregate expert opinions, see for example the seminal work Jouini and Clemen (1996).

tried to introduce the use of copulas - which gained increasing popularity in the finance world in the last decade - also in loss reserving[2].

The question of how to cope with dependent risks such as the losses an insurance company has to face in its different lines of business (LoBs) is surely of utmost importance. Current practice and Solvency II standard formulas account for diversification by means of linear correlation matrices estimated on a market-wide basis. Obviously, these correlation matrices can fail to capture the specificities insurance companies can present, due to geographical reasons or management choices.

A few papers studied the application of copulas to run-off triangles estimation. Tang and Valdez (2005) used simulated loss ratios to aggregate losses from different LoBs. Li (2006) compared aggregation through the use of different copula functions, given distributional assumptions on the marginals. More recently, De Jong (2009) introduced a Gaussian copula model to describe dependence between LoBs.

This paper aims at combining both these two important aspects: bayesian methods and the use of copulas. The bayesian approach, introducing data coming from expert judgement, allows to include additional reliable information when estimating reserves and to derive full predictive distibutions. Copulas allow to obtain joint distributions in an easily tractable way, separating the process of defining the marginals and the dependence structure. Hence, we introduce prior information on the dependence structure, using Bayesian copulas in the aggregation of losses across LoBs. Up to our knowledge, this paper is the first attempt in introducing Bayesian copulas in stochastic claims reserving. Dalla Valle (2009) applied a similar technique to the problem of the estimation of operational risks. We adapt their approach to the aggregation of OLLs from different LoBs.

Combining a Bayesian approach to derive the marginal distributions of OLLs for each single LoB and the use of Bayesian copulas to aggregate them, one obtains a fully Bayesian model that incorporates expert judgement on the ultimate costs and development pattern of each LoB as well as on the dependence structure between them.

We apply this model to four lines of business of an Italian insurance company. We compare results obtained from the Bayesian copula model with those obtained from a standard copula approach.[3].

The outline of the paper is the following. Section 2.2 presents a simple

[2]Copulas have been recently used in individual claim models (Zhao and Zhou (2010)).

[3]Financial literature offered only few examples of application of non bivariate copulas. This paper, testing the theoretical framework on a multi-line insurance company, provides a four-dimensional application of our model of aggregation through copulas and a comparison of results for different copula choices

Bayesian model, which uses Markov Chain Monte Carlo (MCMC) simulation methods to derive the predictive distribution of OLLs for each LoB. Section 2.3 motivates the choice of modeling dependence between LoBs and briefly reviews the most important notions on the theory of copulas. Section 2.4 presents the Bayesian copula approach. Section 2.5 applies the model to a large insurance company, reports and compares the results. Section 2.6 concludes.

2.2 A Bayesian approach for computing Lob's reserves

First, we briefly present a bayesian model for the estimation of the OLLs for the single LoBs. We assume that an Over-Dispersed Poisson (ODP) distribution models incremental claims in the run-off triangle. Then, denoting with X_{ij} the claim payments in the development year (d.y.) j concerning accident year (a.y.) i and with ϕ_i the overdispersion coefficient for accident year i, we assume that $\frac{X_{ij}}{\phi_i}$ are independently Poisson distributed with mean $\frac{\mu_i \gamma_j}{\phi_i}$:

$$\mathbb{E}\left[\frac{X_{ij}}{\phi_i}|\Theta\right] = \frac{\mu_i \gamma_j}{\phi_i},$$

$$\text{Var}\left[\frac{X_{ij}}{\phi_i}|\Theta\right] = \frac{\mu_i \gamma_j}{\phi_i},$$

$$\phi_i > 0, \mu_i > 0$$

$\forall\ i = 1, ..., I, \gamma_j > 0\ \forall\ j = 1, ..., J,$

$$\Theta = \mu_1, ..., \mu_I, \gamma_1, ..., \gamma_J, \phi_1, ..., \phi_I,$$

We renormalize the model setting the (observed) parameter $\mu_1 = 1$. μ_i' s then represent ultimate claims relative to year 1, while the γ_j' s represent the development pattern in monetary terms relative to ultimate cost of a.y. 1. This renormalization allows to increase flexibility in distributional assumptions, avoiding the awkward constraint that the parameters of the development pattern by d.y. have to sum up to 1.

We estimate the overdispersion parameter ϕ using the Pearson's residuals obtained from the triangle and assume it constant across accident years. Although one can object that this prevents the model from being fully bayesian, this choice is backed by two important considerations: first, ϕ has hardly a simple economic interpretation and, consequently, it will be hard to define a reasonable prior distribution to model it. Moreover, the MCMC algorithm turns out to be considerably more stable if ϕ is not bayesian. We choose the prior distribution of both the μ's and the γ's to be independently gamma distributed. We don't have analytical expressions for the posterior distribution. Hence, we set up a Markov Chain Monte Carlo algorithm in order to simulate the posterior distribution of parameters. The prior distribution is set through coefficient of variation (cv) and mean; at each step, we update the parameters to match the current mean values of $\mu^{(t)}$ and $\gamma^{(t)}$, where t is the iteration step in the simulation algorithm. Hence,

$$\mu_1 = 1$$

$$\mu_i \sim \Gamma(a, b_i) \ i = 2, ..., I$$

$$\text{with } a = \frac{1}{cv(\mu)^2} \text{ and } b_i = \frac{\mu_i^P}{a}$$

and

$$\gamma_j \sim \Gamma(c, d_j) \ j = 1, ..., J$$

$$\text{with } c = \frac{1}{cv(\gamma)^2} \text{ and } d_j = \frac{\gamma_j^P}{c}$$

We implement the MH algorithm with gamma proposal distributions, whose coefficient of variation is kept fix throughout the algorithm. The lower part of the triangle is obtained through simulation and then discounted using the term structure of interest rates at the end of the last a.y..

2.3 A copula approach to aggregate across LoBs

In the previous section we presented a way of retrieving the predictive distribution of OLLs for a single line of business. From now on, we address the problem of generating the joint distribution of OLLs from different LoBs, in order to estimate prudential reserves for multi-line insurance companies. Notice that what follows can be applied independently of the choice of the method used to obtain the predictive distribution of reserves.
Correctly capturing the presence of dependence between the losses in different LoBs is intuitively a desirable feature of a good model for claims reserving. The following Table compares the correlation matrix between the LoBs of an Italian insurance company, estimated from a time series of loss ratios, and

the one the CEIOPS mandated to use when calculating reserve risk with the standard formula in the Quantitative Impact Studies (QIS):

LoB				
	MTPL	**MOC**	**FP**	**TPL**
MTPL	1 (0)	0.4751 (0.0463)	0.4598 (0.0549)	0.5168 (0.0281)
MOC	0.4751 (0.0463)	1 (0)	0.8789 (0.000001)	0.7331 (0.0005)
FP	0.4598 (0.0549)	0.8789 (0.000001)	1 (0)	0.8748 (0.000002)
TPL	0.5168 (0.0281)	0.7331 (0.0005)	0.8747 (0.000002)	1 (0)

Table 2.1: Linear correlation between LoBs. The brackets report p-values.

LoB				
	MTPL	**MOC**	**FP**	**TPL**
MTPL	1	0.25	0.25	0.5
MOC	0.25	1	0.5	0.25
FP	0.25	0.5	1	0.25
TPL	0.5	0.25	0.25	1

Table 2.2: This Table reports the correlation matrix the CEIOPS estimated and requires the participants to the Quantitative Impact Studies (QIS) to use in the evalutaion of reserves.

Table 2.1 clearly shows that the "industry-wide" estimate proposed by CEIOPS and the industry-specific ones differ. Results on the correlation of a time series of realized losses, which we will present in Section 2.5 further support this evidence.
We then turn to the use of copulas in order to model the dependence between LoBs. Indeed, they allow us to separate the estimation of the characteristics of the dependence structure from the modeling of marginal distributions.

2.3.1 Copulas

In this section we briefly give the basic definitions and fundamental notions about copulas. We are interested in modelling the joint distribution $F(L_1, ..., L_n)$, where L_i denotes the OLLs of the i-th LoB of a company whose business involves n sectors, since our object of interest is

$$L_{tot} = \sum_{i=1}^{n} L_i$$

and its related percentiles. Copula functions permit us - as we will briefly show in this section - to separate the process of estimating the marginal distributions $F(L_1), ..., F(L_n)$ of the OLLs of each LoB from the estimation of the dependence structure. Moreover, the latter can be modeled in a highly flexible way, since many copula functions are available to describe it and capture its (also non-linear) properties. We recall the most important results on multivariate copulas, which we will use in the construction of our model.[4] First of all, we define multivariate copulas:

Definition 2.3.1 *An n-dimensional subcopula is a function C:*$A_1 \times A_2 \times ... \times A_n \to \mathbb{R}$*where, for each* $i, A_i \subset I$ *and contains at least 0 and 1, such that:*

1. *C is grounded*[5]
2. *its one-dimensional margins are the identity function on I:* $C_i(u) = u$, $i = 1, 2, ..., n$
3. *C is n-increasing.*[6].

C is a copula if it is an n-dimensional subcopula for which $A_i = I$ *for every i.*

The following (Sklar's) theorem proofs the link between a copula and the marginal distribution functions[7]:

[4] For a comprehensive review of the theory of copulas and their use in finance the reader can refer to Cherubini, Luciano, and Vecchiato (2004) and Nelsen (2006).

[5] Let $C : \mathbb{R}^{*n} \to \mathbb{R}$ be a function with domain $A_1 \times ... \times A_n$, where A_i are non-empty sets with a least element a_i. C is grounded iff it is null for every $\mathbf{v} \in$ Dom C, with at least one index k such that $v_k = a_k$:

$$C(\mathbf{v}) = C(v_1, v_2, ..., v_{k-1}, a_k, v_{k+1}, ..., v_n) = 0$$

[6] $C : A_1 \times A_2 \times ... \times A_n \to \mathbb{R}$ is n-increasing if:

$$\sum_{\mathbf{w} \in ver(A)} C(\mathbf{w}) \prod_{i=1}^{n} sgn(2w_i - u_{i1} - u_{i2}) \geqslant 0$$

where *ver(A)* is the set of vertices of A.

[7] For a proof of this theorem in the multivariate case we refer the reader to Schweizer and Sklar (1983).

Theorem 2.3.2 *Let $F_1(x_1), ..., F_n(x_n)$ be marginal distribution functions. Then, for every $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^{*n}$:*

1. *If C is any subcopula whose domain contains $RanF_1 \times ... \times RanF_n$, $C(F_1(x_1), ..., F_n(x_n))$ is a joint distribution function with margins $F_1(x_1)$, ..., $F_n(x_n)$.*

2. *Conversely, if F is a joint distribution function with margins $F_1(x_1)$, ... ,$F_n(x_n)$ there exists a unique subcopula C, with domain $RanF_1$ $\times ... \times RanF_n$ such that $F(\mathbf{x}) = C(F_1(x_1), ..., F_n(x_n))$.*

If the marginals are continuous, the subcopula is a copula; if not, there exists a copula $\mathcal{C}$ such that

$$\mathcal{C}(u_1, ..., u_n) = C(u_1, ..., u_n) \text{ for every } (u_1, ..., u_n) \in RanF_1 \times ... \times RanF_n.$$

This is the most important result in the theory of copulas: it states that - as we pointed out before - starting from separately determined marginals and dependence structure copula functions allow to represent a joint distribution function of the variables involved. Moreover, overall uniqueness is ensured when marginals are continuous; when they are discrete, uniqueness is guaranteed on the domain $RanF_1 \times ... \times RanF_n$.

2.3.2 Applying copulas to claims reserving

We outline first a simple procedure to obtain a joint distribution of OLLs for an n-dimensional non-life insurance company through the use of copulas:

1. derive the marginal distributions of the OLLs $F(L_1), ..., F(L_n)$ for each LoB independently. For this task, it is possible to resort to classical methods, simulation, as well as to the Bayesian technique we described in Section 2.2.[8]

2. estimate the dependence structure between the L_i's for $i = 1, ...n$.

3. choose a convenient copula function and evaluate its parameter(s). The copula will satisfy the uniqueness properties of Theorem 2.3.2, depending on the form of its marginals.

[8] Tang and Valdez (2005) used simple distributional assumptions on the marginals

Sampling from any n-dimensional copula obtained can be done exploiting the properties of conditional distributions. Then, we can easily evaluate the quantities of interest on the simulated sample. Difficulties in the procedure above lie mainly on the correct estimation of the dependence structure, which is a complicated task given the low (annual) frequency of the input data used in stochastic claims reserving models based on run-off triangles. The same observation applies to the choice of the most appropriate copula function. In section 2.5.2 we compare the results of evaluating the OLLs of a multi-line insurer under different copula assumptions.

2.4 A Bayesian copula approach

As we saw in Tables 2.1 and 2.2, industry-wide estimates of the dependence between the different LoBs and own assessments based on companies' tracks can differ. This can be due - for example - to geographical issues as well as to management actions or policies. Hence, including company-specific measures of dependence in reserves' estimation as expert judgement together with industry-wide estimates can help in improving the quality of the predictions of future losses. Hence, we present a Bayesian approach to the use of copulas, by adding uncertainty on the parameters of the copula function.
The procedure which has to be applied to implement the Bayesian copula model is similar to the one we described for standard copulas in the previous Section 2.3.2, but a few more steps are required:

1. choose a convenient distributional assumption for the prior of the copula parameter(s) θ, $\pi(\theta)$
2. compute, using Bayes' theorem, the posterior distribution of the parameter given the input data:

   $$f(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)\pi(\theta),$$

   where $\mathbf{x}$ denotes the $n \times T$ matrix of observations (T is the number of observations).

A convenient choice of the prior distribution requires the choice of priors whose densities are conjugate to the one of the distribution of the estimation object - in our context, OLLs per a.y.. We now present - as an example - the application of the procedure to a Gaussian copula choice.

2.4.1 Bayesian Gaussian copula

This Section introduces the use of Bayesian Gaussian copulas for aggregating marginal distribution of OLLs. The choice of a Gaussian copula is the most immediate one and it entails using a linear measure of dependence - linear correlation - to represent the link between LoBs. Hence, we assume that OLLs across different LoBs are distributed following a multivariate Gaussian density. The multivariate Gaussian copula density is

$$c(u_1, ..., u_n|\Sigma) = |\Sigma|^{-1/2} exp\left\{-\frac{1}{2}\mathbf{x}'(\Sigma^{-1} - I_n)\mathbf{x}\right\}$$

where Σ denotes the $(n \times n)$ covariance matrix between the LoBs, I_n is an identity matrix of dimension n and $\mathbf{x}$ is a matrix of observed OLLs. Dalla Valle (2009) applied Bayesian Gaussian copulas to the estimation of operational losses. However the paper considered correlation matrices, incurring in the problem of requiring the priors to be vague - and thus uninformative. Starting from equation 2.4.1, we take the product over the T observation of the sample to obtain the likelihood function:

$$f(\mathbf{x}|\Sigma) = |\Sigma|^{-T/2} exp\left\{-\frac{1}{2}\sum_{i=1}^{T}(\mathbf{x_i} - \mathbf{m})'(\Sigma^{-1})(\mathbf{x_i} - \mathbf{m})\right\}$$

We choose the Inverse Wishart as a prior distribution for Σ. Inverse Wishart distributions are commonly used to represent covariance matrices and the attractive property of being conjugate to the multivariate Gaussian. Hence we set

$$\pi(\Sigma) \sim W^{-1}(\alpha, B).$$

, where $W^{-1}(\alpha, B)$ denotes the Inverse Wishart Distribution with α degrees of freedom and precision B. We can write its probability density function

$$\frac{|B|^{\alpha/2}\, |\Sigma|^{-(\alpha+n+1)/2}\, e^{-tr(B\Sigma^{-1})}}{2^{\alpha n/2}\Gamma_n(\alpha/2)}$$

and apply Bayes' theorem to compute its posterior distribution.

Following DeGroot (2004), we can easily conclude that the posterior distribution of Σ follows again an inverse Wishart distribution:

$$\pi(\Sigma|\mathbf{x}) \sim W^{-1}(T + \alpha, B + \sum_{i=1}^{T}\mathbf{x}_i\mathbf{x'}_i)$$

The precision parameter of the posterior inverse Wishart is then given by the sum of the precision parameter in the prior distribution and $T - 1$

times the sample covariance matrix. Since the mean of the inverse Wishart distribution is

$$\mathbb{E}[\Sigma] = \frac{B}{\alpha - n - 1}$$

we set α to $n+2$, making the precision matrix B coincide with our prior mean choice.
Then, in order to simulate a random sample from the Gaussian copula for OLLs, we first draw a covariance matrix from the inverse Wishart distribution and then we use it to generate outcomes from the Gaussian copula. Generating a sufficient number of outcomes from the Bayesian Gaussian copula is then easy and permits to derive a full distribution of aggregate OLLs.

2.5 An application to an Italian Insurance Company

In this section we apply the methodologies described in the previous parts of the paper to obtain and compare the predictive OLLs distribution of a multi-line insurance company for different (standard and bayesian) copula choices. First, in section 2.5.1 we derive the marginal distribution for each LoB as described in Section 2.2, then we compare the results obtained by aggregating the marginals using both standard and Bayesian copulas. Our dataset is composed by the paid claims triangles of a large insurance company from 2001 to 2008. We restricted our attention to its 4 most important LoBs, whose linear correlation estimates based on a time series of loss ratios - reported in Table 2.1 - were significant at least at a 10% level.
It is important to remark that we derive full predictive distributions and, as a consequence, that we can easily compute not only best estimates, but all the relevant percentiles[9]. Notice that the approach can be easily extended to derive a predictive distribution of the overall losses of a company, considering all the LoBs in which it is involved. However, we recognize that data quality must be high enough to return reliable estimates of the dependence structure.

2.5.1 Estimation of the marginals

Applying the method described in section 2.2 we derive the marginal distribution of OLLs 4 LoBs: Motor Third Party Liability (MTPL), Motor Other Classes (MOC), Fire and Property (FP) and Third Party Liability (TPL).

[9]This is important in terms of the VaR-approach followed by Solvency II.

Table 2.3 reports the most important figures of the predictive distribution for different choices of the precision of the priors. We define the mean of the prior distribution of ultimate costs (before renormalization) as

$$\mathbb{E}[\mu_i'] = \sum_{j=1}^{r-i+1} X_{ij} + R^S(i)$$

This means that, for each a.y., the prior mean is given by the sum of the last observed cumulative claim payment and the statutory reserves $R^S(i)$. The mean prior on the development pattern is simply the chain ladder estimate:

$$\mathbb{E}[\gamma_j'] = \beta_j^{CL} - \beta_{j-1}^{CL}$$

where

$$\beta_j^{CL} = \prod_{k=j}^{J-1} \frac{1}{\hat{\lambda}_k}, k = 1, ..., J-1$$

and $\hat{\lambda}_k$ are the chain ladder estimates of the development factors. Model parameters' prior mean μ_i, i=2,...,I and γ_j, j=1,...J are obtained as

$$\mathbb{E}[\mu_i] = \frac{\mathbb{E}[\mu_i']}{\mu_1}$$

$$\mathbb{E}[\gamma_j] = \mathbb{E}[\gamma_j'] * \mu_1$$

Table 2.3 compares the key figures of the predictive distribution for each LoB for different choices of the coefficient of variation of the priors.

It shows largely different OLL distributions for different prior choices. As soon as the priors become more vague the estimates converge to the chain ladder ones, while convergence to the prior is achieved when the priors themselves are precise. Hence, differences arise, since statutory reserves are not computed using the Chain Ladder method, but a different one which accounts for the speed of finalization and mean costs also. It is easy to notice from the Table that - as one could expect - as soon as the prior information becomes less precise the standard deviation of the distribution of OLLs increases.

2.5.2 Estimation of reserves through copulas

In this Section we present model results obtained from classical copula methods and compare the figures obtained with different copula choices. We first estimate copula parameters from adequately chosen time series data. We

decided to use loss ratios, following Tang and Valdez (2005), since they constitute the most reliable source of information. Our choice of using the historical loss ratio series is motivated by the lack of qualitatively useful data about our direct object of interest, losses, for which historical data are either unavailable or too far away in time.[10] Hence we implicitly assume that the correlation between loss ratios is a good proxy for the correlation of losses themselves. Moreover, loss ratios are a non-monetary measure, allowing us to abstract from the challenges of correctly capturing overall and LoB-specific inflation when estimating.
While Tang and Valdez (2005) used industry-wide estimates, we use a company-specific time series of loss ratios. We compare the results obtained from these industry-specific estimates with those obtained using the matrix proposed by the CEIOPS in the Quantitative Impact Study 5 (QIS 5).

We first deal with the Gaussian and the t copulas, using the (ML) linear correlation estimated matrix reported in Table 2.1 as the parameter. The Upper Panel of Table 2.4 compares the results from the Independence, the Gaussian and the Student's t copula with 4 degrees of freedom.

[10]In Section 2.5.3, however, we will be forced to derive a measure of losses per a.y. using observed data and statutory reserves

Precise Priors (cv 0.05)				
LoB	**4**	**5**	**7**	**8**
Mean	404388773.3	13519713.9	34505262.	65461545
Std Deviation	9732942.751	441754.4416	1081978.667	1580984.48
VaR 75	410995838.9	13816484.7	35222255.7	66524262.8
VaR 97.5	423699508	14392490.3	36666395.9	68564682.2
VaR 99	427297420.1	14567031.5	37072985.5	69171034.6
VaR 99.5	429308413.2	14681856.6	37329907.2	69600078.3
RC	24919640.0	1162142.6	2824644.6	4138533.1
Intermediate Priors (cv 0.1)				
LoB	**4**	**5**	**7**	**8**
Mean	433049373.5	13748473.7	35500766.5	68943031.8
Std Deviation	14963435.3	515438.8	1436387.9	2699322.4
VaR 75	442997813.3	14099498.8	36452316.5	70747579.4
VaR 97.5	463012090.3	14776090.2	38371337.7	74270158.7
VaR 99	469270928.2	14972190.1	38966614.0	75294702.3
VaR 99.5	473098522.3	15109019.8	39376102.7	76096070.4
RC	40049148.8	1360546.2	3875336.2	7153038.6
Vague Priors (cv 0.5)				
LoB	**4**	**5**	**7**	**8**
Mean	466678568.2	13823953.1	36158247.4	76376715.6
Std Deviation	23091344.0	589537.2	1798075.8	4849745.317
VaR 75	482416095.4	14222419.8	37337075.1	79526592.3
VaR 97.5	513180633.3	15015262.8	39834101.3	86254875.9
VaR 99	521989486.5	15246807.0	40609193.4	88270812.9
VaR 99.5	528830674.6	15412128.7	41120258.5	89657400.2
RC	62152106.4	1588175.7	4962011.2	13280684.6
Very Vague Priors (cv 1.5)				
LoB	**4**	**5**	**7**	**8**
Mean	469457550.6	13889439.6	36537741.4	77348490.4
Std Deviation	23657712.1	631536.8	1930043.5	4823628.448
VaR 75	485141513.3	14297531.7	37813249.2	80594262.3
VaR 97.5	516938239.6	15181185.6	40468961.1	87318681.5
VaR 99	527327189.6	15420777.1	41177121.6	89221792.3
VaR 99.5	534835778.5	15568471.2	41727870.5	90309880.2
RC	65378227.9	1679031.6	5190129.0	12961389.8

Table 2.3: This Table reports mean and VaR measures for the OLL distributions for each LoB for different choices of cv of the priors

Figures	Copula Type				
	Independence	**Gaussian MLE**	**Gaussian QIS**	**Student t MLE**	**Student t QIS**
Mean	593012692.6	592999468	593078422.2	593107009.4	593089632.8
Std Dev	23661871.8	27330843.4	26588306.7	27336552.8	26466577.4
VaR 75	609074201.8	611478073.8	610981167.6	611481550.3	610664068.4
VaR 97.5	640745885.5	647956557.4	646600022.3	648294764.2	646824832.7
VaR 99	649693108.9	658180123.7	657255148.1	659325286.3	658173341.8
VaR 99.5	656323827.6	665951381.4	664323502.5	666734284.0	665631981.4

Figures	Archimedean Copula Type		
	Clayton	**Frank**	**Gumbel**
Mean	592311157.2	592329320.5	592142996.5
Std Dev	25889681.9	25571649.1	25744039.5
VaR 75	610002820.2	609871400.4	609129608.0
VaR 97.5	642228178.2	643249755.5	644850509.7
VaR 99	651706417.7	652817211.1	656423299.3
VaR 99.5	657929168.7	659126935.2	664429091.2

Table 2.4: This Table compares mean, standard deviation and various percentiles of the joint OLL distribution of the 4 LoBs, obtained from different copula choices. The Upper Panel reports results from the Independence, the Gaussian and the t copula. In the "MLE" columns the parameter of the copula is the Maximum Likelihood estimator of the correlation matrix - the sample correlation - while in the "QIS" column the QIS5 correlation matrix reported in Table 2.2 was used as the copula parameter. The Lower Panel compares figures from different Archimedean copula choices.

Since no negative correlation between the LoBs is captured by the time series of loss ratios nor by the QIS 5 matrix, the Independence copula obviously offers the lowest level of Prudential Reserves (656 millions). Using the own assessment of correlation (in the Gaussian or t MLE column) returns a distribution with slightly fatter tails than the one obtained when using the CEIOPS QIS matrix. The consequent "risk capital" - which is the capital kept in excess of the best estimates of OLLs - is also higher when using the own assessment of correlation. We then turn our attention to Archimedean copulas. First, we estimate the Kendall's τ matrix from the time series of loss ratios. Since the corresponding parameters of the bivariate copulas involving the various LoBs differ, we used the recursive procedure of Genest (1987) and Genest and Rivest (1993) in order to generate random samples from the multivariate copulas. The procedure exploits the properties of conditional distribution functions. We describe the technical details and the algorithm of the procedure in the Appendix. The Lower panel of Table 2.4 reports also the figures obtained when using the Archimedean Clayton, Gumbel and Frank's copulas.

The 99.5th percentile of the OLL distribution - which is usually indicated as a standard measure for prudential reserves in the Solvency II framework - computed with the Clayton copula is lower than the one obtained with any other copula type, with the only exception of the Independence copula. On the contrary, Gumbel's copula predicted $VaR_{99.5\%}$ is the highest among the Archimedean families we compared, but its estimate is however lower than the one obtained from the Gaussian and the t copulas.

2.5.3 Estimation of reserves using Bayesian Copulas

Unfortunately, the procedure described in Section 2.4 can not be applied when the record of past losses is not sufficiently long and homogenous across years. In our data sample, information about past claim payments lacked the deepness to allow us to use a significant time series of observed OLLs. To overcome this problem, we derived a time series of losses adding the observed paid claims by a.y. for the d.y. available (inflated at a monetary rate of inflation provided by ISTAT) and the reserved amount at the end of the observation period obtained from the balance sheet. Using this time series, we obtained an own assessment of correlation which is reported in Table 2.5. This matrix evidently differs sharply from both the QIS5 one and the one estimated from the time series of loss ratios. In particular, losses in the MTPL LoB appear to be negatively correlated with the other 3 LoBs, which, as happened in terms of loss ratios, show instead a very high degree of positive correlation. It is worth noticing however that these estimates,

computed on a small series of data, can be inaccurate, as the p-values in the Table highlight. Hence, the idea of coupling this information with some more reliable assessments, such as the market-wide one provided by some the CEIOPS seems particularly appropriate.
We assume that losses across LoBs follow a multivariate Gaussian distribution. We include uncertainty on the variance/covariance matrix and we assume a prior Inverse Wishart distribution with precision parameter equal to the QIS5 implied variance/covariance matrix[11] and $n+2$ degrees of freedom. As we showed in the previous section, we can then derive easily the posterior distribution of this variance/covariance matrix and sample from it to generate the multivariate Gaussian copula outcomes. This posterior distribution accounts for both the mean prior (the QIS5 matrix) and the estimated variance/covariance matrix of Table 2.5. We then obtained the aggregate distribution of OLLs, using the "vague priors" marginals derived in 2.2.
Table 2.6 reports best estimates and relevant quantiles of predicted OLLs.

	LoB			
	MTPL	**MOC**	**FP**	**TPL**
MTPL	1 (0)	-0.5275 (0.1791)	-0.5389 (0.1682)	-0.3530 (0.3910)
MOC	-0.5275 (0.1791)	1 (0)	0.9728 (0.000001)	0.8945 (0.0027)
FP	-0.5389 (0.1682)	0.9728 (0.000001)	1 (0)	0.8560 (0.0067)
TPL	-0.3530 (0.3910)	0.8945 (0.0027)	0.8560 (0.0067)	1 (0)

Table 2.5: Linear correlation between LoBs estimated from a time series of losses.Brackets report p-values.

We first compare standard Gaussian copula results when using this new estimated matrix with those obtained using the previously reported assessments of correlation. The distribution of joint OLLs obtained using a standard Gaussian copula with the matrix showed in 2.5 as a parameter is leptokurtic with respect to the one obtained assuming an Independence copula. The $VaR_{99.5\%}$, in particular, is 7 millions lower (649 vs. 656 millions of euros).
The Bayesian Gaussian copula approach, instead, "mixing" between the use of this own assessment of correlation and the QIS matrix reported in Table

[11]We get this precision matrix by transforming the correlation matrix using the estimated sample variance

Figures	**Copula Type**		
	Independence	**Gaussian MLE**	**Bayesian Gaussian**
Mean	593012692.6	593072251.84	593108913.85
Std Dev	23661871.8	21099212.30	21859435.62
VaR 75	609074201.8	607307094.19	607782835.70
VaR 97.5	640745885.5	635540055.89	637227064.46
VaR 99	649693108.9	643711062.03	646107459.86
VaR 99.5	656323827.6	649551647.17	652061589.16

Table 2.6: This Table reports the simulated joint distribution of the OLLs for an Independence copula, a Gaussian copula whose MLE of the correlation matrix is estimated from the time series of losses as described in Section 2.5.3 and for the Bayesian Gaussian copula.

2.2, presents a less sharp diversification effect (its $VaR_{99.5\%}$ is 652 millions of euros). Figures 2.1 and 2.2 compare the densities of the predictive distribution of OLLs obtained from the standard Gaussian copula with the estimated and the QIS correlation matrix and the one obtained with the Bayesian Gaussian technique. They clearly show that this latter distribution is "intermediate" between the other two. Its standard deviation increases with respect to the one reported in the Gaussian MLE, as reported in Table 2.2. The distribution obtained using a Gaussian copula with the QIS 5 correlation matrix as the parameter presents, as expected, fatter tails with respect to the Bayesian Gaussian one. This is due to the different dependence structure and mainly to the negative correlation between the biggest LoB in terms of volume (the MTPL one) and the other ones.

2.6 Conclusions

In this paper we proposed a way to couple Bayesian methods and copulas for stochastic claims reserving method. We showed how to account for expert judgement through Bayesian techniques not only in the estimation of the marginal distribution of losses, but also in the process of aggregating these estimates across multiple Lines of Business.

We made use of copula functions, which allowed us to treat separately the marginals and the dependence structure. We examined how to introduce uncertainty on copula parameters. In particular - due to their analytical tractability - we focused on Bayesian Gaussian copulas. We presented an application of the methodology to a large multi-line Italian insurance company and we compared the results obtained with standard copula aggregation - under different assumptions on the copula type - and the Bayesian copula

model.
Unfortunately, at the moment, lack of enough statistically reliable data mine the possibility of backtesting the model and the performance of goodness of fit tests for establishing best fit copulas. However, we plan to explore these aspects in the future. Extension of the Bayesian approach to other copula functions (t or Archimedean) can also be addressed.

2.7 Appendix - Generating n-dimensional copulas

2.7.1 The Genest and Rivest approach

In this Appendix we present Genest and Rivest's approach to the simulation of n-dimensional Archimedean copulas with one parameter. This method naturally encompasses the special case in which all the bivariate copulas present the same parameter. The objective is to generate variates from the distribution function

$$F(x_1, x_2, ..., x_n) = C(F_1(x_1), F_2(x_2), ...F_n(x_n))$$

where C is a copula function. If C is Archimedean, it admits this representation:

$$C(u_1, u_2, ..., u_n) = \phi^{-1}(\phi(u_1) + ... + \phi(u_n)),$$

where ϕ is called the *generator* of the Archimedean copula. Genest (1987), Genest and Rivest (1993) and Lee (1993) showed how to generate full distributions by recursively simulating conditional ones. Assume a joint probability density function of a multivariate distribution $(X_1, X_2, ..., X_n)$ exists. Then, for $i = 2, ..., n$ the density function of $X_1, ..., X_i$ can be written as:

$$f_i(x_1, x_2, ..., x_i) = \frac{\partial^i}{\partial x_1 ... \partial x_i} \phi^{-1}\{(\phi(F_1(x_1)) + ... + \phi(F_i(x_i))\} =$$

$$= \phi^{-1(i)}\{(\phi(F_1(x_1)) + ... + \phi(F_i(x_i))\} \prod_{j=1}^{i} \phi^1 [F_j(x_j)] F_j^{(1)}(x_j)$$

where the superscript (j) denotes the j-th partial derivative. Hence, we can compute the conditional density of X_i given $X_1, ..., X_{i-1}$

$$F_i(x_i|x_1, ..., x_{i-1}) = \frac{f_i(x_1, ..., x_i)}{f_{i-1}(x_1, ..., x_{i-1})} =$$

$$= \phi^{(1)}[F_i(x_i)]F^{(1)}(x_i)\frac{\phi^{-1(i-1)}\{(\phi(F_1(x_1)) + ... + \phi(F_i(x_i))\}}{\phi^{-1(i-1)}\{(\phi(F_1(x_1)) + ... + \phi(F_{i-1}(x_{i-1}))\}}$$

Then, we obtain the conditional distribution function of X_i given $X_1, ..., X_{i-1}$:

$$F_i(x_i|x_1, ..., x_{i-1}) = \int_{-\infty}^{x_i} f_i(x|x_1, ..., x_{i-1})dx =$$

$$= \frac{\phi^{-1(i-1)}\{(c_{i-1} + \phi(F_i(x_i))\}}{\phi^{-1(i-1)}(c_{i-1}}$$

where $c_i = \phi[F_1(x_1)] + ... + \phi[F_i(x_i)]$.

Starting from these considerations, we present the algorithm that Lee (1993)proposed to generate outcomes from an n-dimensional Archimedean copula:

1. Generate n independent uniform random numbers $U_i \sim U[(0,1)]$ for $i = 1, ...n$
2. Set $X_1 = F_1^{-1}(U_1), c_0 = 0$.
3. Calculate X_i for $i = 2, ...n$ recursively exploiting the properties of the conditional distribution:

$$U_i = F_i(X_i|x_1, ..., x_{i-1}) = \frac{\phi^{-1(i-1)}\{(c_{i-1} + \phi(F_i(x_i))\}}{\phi^{-1(i-1)}(c_{i-1})}$$

2.7.2 Generating 4-dimensional Archimedean copulas

In the following subsections we report the algorithms to generate the 4-dimensional copulas we used in the paper: the Clayton, the Frank and the Gumbel ones. Please notice that closed form solutions can be obtained for the former, while for the Frank and the Gumbel copula, numerical methods are necessary to recover the variates. Throughout the section, θ_1, θ_2 and θ_3 refer to the corresponding bivariate copula parameter between dimension 1 and, respectively, dimensions 2, 3 and 4, while $(U_1, ...U_4)$ refers to a 4-dimensional vector where where each of the U_i's, $i = 1, .., 4$ generated independently from a uniform distribution with values in (0,1). For the copulas we considered, these parameters can be obtained simply from their relationship with the sample Kendall's τ:

1. Clayton: $\theta = \frac{2\tau}{1-\tau}$
2. Gumbel: $\theta = \frac{1}{1-\tau}$

3. Frank: $\tau = 1 - \frac{4}{\theta}[D_1(-\theta) - 1]$, where D denotes the Debye function of order 1.

Notice that the parameters for Frank's copula are obtained numerically finding the zero of the equation that links them to Kendall's τ.

4 dimensional Clayton copula

First, we recall the generator of the Clayton copula:

$$\phi(t) = t^{-\theta} - 1$$

We then compute its inverse and its first three derivatives:

$$\phi^{-1}(s) = (1+s)^{-\frac{1}{\theta}}$$

$$\phi^{-1(1)}(s) = -\frac{1}{\theta}(1+s)^{-\frac{1}{\theta}-1}$$
$$\phi^{-1(2)}(s) = -\frac{1}{\theta}(-\frac{1}{\theta}-1)(1+s)^{-\frac{1}{\theta}-2}$$
$$\phi^{-1(3)}(s) = -\frac{1}{\theta}(-\frac{1}{\theta}-1)(-\frac{1}{\theta}-2)(1+s)^{-\frac{1}{\theta}-3}$$

Then, after having drawn $(U_1, ..., U_4)$, we compute the random variates from the Clayton copula in the following way:

$$X_1 = F_1^{-1}(U_1)$$

$$X_2 = F_2^{-1}\left[\sqrt[\theta_1]{\frac{1}{\left[\sqrt[\frac{1}{\theta_1}+1]{\frac{1}{U_2}} - 1\right] F_1(x_1)^{-\theta_1} + 1}}\right]$$

$$X_3 = F_2^{-1}\left[\frac{1}{\left\{\left[\sqrt[\frac{1}{\theta_2}+2]{\frac{1}{U_3}} - 1\right][F_1(x_1)^{-\theta_2} + F_2(x_2)^{-\theta_2} - 1] + 1\right\}^{\frac{1}{\theta_2}}}\right]$$

$$X_4 = F_4^{-1}\left[\sqrt[\theta_3]{\frac{1}{\left[\sqrt[\frac{1}{\theta_3}+3]{\frac{1}{U_4}} - 1\right][F_1(x_1)^{-\theta_2} + F_2(x_2)^{-\theta_2} + F_3(x_3)^{-\theta_3} - 2] + 1}}\right]$$

As we remarked above, the form of the Clayton copula permits us to find these analytical expressions for (X_1, X_2, X_3, X_4) generated with the Genest and Rivers' approach.

4-dimensional Frank copula

Frank's copula generation is somewhat more difficult than the generation of the Clayton copula. We recall the copula generator, its inverse and its derivatives up to the third order:

$$\phi(t) = -\ln \frac{e^{-\alpha t} - 1}{e^{-\alpha} - 1}$$

$$\phi^{-1}(s) = \frac{1}{\alpha} \ln\left[1 + e^s(e^\alpha - 1)\right]$$

$$\phi^{-1(1)}(s) = -\frac{1}{\alpha}\frac{1}{1 + e^s(e^\alpha - 1)}(e^\alpha - 1)e^s$$

$$\phi^{-1(2)}(s) = \frac{1}{\alpha}\frac{1}{(1 + e^s(e^\alpha - 1))^2}(e^\alpha - 1)e^s$$

$$\phi^{-1(3)}(s) = \frac{1}{\alpha}\frac{1 - e^s(e^\alpha - 1)}{(1 + e^s(e^\alpha - 1))^3}(e^\alpha - 1)e^s$$

From the expressions we derive below, it is easy to see that the high non-linearity of the equations does not allow us to obtain closed-form solutions, contrary to the Clayton copula:

$$X_1 = F_1^{-1}(U_1)$$

$$U_2 = \frac{e^{\theta_1 F_1(x_1)}\left[e^{\theta_1 F_1(x_1)} - 1\right]}{e^{\theta_1} - 1 + \left[e^{\theta_1 F_1(x_1)} - 1\right]\left[e^{\theta_1 F_2(x_2)} - 1\right]}$$

$$U_3 = \frac{\left[e^{\theta_2 F_3(x_3)} - 1\right]\left[e^{\theta_2} - 1\right]\left\{e^{\theta_2} - 1 + \left[e^{\theta_2 F_1(x_1)} - 1\right]\left[e^{\theta_2 F_2(x_2)} - 1\right]\right\}^2}{\left\{\left[e^{\theta_2 - 1}\right]^2 + \left[e^{\theta_2 F_1(x_1)} - 1\right]\left[e^{\theta_2 F_2(x_2)} - 1\right]\left[e^{\theta_2 F_3(x_3)} - 1\right]\right\}^2}$$

$$U_4 = \frac{\left[e^{\theta_3 F_4(x_4)} - 1\right]\left[e^{\theta_3} - 1\right]\left\{\left[e^{\theta_3} - 1\right]^3 - P\right\}\left\{\left[e^{\theta_3} - 1\right]^2 + Q\right\}^3}{\left\{\left[e^{\theta_3} - 1\right]^3 + P\right\}^3\left\{\left[e^{\theta_3} - 1\right]^2 - Q\right\}}$$

where

$$P = \prod_{i=1}^{4} e^{\theta_3 F_i(x_i)} - 1$$

and

$$Q = \prod_{i=1}^{3} e^{\theta_3 F_i(x_i)} - 1$$

We find the zeros of the above equations to obtain (X_1, X_2, X_3, X_4).

4 dimensional Gumbel copula

As for the Frank copula, the high non-linearity of the relationship between U_i's and the X_i's do not permit to obtain analytical expressions. Resorting to numerical methods, anyway, one can easily solve the equations derived from the properties of conditional distributions and generate variates from a 4 dimensional Gumbel copula. First, we recall the generator of the bivariate Gumbel copula with parameter θ, its inverse and the derivative of the latter up to the third order:

$$\begin{aligned}
\phi(t) &= (-\ln t)^{\theta} \\
\phi^{-1}(s) &= e^{-t^{\frac{1}{\alpha}}} \\
\phi^{-1(1)}(s) &= -\frac{1}{\alpha} t^{\frac{1}{\alpha}-1} e^{-t^{\frac{1}{\alpha}}} \\
\phi^{-1(2)}(s) &= -\frac{1}{\alpha^2} t^{\frac{1}{\alpha}-2} e^{-t^{\frac{1}{\alpha}}} (1-\alpha-t^{\frac{1}{\alpha}}) \\
\phi^{-1(3)}(s) &= -\frac{1}{\alpha^3} t^{\frac{1}{\alpha}-3} e^{-t^{\frac{1}{\alpha}}} (1-3\alpha+(2\alpha-3)t^{\frac{1}{\alpha}}+2t^{\frac{2}{\alpha}})
\end{aligned}$$

Applying Genest and River' s procedure we get:

$$X_1 = F_1^{-1}(U_1)$$

$$U_2 = \frac{e^{-[\phi(F_1(x_1))+\phi(F_2(x_2))]^{\frac{1}{\theta_1}}} [\phi(F_1(x_1))+\phi(F_2(x_2))]^{\frac{1}{\theta_1}-1}}{e^{-\phi(F_1(x_1))^{\frac{1}{\theta_1}}} \phi(F_1(x_1))^{\frac{1}{\theta_1}-1}}$$

$$U_3 = \frac{e^{-[P+\phi(F_3(x_3))]^{\frac{1}{\theta_2}}} [P+\phi(F_3(x_3))]^{\frac{1}{\theta_2}-2} \left[1-\theta_2-(P+\phi(F_3(x_3)))^{\frac{1}{\theta_2}}\right]}{e^{-P^{\frac{1}{\theta_2}}} P^{\frac{1}{\theta_2}-2} \left[1-\theta_2-P^{\frac{1}{\theta_2}}\right]}$$

$$U_4 = \frac{e^{-[Q+\phi(F_4(x_4))]^{\frac{1}{\theta_3}}} [Q+\phi(F_4(x_4))]^{\frac{1}{\theta_3}-3} \left[2\theta_3^2-3\theta_3+1+(2\theta_3-3)(Q+\phi(F_4(x_4)))^{\frac{1}{\theta_3}}+2(Q+\phi(F_4(x_4)))^{\frac{2}{\theta_3}}\right]}{e^{-Q^{\frac{1}{\theta_3}}} Q^{\frac{1}{\theta_3}-3} \left[2\theta_3^2-3\theta_3+1+(2\theta_3-3)Q^{\frac{1}{\theta_3}}+2Q^{\frac{2}{\theta_3}}\right]}$$

where

$$P = \sum_{i=1}^{2} \phi(F_i(x_i))$$

and

$$Q = \prod_{i=1}^{3} \phi(F_i(x_i))$$

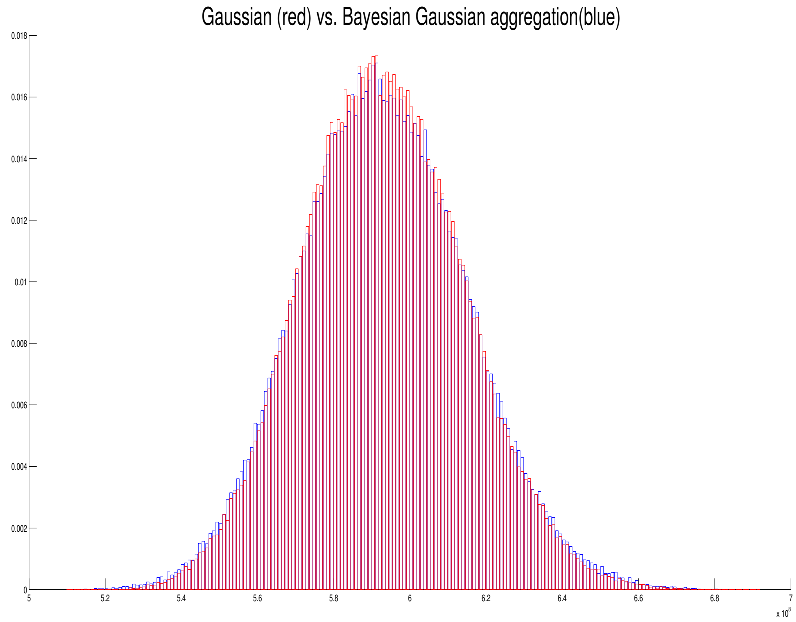

Figure 2.1: This Figure shows the density of the predictive distribution of OLLs obtained using a Gaussian copula with correlation matrix estimated from company data on losses (red) and using the Bayesian Gaussian model (blue).

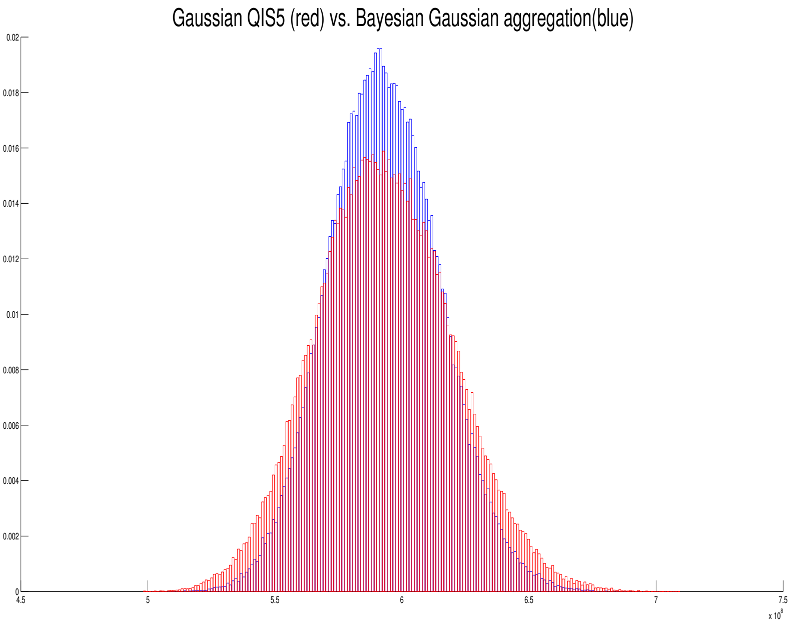

Figure 2.2: This Figure shows the density of the predictive distribution of OLLs obtained using a Gaussian copula with the correlation matrix given by the CEIOPS (red) and using a Bayesian Gaussian model (blue).

Part II

A theoretical model of Capital Structure for Financial Conglomerates

In this second part of my Dissertation I propose a theoretical model that studies optimal capital and organizational structure choices of financial groups which incorporate two or more business units. The group faces a VaR-type regulatory capital requirement. Financial conglomerates incorporate activities in different sectors either into a unique integrated entity, into legally separated divisions or in ownership-linked holding company/subsidiary structures. I model these different arrangements in a structural framework through different coinsurance links between units in the form of conditional guarantees issued by equityholders of a firm towards the debtholders of a unit of the same group. I study the effects of the use of such guarantees on optimal capital structural and organizational form choices. I calibrate model parameters to observed financial institutions' characteristics. I study how the capital is optimally held, the costs and benefits of limiting undercapitalization in some units and I address the issues of diversification at the holding's level and regulatory capital arbitrage.

Chapter 3

Optimal Capital Structure and Organizational Choices in Financial Conglomerates

3.1 Introduction

Conglomeration in the financial sector has gained growing importance since the beginning of the Nineties. As soon as the legal barriers which prevented firms involved in different financial sectors from merging disappeared almost everywhere, institutions of large size and involved in different sectors started to merge. Even the United States, that since the Thirties imposed separation between the three sectors, favoured this phenomen through the adoption of the Gramm-Leach-Bliley Act in 1998[1], which allowed banks, insurance companies and securities to merge. The consequence was the rise of financial conglomerates[2], which are now widespread and represent a large part of the actors in the market. The possible diversification benefits arising from the combination of non-perfectly correlated activities seem to provide *per se* a first rationale for their success[3]. Diversification indeed should reduce the probability of entering financial distress - as a consequence of lower cash flow volatility - and the need for external, more costly, financing. Another

[1]For a detailed description of the Act and its consequences on the banking sector we refer the reader to Macey (1999).

[2]Financial conglomerates are formally defined(Dierick (2004)), as financial institutions that incorporate activities in at least two out of the three financial sectors (banking, securities and insurance).

[3]See, for instance, Dierick (2004) for a comprehensive review of financial conglomerates' main characteristics.

potential benefit of conglomeration is the creation of synergies in costs and revenues[4]. However, the extent to which these economies of scale and scope can effectively be realized is still an open issue in empirical studies[5].
What is not arguable is that the trend towards financial conglomeration has posed new challenges to the regulators. Coordination between supervisors of the different sectors at the national level, besides international harmonization of the rules in place, has become a key issue. Moreover, the recent financial crisis brought to anyone's attention that the interconnection between markets, the increasing concentration, the high leverage[6] and the existence of *too big to fail* or *systemic important* entities pose a serious threat to the whole economic system. Regulators are trying to adapt the regulatory framework and mitigate these problems. At the end of 2009 the Solvency II Directive on the capital adequacy of insurance companies has been approved and will become effective in the next few years and the Basel committee is already working on a revised set of rules for the banking industry.
The aim of this paper is to investigate theoretically the capital structure decisions of financial firms subject to capital requirements and how they differ across the different organizational structures financial conglomerates can adopt. We do this in the context of a two-period static framework model. Firms face a unique source of risk, operating cash flow volatility. We consider frictions in the market in the form of proportional taxes and default costs. The units are financed through debt or equity. Since debt is tax advantaged, a trade-off between tax savings through leverage and default costs emerges. Firms maximize their value through the choice of principal debt and have to meet an equity capital requirement. We reproduce the different arrangements financial conglomerates can take through the presence of different coinsurance links, in the form of guarantees of rescue in case of default, between divisions.
First, we focus on the optimal behavior of the different arrangements a financial group can assume and compare their properties in terms of value, debt capacity and default costs. We characterize analytically the case in which the firm does not lever up.
Second, we address the issue of the optimal allocation of equity capital among

[4]Vennet (2002) carried out an empirical analysis of cost and revenue synergies of European conglomerates, concluding that revenue efficiency is improved after conglomeration, but there is no evidence of scale effects on the cost side.

[5]See for instance Laeven and Levine (2007), Schmid and Walter (2009) and Van Lelyveld and Knot (2009).

[6]Hellwig (2009) reports that institutions like Deutsche Bank and UBS had narrowed their equity capital buffer down, reaching a level between 2 and 3% at the time of the subprime crisis.

divisions and of how the capital requirement affects it. We find that when consolidated supervision is in place - i.e. capital adequacy is required at the group's level - an uneven distribution of equity capital among divisions optimally arise. In this case, financial conglomerates isolate the risk of default in one thinly capitalized unit, which exploits tax savings through leverage and bears high default costs. This optimal arrangement is also the one that maximizes default costs, which are a deadweight cost to the economy. Solo supervision - i.e. capital adequacy is prescribed at the single entities' level - avoids this problem and can make holding company/subsidiary (HS) structures emerge as both the privately (value maximizing) and the welfare (default costs minimizing) optimal arrangements.
Third, we address the issue of capital arbitrage when units in the same group are subject to different regulatory constraints. We give a characterization of the optimal arrangement under such an asymmetry.
The paper is structured as follows: Section 3.2 briefly reviews the related literature on financial conglomerates and their modelling, section 3.3 describes how financial groups are legally organized and supervised in the real world. Section 3.4 presents the model and some analytical results. Section 3.5 and Section 3.6 provide numerical analysis for different organizational structures. Section 3.7 concludes.

3.2 Related Literature

Recently, some papers focused on the optimal organizational choice of financial conglomerates. Dell'Ariccia and Marquez (2010) analyzed the trade-off between branches or subsidiaries structures for cross-border banking groups facing macroeconomic risks and exogenous capital constraints. Harr and Rønde (2003) studied a similar problem for multinational banks facing capital requirements, moral hazard and adverse selection problems. Gatzert and Schmeiser (2008) analyzed the different arrangements financial conglomerates can take and focused on the diversification benefits exploitation and shareholder value, under an exogenous capital structure. We contribute to this stream of theoretical literature by studying simultaneously the endogenously optimal organizational and financing choices of financial conglomerates subject to capital requirements.
The literature has also devoted a strong effort to study the interplay between capital regulation and the risk-taking incentives of banks. Freixas, Lóránth, and Morrison (2007) focus on endogenous risk taking choices of banks under asymmetric information. They argue that socially excessive risks are undertaken by financial institutions in the absence of capital requirements. In this

paper, we give theoretical support to their hypothesis, describing the interactions between financing constraints and firms' optimal decisions and their impact on endogenously determined default costs, default probabilities and recovery rates.
We draw our modelling set up from Leland (2007) and Luciano and Nicodano (2010). They both study the optimal capital structure choices of unconstrained commercial firms in a structural model. We adapt their framework to regulated firms by introducing financing constraints. We also perform some numerical analysis and we calibrate the model to the observed features of financial institutions. By assuming perfect information and no agency costs, we depart from Kahn and Winton (2004) and Freixas, Lóránth, and Morrison (2007), who study the problem of interdivisional optimal capital allocation for financial institutions under different market frictions, related to informational asymmetries.
A number of papers focused on the creation of coinsurance opportunities through conglomeration or capital and risk transfer instruments. Lewellen (1971) first argued that imperfect correlation between units can increase debt capacity. Flannery, Houston, and Venkataraman (1993) extended this result analyzing joint vs. separate incorporation in a model with corporate taxation and trade-off between the tax advantage and agency costs - due to underinvestment - of debt. Leland (2007) carefully explored the purely financial synergies of merging two - possibly asymmetric[7] - units, finding that diversification between cash flows and pooling (i.e. risk sharing among units) allows the firm to increase debt capacity and value relative to comparable stand alone companies, if divisions' characteristics are not too different. When this is the case, instead, conglomerates are penalized by their reduced capital structure flexibility. We analyze the effects of cash flow pooling - which is implicitly an unconditional rescue guarantee among units - in financial conglomerates. We explore different ways of recognizing diversification benefits when computing the capital requirement. We find that joint incorporation can be sub-optimal also when units have the same characteristics and the same cash flow distribution.
Following Luciano and Nicodano (2010) we model holding company/subsidiary structures as issuers of binding unilateral or mutual support commitments, conditional on the ability of the guarantor unit to cover both the obligations of the insolvent beneficiary and its own ones without experience default. The existence of such formal or informal guarantees is supported by previous

[7]Throughout the paper, the term "asymmetric" units will refer to divisions whose characteristics differ for at least one parameter. It does not mean that the distributions of their cash flows are asymmetric in a statistical sense.

studies (Boot, Greenbaum, and Thakor (1993), Khanna and Palepu, 2000). Luciano and Nicodano (2010) show that these conditional guarantees are always optimal with respect to the no guarantee case and that the units of a group always specialize either as a guarantor or as a beneficiary. In this paper we find that, when units are subject to capital requirements, mutual guarantees are in general more effective than unilateral ones: in most cases, each unit of a "regulated" holding/subsidiary structure both provides and receives support optimally.

3.3 Financial Groups and their supervision

3.3.1 Financial conglomerates and coinsurance

In this section we briefly discuss how financial conglomerates set up and how they are supervised.

Dierick (2004) surveys the corporate structures of financial conglomerates and identifies four main models:

1. horizontal groups (HG), in which no direct link exists between the different units,
2. integrated conglomerates (IC), which operate as a unique company,
3. holding company-subsidiary structures (HS), in which units are legally and operationally separated, but are connected through ownership or control links,
4. holding company models (HCM), in which a non-operating entity[8] controls one or more operating subsidiaries[9].

Stand alone companies (SA) which are not linked through any control structure constitute HGs[10]. They do not exploit any coinsurance opportunity, but they can take advantage of full capital structure flexibility. Integrated conglomerates (IC) set up a unique entity which merges the cash flows of its

[8]On top of the structure lies a Non Operating Holding Company (NOHC), which is paradoxically sometimes less strictly regulated than standard holdings.

[9]Recently, some OECD economists (Blundell-Wignall et al., 2009), hinted at NOHC as a possible instrument to avoid systemic and counterparty risk in a crisis. Anyway, they assume no coinsurance and capital transfer allowed between the units. So, what they have in mind is equivalent to horizontal groups, with the on-top company providing economies of scale and scope through centralized IT and back-office functions.

[10]The structure of an HG consists usually also of an "umbrella" corporation on top of it, which has the purpose of establishing their common ownership.

units and files a unique balance sheet. Hence, divisions provide unconditional coinsurance for each others' debt repayment. Coinsurance in HS structures is provided by the possibility of risk and capital transfers between the holding and its subsidiaries.

3.3.2 The supervision of financial groups

A fundamental difference between financial and non-financial companies is that, given the key role they play in the economy, the former are subject to more stringent capital standards. In particular, after the first Basel accord the focus of supervision of financial groups has primarily concerned the avoidance of double gearing - i.e. using the same capital to hedge against different risks[11] - and the development of internal models that could link capital standards to the risk characteristics of firms. After the recent financial crisis, the focus of regulators seem to have shifted towards the avoidance of excessive leverage and the limitation of systemic risk. Nonetheless, the growing importance of conglomeration[12] and bancassurance recently led to focus on cross-sectoral regulation. In particular, two main issues seem to be particularly important: how to account for within and between sector diversification at the holding level[13] and how to prevent regulatory capital arbitrage. Firms can in fact exploit diversification benefits and asymmetries in the supervisory rules at their own advantage[14].
According to the Joint Forum document 2010, there are two possible approaches to the supervision of a group: accounting consolidation, which treats the group as a single entity, or risk-based aggregation, according to which capital requirements apply to the single entities and are then aggregated in order to give a group-wide figure. Even if under Basel II requirements have to be met at a consolidated level, insurance undertakings in a financial group have to be treated separately from their banking affiliates. Hence, both approaches coexist in the Basel/Solvency framework: when firms in the group belong to the same sector a consolidated approach is followed, while capital requirements are computed separately for affiliates involved in different in-

[11] In particular, from an economic capital perspective, this means using the same capital to hedge against risks in two different entities of the group.

[12] The FSA agency in 2007 recognized that, for example, the four largest Japanese banks are financial conglomerates, combining undertakings in each sector.

[13] EU regulators, prescribing de-consolidation between banking and insurance undertakings belonging to the same group, decided to rule out inter-sectoral diversification benefits, while they can be taken into account in internal models when units belong to the same sector.

[14] In particular, banking and securities on one side and insurance on the other are subject to different capital adequacy standards.

dustries. This situation can provide incentive to shift assets from banking to insurance undertakings and viceversa when capital arbitrage is possible.
Recently, during the debate on the Solvency II framework, an interesting Group Support Regime framework has been proposed and highly discussed. It intended to legally bind holding companies to intervene to meet their subsidiaries' solvency capital requirements. Its conceptual innovations lied in stressing the role of intergroup guarantees and capital transfers and in the explicit regulation of capital fungibility within the units. In particular, the proposal prescribed that each subsidiary should hold enough capital to meet its MCR, but the capital buffer between the MCR and the SCR could be covered through transfers from the holding company, provided that the overall group SCR, equal to the sum of the solo SCRs, was reached. Thus, the proposal explicitly tackled two of the problems we will discuss in the following sections. First, it somehow gave guidelines on where to hold the capital inside a group. Second, it highlighted the possibility of using intra-group guarantees as regulatory instruments. However, the Group Support Regime was finally dropped in the final draft of the Directive, in favour of a consolidated approach already used in the Basel II set up. In the paper, we analyze theoretically the potential benefits and shortcomings of adopting both these approaches.

3.4 Modelling financial conglomerates

In the following sections we model the different arrangements can take.
An entrepreneur wants to set up a financial business made up of N units at time 0. The divisions can either be incorporated into legally separated - stand alone - entities which create an horizontal group, into a unique entity, the integrated conglomerate, or into a control-linked holding/subsidiary structure.

3.4.1 Basic set up: the stand alone firm and the HG

We describe our basic model for stand alone firms (SA). This section reviews Leland (2007)'s model, establish some of its properties and describes how we adapt it to financial firms through the introduction of a capital requirement.
An entrepreneur sets up a SA (which we denote with the subscript i) which has a future exogenous random operating cash flow X_i over an horizon of T years. The problem he faces concerns how to finance this future cash flow at time 0, being subject to proportional taxes and dissipative default costs. The institution can choose between two types of financing instruments: debt and

equity. Debt is modelled as a zero-coupon risky bond with maturity T[15]. The entrepreneur chooses the nominal value P_i the firm issues and it has to repay at time T. Debt has a fiscal advantage, since interests are tax-deductible. We define X_i^Z as the tax shield, i.e. the level of operating cash flows under which the firm pays no taxes, since interest deductions offset them[16]:

$$X_i^Z(P_i) = P_i - D_{0i} \tag{3.1}$$

where $D_{0i}(P_i)$ represents the value of debt at time 0.

Due to the presence of the tax shield, a firm can increase value through the use of leverage. The expected value of tax savings is

$$TS_{0i}(P_i) = \tau_i \phi \mathbb{E}\left[X_i 1_{\{X_i < X_i^Z\}} + X_i^Z 1_{\{X_i \geq X_i^Z\}}\right] \tag{3.2}$$

where τ_i is the proportional effective tax rate, $\phi = \dfrac{1}{(1+r)^T}$ is the discount factor and r is the exogenous risk-free interest rate in the economy and $\mathbb{E}[\cdot]$ denotes the mathematical expectation operator.

Equity is a residual claim and represents the own funds the firm has.

Debtholders and equityholders evaluate their claims on the firm at time 0 in a risk-neutral environment as the discounted present value of their expected future payoffs. At time T, cash flows are realized and proceeds are distributed to the stakeholders of the firm. The firm enjoys limited liability. The tax authority has absolute priority and it is repaid first. Debt claimants receive income net of taxes up to the principal P_i, while equityholders are entitled to proceeds only if the firm is solvent and receive cash flows after taxes and debt repayment. If cash flows realized at time T are not sufficient to meet in full the debt obligation to bondholders the firm is declared insolvent.

Leland (2007) defines X_i^D as the distress threshold, i.e. the level of cash flows necessary to repay debtholders in full at the end of the horizon:

$$\begin{aligned} (1-\tau_i) X_i + \tau_i X_i^Z &= P_i && (3.3) \\ X_i^D &= \frac{P_i - \tau_i X_i^Z}{1-\tau_i} = P_i + \frac{\tau_i D_{0i}}{1-\tau_i} && (3.4) \end{aligned}$$

[15]Our modelling of debt as a risky long-term zero-coupon bond cannot accurately represent the whole bunch of debt instruments (deposits, policies, bond issues) a financial firm can issue. Anyway, we remark that short-term forms of debt can be thought of as continuously rolled over up to the end of the horizon and we can account for deposit insurance and other investors' protection systems when calibrating model parameters. In the end, we do not seem to lose much by modelling debt parsimoniously as a zero-coupon bond.

[16]Notice that interests are deductions: if $P_i - D_{0i}$ exceeds realized cash flow X_i, no reimbursement is obtained: no carry-forward is allowed.

We can easily notice that X_i^D is always greater than X_i^Z, since $D_{0i} \geq 0,\ 0 \leq \tau_i \leq 1$.

Equityholders receive then

$$E_i = (X_i^n - P_i)^+ \tag{3.5}$$

were X_i^n indicates cash flow net of taxes

$$X_i^n = (1 - \tau_i)\, X_i + \tau_i X_i^Z. \tag{3.6}$$

Debtholders receive

$$D_i = X_i^\alpha 1_{\{0<X_i<X_i^Z\}} + X_i^{n,\alpha} 1_{\{X_i^Z<X_i<X_i^D\}} + P_i 1_{\{X_i>X_i^D\}} \tag{3.7}$$

where X^α is the realized cash flow net of default costs:

$$X_i^\alpha = (1 - \alpha_i)\, X_i \tag{3.8}$$

and $X^{n,\alpha}$ represents realized cash flows net of taxes and default costs:

$$X_i^{n,\alpha} = (1 - \alpha_i)\, X_i - \tau_i (X_i - X_i^Z) = (1 - \alpha_i - \tau_i) X_i + \tau_i X_i^Z \tag{3.9}$$

Default costs are proportional to income: α_i represents the fraction of operating cash flows lost due to the costs associated with bankruptcy.

The present value of these expected costs of default is

$$DC_{0i}(P_i) = \alpha_i \phi \mathbb{E}\left[X 1_{\{X_i<X_i^D\}}\right]. \tag{3.10}$$

These costs represent a deadweight loss to the economy, while expected levied taxes are

$$T_{0i}(P_i) = \tau_i \phi \mathbb{E}\left[(X - X^Z)^+\right]. \tag{3.11}$$

The discounted present values of payoffs to equityholders and debtholders represent the market values of equity and debt respectively:

$$E_{0i}(P_i) = \phi \mathbb{E}\left[(X_i^n - P_i)^+\right] \tag{3.12}$$

$$D_{0i}(P_i) = \phi \mathbb{E}\left[X_i^\alpha 1_{\{0<X_i<X_i^Z\}} + X_i^{n,\alpha} 1_{\{X_i^Z<X_i<X_i^D\}} + P_i 1_{\{X_i>X_i^D\}}\right] \tag{3.13}$$

We can establish some properties of equity and debt values for a stand alone firm. In particular, the following properties hold:

Proposition 3.4.1 *1. The value of debt $D_0(P)$ is concave in P and its derivative with respect to P is bounded above by 1 and below by $-\frac{1-\tau}{\tau}$.*

2. The value of equity $E_0(P)$ is decreasing and convex in P.

Proof. See Appendix A. ■ Being in the context of a structural model, we remark that default probabilities, DP_i, and recovery rates, RR_i (conditional on default) are endogenous:

$$DP_i(P_i) = \Pr[X_i < X_i^d] \tag{3.14}$$

$$RR_i(P_i) = \frac{\mathbb{E}\left[X_i^{\alpha} 1_{\{0<X_i<X_i^Z\}} + X_i^{n,\alpha} 1_{\{X_i^Z<X_i<X_i^P\}}\right]}{P_i \ \Pr[X_i \leq X_i^d]} \tag{3.15}$$

Total value V_{0i} is obtained summing up debt and equity market value at time 0 (D_{0i} and E_{0i}, respectively).

Capital requirement for a SA

Firms have to fulfill a minimum own-funds (equity) requirement at time 0. We model it as a Value-at-Risk bound evaluated on operating cash flows - which, being the only random variable[17], captures the whole uncertainty the company faces. Firms that do not match this equity standard at time 0 are not allowed to exercise their activity and hence cannot set up. Such a capital requirement is in the spirit of the Solvency/Basle II Solvency Capital Requirement (SCR) but, at the same time, it is also a Minimum Capital Requirement (MCR), since firms are not allowed to set up if they do not meet it.
We define the distribution of future losses as $L_i = -X_i$ and we formally introduce the capital adequacy constraint as a Value at Risk on their distribution at a confidence level β_i at a certain time horizon:

$$E_{0i}(P_i) \geq \max(0, VaR_{\beta_i}(L_i)) \tag{3.16}$$

[17]Since the distribution of cash flows is known, we depart from the stream of literature that focused on the asymmetric information problems related to the disclaim of financial institutions' risk to the regulators. Here, we assume that the capital requirement is computed in an internal-model fashion and, as it happens in reality, that the information disclaim, calibration and validation of the model is subject to the approval of the regulator. Hence, both the company and the regulator have of full knowledge of this distribution.

Value maximization

The entrepreneur chooses the amount of principal debt P_{0i} to issue in order to maximize firm value[18]. Hence, optimal capital structure is determined as the solution of the following program:

$$\max_{P_i} V_{0i}(P_i) = \max_{P_i} [E_{0i}(P_i) + D_{0i}(P_i)] = V_{0i}(P_i^*) \tag{3.17}$$

$$s.t.\ \ E_{0i}(P_i) \geq \max(0, VaR_{\beta_i}(L_i)) \tag{3.18}$$

$$P_i \geq 0 \tag{3.19}$$

Numerical solution is required, since (3.13) is an implicit equation (X_i^Z and X_i^D both depend on D_{0i}).
The VaR constraint counterbalances the risk-taking incentive provided by limited liability[19]. The higher the cash flow volatility, the higher the level of prudential equity capital required. The introduction of a VaR-type constraint, then, results in an important departure from Leland's model where, due to limited liability, optimal firm value is convex in cash flow volatility and hence increasing from a certain level onwards.
We can rewrite the objective function highlighting the trade-off between tax savings and default costs. Problem 3.17 is equivalent either to maximize the sum of unlevered value and tax savings, net of default costs or to minimize the sum of tax burdens and default costs:

$$V_{0i}(P_i^*) = \max_{P_i} [V_{0i}(0) + TS_{0i}(P_i) - DC_{0i}(P_i)] = \tag{3.20}$$

$$= \min_{P_i} [T_{0i}(P_i) + DC_{0i}(P_i)] \tag{3.21}$$

Obviously, the constraint is binding only when the optimal unconstrained solution implies a level of equity which lies below the requirement. When this is not the case, the firm is unaffected by regulation, since it finds it optimal to hold more capital than what is prescribed and the unconstrained solution coincides with the constrained one.
We now establish some properties of the constrained maximization program's solutions:

[18]The entrepreneur maximizes this sum because it represents the amount of money he cashes in at time 0 as a transfer received by the debtholders (D_{0i}) and as the present value of his equity share in the firm (E_{0i}), which is the amount he would earn from selling the stocks.

[19]Increasing cash flow volatility allows to exploit the asymmetry induced by limited liability, which shields the entrepreneur from downside losses.

Proposition 3.4.2 *There exists no unlevered ($P^* = 0$) solution in which the constraint is non-binding when $\tau > 0$. When the constraint is non-binding, feasible solutions satisfy a property on the ratio $\frac{\alpha}{\tau}$.*

Proposition 3.4.3 *The constraint is binding at the optimum if the ratio $\frac{\alpha}{\tau}$ lies below a certain function of the optimal solution $g(P^*)$.*

Proof. See Appendix A. ■

HGs are constituted by SA firms. We simply notice that the absence of any linkage between the units results in completely independent financing decisions. The N units have siloed and not fungible capital. Thus, the total value of the HG is given by the sum of the values obtained by solving the maximization program (3.17) for each division.

3.4.2 Integrated conglomerates

An IC merges N units into one legal entity. Its capital structure is unique: debt and equity are issued on the name of the merger entity rather than on the single divisions' ones. The cash flows realized by each division are used to service the debt raised at the group's level[20]: implicitly, each unit is liable for each others' debt[21] and coinsurance is present in the form of a

[20]Total operating cash flows are given by the sum of the ones realized by each division:

$$X_C = \sum_{i=1}^{N} X_i. \tag{3.22}$$

Hence, for what concerns the first two moments:

$$\begin{aligned}
\mathbb{E}[X_C] &= \sum_{i=1}^{N} \mathbb{E}[X_i], && (3.23)\\
\sigma^2[X_C] &= \sum_{i=1}^{N} \sigma^2\left[X_i\right] + 2\sum_{i=1}^{N}\sum_{i<j} \rho_{i,j}\sigma_i\sigma_j && (3.24)\\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} Cov(X_i, X_j) && (3.25)
\end{aligned}$$

where $\sigma^2[X_i]$ denotes the variance of cash flows for the $i-$th unit and ρ_{ij} the (Pearson's) correlation coefficient between cash flows of units i and j.

[21]Financial literature often use the term "branches" structure to indicate an IC, in which capital structure is unique and units are liable for each others.

mutual unconditional rescue guarantee between divisions. As a consequence, there can then be states of the world in which one or more units would be insolvent if taken individually, but the IC is not, since units with positive net income provide enough cash flows to repay debtholders. On the contrary, there might be divisions which would be solvent individually and are dragged into default by the bad performances of the other units.

The VaR capital requirement consists of a unique constraint, computed either on the joint distribution of losses or as the sum of the constraints computed individually on each unit:

$$V^*_{0C}(P^*_C) = \max_{P_C} V_{0C}(P_C) = \max_{P_C} \left[E_{0C}(P_C) + D_{0C}(P_C)\right] \quad (3.26)$$

$$s.t.\ E_{0C}(P_C) \geq \max\left(0, VaR_{\beta_C}(L_C)\right)\ \ or \quad (3.27)$$

$$E_{0C}(P_C) \geq \sum_{i=1}^{N} \max\left(0, VaR_{\beta_C}(L_i)\right), \quad (3.28)$$

$$P_C \geq 0 \quad (3.29)$$

where P_C, $E_{0C}(P_C)$ and $D_{0C}(P_C)$ denote respectively the principal, the equity and the debt value of the IC.
Notice that the constraint (3.27) accounts for diversification effects across units when computing the constraint, while the (3.28) one does not. The choice of whether to account for these effects in the capital requirement is a highly debated theme. Diversification at the group's level is a real economic phenomenon if it captures diversification benefits which are not already accounted for at the single risk aggregation level. The extent to which this holds true is not clear[22]. Hence, these two constraints model the two opposite extreme situations: recognizing diversification effects according to the correlation between cash flows and not recognizing them at all when computing the capital requirement.
Payoffs to debt and equityholders are derived in the same way as in the stand alone case.

[22]Kuritzkes et al. (2003) try to compute the possible diversification effects within single risk-factors, within business lines and across business units, possibly involved in different sectors, suggesting that incremental benefits from this last diversification type are modest (5%-10%). An HM Treasury (2008) technical report recognizes that group SCR is generally expected to be lower than the group's constituent parts solos.

3.4.3 Holding/Subsidiary structures

In a HS, units are linked together through a control structure, which is exercised through an infinitesimal share holding. Hence, no cash dividend transfers involve the subsidiary and the equityholders of the holding company[23]. Parent companies drive their subsidiaries' financing choices. Each of the N units in the HS can issue an amount of debt principal P_i on its own name and these amounts are decided by the same individual, the entrepreneur who owns the HS.

Group Support - Conditional Guarantees

As we already mentioned, we follow Luciano and Nicodano (2010) and we assume that units are linked by legally enforceable rescue guarantees. From now on, we consider an HS in which there are two units ($N = 2$), that we will simply refer to as a holding company (H) and a subsidiary (S). Guarantees can be either unilateral or mutual. In the first case one of the two firms, the issuer, commits itself to transfer cash to the beneficiary when it is able to avoid its default without experiencing it itself. In the second case, both firms can act as guarantors or beneficiaries.

We formally describe the event of rescue R_H associated to the unilateral conditional guarantee issued by H:

$$R_H(P_H, P_S) = \left\{ \begin{array}{c} X_S < X_S^D \\ (X_H^n - P_H) \geq (P_S - X_S^n) \end{array} \right\} \tag{3.30}$$

We rewrite the second condition as $X_H \geq h(X_S)$, where

$$h(X_S) = \left\{ \begin{array}{ll} \dfrac{P_S + P_H - \tau X_H^Z - X_S}{1-\tau}, & 0 < X_S < X_S^Z \\ \dfrac{P_S + P_H - \tau(X_S^Z + X_H^Z)}{1-\tau} - X_S, & X_S \geq X_S^Z \end{array} \right\}. \tag{3.31}$$

and $h(X_S)$ is then the minimum level of X_H which makes the transfer possible.

[23]This assumption does not mean that holding's shareholders are owners of the whole capital of the subsidiary. We can extend the model including the existence of an ownership share ω, that entitles the holding to subisidiary's dividends. Dividends are transferred to the holding when subsidiary's operating income exceeds its solvency threshold.

The parent company can use these proceeds to repay its debtholders. Hence, dividend transfers provide hedging against some bad states of the world, in which the holding is rescued through dividends.

We obtain a symmetrical expression for the rescue event R_S associated to the case in which S issues a unilateral guarantee:

$$R_S(P_H, P_S) = \left\{ \begin{array}{c} X_H < X_H^D \\ (X_S^n - P_S) \geq (P_H - X_H^n) \end{array} \right\} \tag{3.32}$$

where the second condition can be written as

$$h(X_H) = \left\{ \begin{array}{ll} \dfrac{P_S + P_H - \tau X_S^Z - X_H}{1-\tau}, & 0 < X_H < X_H^Z \\ \dfrac{P_S + P_H - \tau (X_H^Z + X_S^Z)}{1-\tau} - X_H, & X_H \geq X_H^Z \end{array} \right\}. \tag{3.33}$$

Notice that the two events R_H and R_S are disjoint but have both to be considered when the guarantee is mutual.

The presence of the conditional guarantee enhances HS value with respect to equivalently indebted SA companies[24].

We denote as 1 and 2 the SA units which have characteristics (cash flow distribution, default cost, tax rates) equivalent to H and S respectively. Following Luciano and Nicodano (2010), we can decompose the value of the conditional guarantee into two components [25]:

$$G(P_H, P_S) = \underbrace{\Gamma(P_1^*, P_2^*)}_{\text{rescue effect}} + \underbrace{\nu_{0HS}(P_H^*, P_S^*) - \nu_{0HS}(P_1^*, P_2^*)}_{\text{leverage effect}} \tag{3.34}$$

where

$$\Gamma(P_1^*, P_2^*) = \mathbb{E}\left[\alpha_S X_S 1_{\{R_H\}} + \alpha_H X_H 1_{\{R_S\}}\right]$$

is the saving in default costs due to the presence of the guarantee[26]. Hence, the effect of the guarantee is both direct (rescue effect), since it helps in saving from default costs, and indirect (leverage effect), since it allows the firm to lever up more and modify its tax savings and default costs.

Payoffs to debt and equity

We now focus on the payoffs to the stakeholders of the two units at time T . Shareholders receive the difference between net cash flows and nominal debt

[24]This result is independent of the correlation between units, of the density functions and of other parameters. Luciano and Nicodano (2010) obtain a sufficient condition for which HS's value with conditional guarantee is higher than mergers'.

[25]They consider also a third component, the "limited liability", which is however always zero in our model.

[26]It is evident from expression 3.4.3 that Γ is positive as soon as R_H or R_S are non-empty

value, when positive, and they transfer the amount of cash needed to rescue their beneficiary company from default when their unit issued a conditional guarantee:

$$E_S(P_H, P_S) = \left[(X_S^n - P_S)^+ - (P_H - X_H^n)1_{\{R_S\}}\right] \tag{3.35}$$

$$E_H(P_H, P_S) = \left[(X_H^n - P_H)^+ - (P_S - X_S^n)1_{\{R_H\}}\right] \tag{3.36}$$

The payoff received by the debtholders of H is

$$\begin{aligned} D_H(P_H, P_S) &= X_H^\alpha 1_{\{0<X_H<X_H^Z\},\{X_S\leqslant h(X_H)\}} + X_H^{n,\alpha} 1_{\{X_H^Z<X_H<X_H^D\},\{X_S\leqslant h(X_H)\}} + \\ &\quad + P_H \left[1_{\{X_H>X_H^D\}} + 1_{\{R_S\}}\right] \end{aligned} \tag{3.37}$$

and, symmetrically, the payoff for the debtholders of S is

$$\begin{aligned} D_S(P_H, P_S) &= X_S^\alpha 1_{\{0<X_S<X_S^Z, X_H\leq h(X_S)\}} + X_S^{n,\alpha} 1_{\{X_S^Z<X_S<X_S^D, X_H\leq h(X_S)\}} + \\ &\quad + P_S \left[1_{\{X_S>X_S^D\}} + 1_{\{R_H\}}\right] \end{aligned} \tag{3.38}$$

When the guarantee is unilateral, the above expressions hold, with $1_{\{R_B\}}=\emptyset$ and $h(X_G) = +\infty$ where B(beneficiary) and G(guarantor) can be either H or S. The market values of equity (E_{0S} and E_{0H}) and debt (D_{0S} and D_{0H}) of both units are simply obtained as the discounted present values of these payoffs.

Capital Requirements in HS

We model the VaR-type constraint for the HS structure either as:

1. a unique capital requirement for the group, according to the consolidated approach[27]

$$E_{0H} + E_{0S} > Var_{\beta_H}(L_H) + VaR_{\beta_S}(L_S) \tag{3.39}$$

2. a unique capital requirement, that accounts for diversification effects in its computation:

$$E_{0H} + E_{0S} > Var_{\beta_{H,S}}(L_H + L_S) \tag{3.40}$$

[27]Under our set of assumptions, consolidation is full. The value of the participation in a subsidiary does not enter the own funds of the holding, but enters the asset side. Notice that in the constraints (3.39) and (3.40), E_{0S} is however added to E_{0H}, to take account of the dividend transfer that would enhance holding's shareholders value.

3. two separate capital requirements for H and S, "solo supervising" the two units:

$$E_{0H} > Var_{\beta_H}(L_H),\ E_{0S} > VaR_{\beta_S}(L_S) \tag{3.41}$$

Notice that the two levels β_H and β_S can differ, to account for the possible diversity of capital adequacy rules when H and S are involved in different sectors. The constraint establishes how much capital must be held by the group and where. Consolidated constraints do not limit the distribution of capital among units. This situation leaves room for two possible concerns. First, potential abuse on subsidiary's minority shareholders interests can be exercised by the entrepreneur. Second, an uneven distribution of capital between units could raise political issues when the entities involved in the group supervision process belong to different countries. Potential solution to this problem is given by the use of solo supervision, which requires a specific level of capital to be held at the subsidiary's level.

Value maximization

The entrepreneur maximizes the total value of the group V_{HS}, by choosing both P_S and P_H. The program he solves is then:

$$\begin{aligned} V^*_{HS}(P^*_S, P^*_H) &= \max_{P_S, P_H} \begin{bmatrix} E_{0S}(P_S,P_H) + D_{0S}(P_S,P_H) + \\ +E_{0H}(P_S,P_H) + D_{0H}(P_H,P_S) \end{bmatrix} && (3.42) \\ & \quad \text{s.t. one of} \\ & \quad (3.39)\ ,\ (3.40),\ \mathit{or}\ (3.41) \\ P_S &\geq 0,\ P_H \geq 0 && (3.43) \end{aligned}$$

Luciano and Nicodano (2010) prove in an unconstrained setting that the subsidiary is never unlevered at the optimum. The following theorem states instead that this can be the case when a capital requirement is introduced. We provide necessary conditions for the subsidiary to be unlevered in the constrained setting when there is a unilateral guarantee from H to S:

Proposition 3.4.4 *a) There exists a set of necessary conditions for a solution of the HS constrained problem in which the subsidiary ($P^*_S = 0$) is unlevered:* $\left(\tau < \frac{1}{1-\phi(1-F(0))}, \frac{\alpha}{\tau} \leq \frac{\frac{dX^Z_H}{dP_H}\left(1-F(X^Z_H)\right)(1-\tau)}{\left(1-\tau\frac{dX^Z_H}{dP_H}\right)}\right)$, *with* $\frac{\alpha}{\tau} = \frac{\frac{dX^Z_H}{dP_H}\left(1-F(X^Z_H)\right)(1-\tau)}{\left(1-\tau\frac{dX^Z_H}{dP_H}\right)}$ *if* $P^*_H > 0$. *b) There exists no solution in which both S and H are unlevered.*

Proof. see Appendix B. ■

3.5 Numerical Analysis: HG and IC

In this section we analyze the numerical application of our model for HGs and ICs. First, we calibrate the parameters of a SA to the observed figures of a financial firm in the medium class of speculative grade-rated companies. Then, we analyze whether joint or separate incorporation of identical units is optimal.

3.5.1 The Stand Alone constrained firm case: calibration

We calibrate our model parameters to the observed figures of financial institutions.
We assume that cash flows X_i are normally distributed, following Leland(2007):

$$X_i \sim N(X_0(1+r)^T, \sigma_i^2 T) \tag{3.44}$$

We set X_0 to 100 and the risk-free rate r to 5%, as he does.

Since deposits and policies can be considered as having virtually infinite maturity, we fix a time horizon T longer than the average maturity of bonds issued by financial companies [28]and we set it to 10 years. Hence, the mean of X_i is 162.89. As Leland (2007) does, we calibrated cash flow volatility to match our model-implied figures to a Ba/B Moody's rating class for financial firms (medium segment of the speculative grade). Indeed, we set σ to 17%, using the annualized equity volatility of financial institutions provided by Gropp and Heider (2010) and a linear transformation from Schaefer and Strebulaev (2008). 10-year model implied default probability is 22.7%, which is close to the Ba rating class observed one (19.1% according to Moody's (2009)). Since mixed evidence is found in empirical literature on the effective tax rates of financial firms[29], we follow Leland's calibration for non-financial companies, and set $\tau = 0.2$. Finally, we set the bankruptcy costs rate α, which is very hard to observe, to 10%, in order to obtain a leverage ratio of 82%[30] (Harding, Liang, and Ross (Harding et al.),Gropp and Heider (2010)) and

[28]Recent Moody's studies indicate average global bond maturity to 4.7 years

[29]Harding, Liang, and Ross (Harding et al.) point out that, based on FDIC Call reports, the average effective tax rate for commercial banks in the U.S. ranges from 19.8% for small firms to 32.7% for large ones.

[30]We use market leverage as a measure of leverage, following Gropp and Heider (2010)

$$L_0(P) = 1 - \frac{E_0(P)}{(E_0(P) + P)} \tag{3.45}$$

Table 3.1: Optimal value of SA with different β

	$\beta = 99\%$	$\beta = 99.5\%$
V_0^*	84.37	84.13
P^*	107	97
L^*	82.1%	77.7%
E_0^*	23.39	27.80
D_0^*	60.98	56.33
DC_0^*	1.26	0.85
TS_0^*	5.62	4.97
RR^*	68.1%	67.7%
DP^*	22.5%	16.8%

a recovery rate of 68.1%, which match the common observed figures of financial institutions. This value, way higher than the usual observed one for bonds (around 40%), is in line with our assumption that debt comprises also deposits and policies. Harding, Liang, and Ross (Harding et al.) report that these instruments, which are fully or almost completely insured, account for 67% of financial institutions' liabilities. Hence, our base case calibrated recovery rate seems to fit well our assumptions. Unlevered firm value is 80.00. Optimal nominal debt level is 107. Following the general practice used for banking firms, we calibrate our base case setting the regulatory requirement's β to 99% and compute the VaR constraint on a one-year horizon[31], under the hypothesis that operating cash flows are identically distributed in time. Annualized standard deviation of cash flows is then 17.00.
Table 3.1 presents the optimal figures of a stand alone company, when β is set to 99% and 99.5%.

3.5.2 Integrated conglomerates

We consider an IC which merges two units having the same characteristics of the SA unit described in the previous section 3.5.1. Cash flows are non-negatively correlated[32] and normally distributed. We explore whether the diversification effects of pooling the cash flows from correlated activities make the IC more valuable than the HG.

[31] Kretzschmar, McNeil, and Kirchner (2010) provide a theoretical framework for economic capital modelling and apply it to a sample of financial institutions. When using a VaR measure, they set the confidence bound to 99%, and the horizon to one year, as we do.

[32] We do not explore negative correlation cases since evidence suggests strong positive correlation between firms involved in the financial sectors.

Table 3.2: Integrated Conglomerates vs. Horizontal Groups

	HG	IC with div $\rho = 0.5$	IC no div $\rho = 0.5$
V_0^*	168.73	168.97	168.63
P^*	214	234	209
BL^*	82.28%	86.62%	81.71%
E_0^*	46.77	36.12	46.76
D_0^*	121.96	132.85	121.87
DP^*	39.86%*	26.47%	17.70%
RR^*	68.10%	71.58%	71.63%
TS_0^*	11.24	12.39	10.68
DC_0^*	2.52	3.42	2.06
Constraint	46.52	35.92	46.52

* The DP of the HG is the probability that at least one of the 2 units experiences default.

The benefits from diversification can derive from two channels: coinsurance between units and the reduction of regulatory capital when the VaR bound is computed on the merged cash flows' distribution. The two effects both have the same sign, as coinsurance decreases with correlation and a less diversified firm requires more prudential own funds.
Figure 3.1 compares the optimal value of an HG and of an IC. When the IC has to satisfy a constraint which accounts for diversification effects, it is more valuable than the SA for all correlation levels. It appears then that - purely financial - diversification opportunities can be exploited. Value gains are smaller for higher correlation between the cash flows of different units. For low correlation levels, value gains in the IC derive more from the tax savings obtained from leverage than from lower default costs. Up to a certain level of correlation, both tax gains and default costs increase with leverage. Above a certain correlation level (0.8 in our simulations), the capital requirement becomes binding. Further increasing correlation results in both tax savings and default costs to decrease. The former shrinks faster, leading to smaller value gains relative to the HG the higher the correlation (0.24 for $\rho = 0.5$, 0.13 for $\rho = 0.8$, 0.06 for $\rho = 0.9$).

The equity capital required for an IC with uncorrelated units (23.35) to set up at time 0 is halved compared to the HG required one (46.52)[33]. For all

[33]Notice that constraints are computed as $VaR_{99\%}$ on a normal with mean $-2\mu/10$ and variance $2\sigma^2/\sqrt{10}$ for the HG case and on a normal with mean $-2\mu/10$ and variance $\sigma^2(2+2\rho)/\sqrt{10}$ for the IC case, where μ, σ^2 and ρ denote mean, variance and correlation of the expected losses respectively.

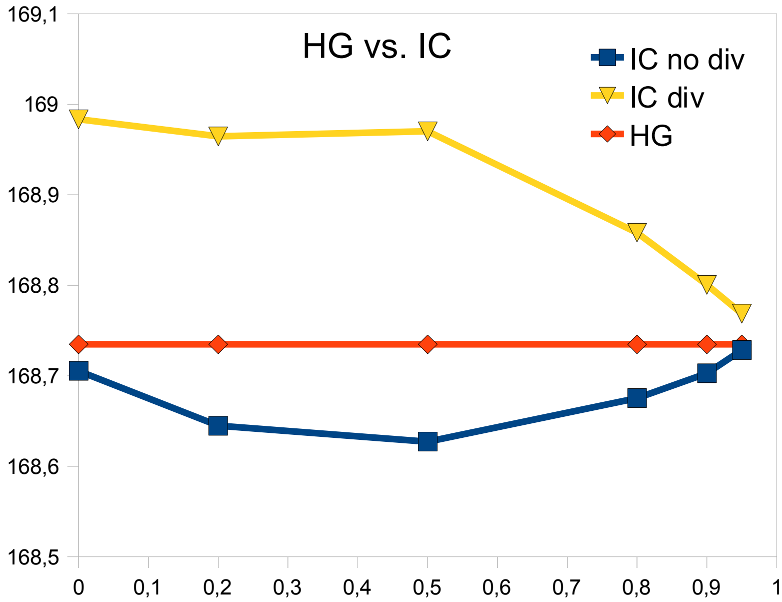

Figure 3.1: This Figure compares the optimal values of an IC and a SA with changing correlation. When the capital requirement takes into account diversification effects ('IC div') IC outperforms SA for all values of correlation. The opposite happens if the constraint is insensitive to different levels of correlation ('IC no div').

levels of correlation the IC faces a lower capital requirement relative to the HG. This allows it to raise a higher principal and enjoy greater tax savings. Leverage increases from 82% to almost 87%. The default probability of an IC is indeed lower than the one of one of the two SA (17% when $\rho = 0.5$ vs. 35%), but higher than the one of a joint default of both units of the HG (5%).
The opposite happens when no diversification effects are considered: HG is more valuable than the IC for all correlation levels. This is perhaps not surprising, since it strengthens the result obtained by Flannery et al. (1993) in a model in which tax advantaged debt generates an underinvestment problem. They argue that, with perfectly correlated activities, jointly incorporated divisions are suboptimal relative to separately incorporated ones, due to the lower flexibility of a unique capital structure. We obtain their result for all positive correlation levels simply imposing a financing constraint that does not allow for a diversification effect[34]. Hence, we shed some light on the role of regulation as a driving factor in the choice of separately vs. joint incorporation. Table 3.2 reports the optimal results for a correlation level $\rho = 0.5$. Expected default costs are remarkably lower than in the HG case (2.06 vs. 2.52), but this is not sufficient to offset the reduction in tax savings (10.68 vs. 11.24) and lead the IC to emerge as the optimal arrangement. Hence, value gains of an IC relative to a HG when units are symmetric derive uniquely from a "constraint effect".

3.6 Numerical Analysis: HS

In this section we study the optimal properties of an HS structure and their sensitivity with respect to different financing constraints. We consider the two cases of consolidated and solo supervision and we analyze the role of legally binding support guarantees. We consider a HS made up of two units with identical cash flow distribution, tax rates, default cost rates and whose correlation between cash flows is $\rho = 0.5$.

3.6.1 Consolidated Supervision

Table 3.6.1 summarizes the results of our numerical analysis when the group is subject to consolidated supervision. It shows optimal HS figures under

[34]Notice that in the unconstrained case and for Leland's calibration for commercial firms, coinsurance results in purely financial synergies to be exploited when symmetric units are merged, for all correlation levels.

Table 3.3: HS and HCM with consolidated constraint, $\rho = 0.5$

	HS	HCM
V_0^*	174.83	169.73
P_S^*	297	230
P_H^*	58	0
E_{0S}^*	0.01	27.69
E_{0H}^*	35.93	15.20
D_{0H}^*	35.12	0
D_{0S}^*	103.77	126.84
DP_H^*	4%	0.12%
DP_S^*	71.4%	26.88%
JD	3.7%	0.12%
DC_{0HS}^*	6.81	2.85
TS_{HS}^*	21.62	12.56

JD is the probability of joint default at T, while DP_i is the selective default probability of unit i. Units of HS are identical.

constraint (3.39).[35]

Unreported results highlight that when ρ is positive and units are symmetric HS always emerges as the optimal arrangement, since it maximizes value with respect to its equivalent IC and HG.

As showed in Luciano and Nicodano (2010), without any limitation on equity capital, the optimal choice for the firm would be to exploit the conditional guarantee keeping the whole equity capital in one unit - which acts as the guarantor - and issue debt only on the name of the other one, which is a beneficiary of support only and never provides it. When we introduce a capital requirement which has to be met at a consolidated level, principals of both units can instead be positive at the optimum. However, the guarantee still turns out to be almost exploited one-way as in the unconstrained case. Equity capital is kept in one unit which specializes in providing conditional support to the other. Debt continues to be issued mostly on the name of the other division, where high tax savings are generated. Optimality of HS relative to the HG and IC structures is due entirely to these tax savings, which are almost twice their HG and IC counterparts (21.62 vs. 11.24 and 12.39, respectively). As a consequence, the 10-year default probability of this subsidiary is huge, 75.1%, and expected default costs are 5 times the ones of a comparable SA (6.71 vs. 1.26). Rescue takes place with a probability of 25%. The high recovery rate (56.85%) leads to moderate implied credit

[35]The intuition behind the results does not change when the financing constraint is met at the consolidated level, but its level is the sum of the "solo" ones, (3.40).

spreads[36] (6.10%). This value is however way (4 times) higher than the credit spread observed for Ba/B firms according to Huang and Huang (2002) (165 basis points). On the contrary, the holding's default probability (3.6%) is 8 times lower than the stand alone one and its expected default costs are minimal (0.1). Summarizing, the HS optimally presents a safe holding company which exploits the high recovery rate and the conditional guarantee to provide support to its subsidiary's debtholders.
Despite being the privately optimal arrangement, HS emerges then in this case as the less desirable one from the regulator's perspective for many reasons. First, because default costs, which are way higher in the HS than in the other competing organizations, represent a deadweight loss in the economy. Second, because one of the two units has optimally almost zero equity value.[37] Group value enhancement is obtained at the expenses of the equityholders of the subsidiary and - in particular - of its minority shareholders. Moreover, imagine the two units are incorporated in different countries: the subsidiary will bring no capital in the market where is set, despite being a very risky firm.[38]

3.6.2 Regulating subsidiaries

The results of the previous section highlight that a more equal distribution of capital should be beneficial for the economy as a whole. In this section we analyze the effects of introducing solo supervision, which is an instrument the regulator has to drive HS's choices.
Hence, we solve the maximization program (3.42) under the constraint (3.41). We set $\beta_H = \beta_S = 99\%$. The first column of Table 3.4 reports the optimal

[36]Model credit spreads are computed as

$$s_i^*(P_i) = y_i(P_i) - r = \sqrt[T]{\frac{P_i}{D_{0i}}} - 1 \qquad (3.46)$$

$y_i^*(P_i^*)$ denotes than the yield of the risky debt of a unit at the optimum.

[37]This finding could point out a role of mutual guarantees in explaining the "mystery of zero leverage firms" Strebulaev and Yang (2006) point out. In particular, it could help reconcile the fact that zero leverage firms do not seem to be concentrated among holding companies or subsidiaries.

[38]Many factors prevent such a situation to arise in reality when the subsidiary is a regulated entity. First, when the subsidiary is a public company, the interests of minority shareholders can not be threatened this way. Second, minimum capital standards or leverage ratios are generally required in order to enjoy the deposit insurance scheme that allows the firm to offer its debtholders a secured amount of instruments and thus to experience the high recovery rate we highlighted above as crucial. However, such a representation can closely mimic what happens in reality when the subsidiary is an unregulated financial company.

Table 3.4: Optimal Value for HS, solo regulation

	HS	2(H),1(S) MG	1(H),2(S) UG
V^*_{0HS}	169.34	169.06	168.85
P^*_{HS}	210	201	201
P^*_S	105	106	106
P^*_H	105	95	95
E^*_{0HS}	46.91	51.13	51.57
D^*_{0HS}	122.28	117.93	117.27
DP^*_S	3.85%	13.84%	13.82%
DP^*_H	3.85%	11.73%	15.16%
TS^*_{0HS}	11.64	11.02	10.23
TS^*_{0S}	5.82	5.85	5.38
TS^*_{0H}	5.82	5.17	4.85
DC^*_{0HS}	1.68	1.34	1.40
DC^*_{0S}	0.84	0.75	0.62
DC^*_{0H}	0.84	0.59	0.78

The 'HS MG' column refers to a HS structure in which a Mutual Guarantee is in place, while the 'HS UG' one refers to a HS structure in which H is the guarantor of S. 1 refers to the unit with $\beta = 99.5\%$, while for 2 $\beta = 99\%$.

figures of this arrangement, when the correlation between units $\rho = 0.5$.

Optimal total value drops from 174.83 when HS is subject to consolidated supervision to 169.34. The optimal guarantee in place is mutual[39]. This decrease in value is due to the reduction of total tax savings which is even higher than the decrease in total default costs, which drop from 6.81 to 1.68 (1.91 when the guarantee is unilateral). This is a highly desirable outcome from a welfare perspective, since the default costs of both units (0.84) are lower than the stand alone ones (1.25) and groups' figures are lower relative to any other arrangement (2.52 in the HG, 3.42 in the IC with diversification effects, 2.05 in the IC when no diversification is accounted for). Moreover, taxes levied are higher than in the stand alone case. The optimal solution presents the same level of principal in both units, which are less levered than their stand alone counterpart (105 vs. 107): total debt capacity is the lowest among all the arrangements. Obviously, the entrepreneur strictly prefers the previous situation, since he cashes in more money at time 0. This is due basically to the higher amount lent by the debtholders of the subsidiary.
However, even when regulated in this way, HS is value maximizing with respect to other organizations. Hence, solo supervision, combined with the presence of a legally binding conditional support guarantee issued by the

[39]The value of a HS with unilateral guarantee (unreported) is 169.07.

holding to its subsidiary, is able to make HS structure both private and welfare optimal, at least for some levels of correlation.

3.6.3 HCM

We reproduce the holding company model in a stylized way, as a parent company which has little or no operating income[40] and owns a merger of its subsidiaries. We set the non-operating holding's operating (NOHC) returns as 9 times smaller than those of its subsidiaries, which are merged into a unique entity and consider a unilateral guarantee from H to S. The group meets its capital requirement at the consolidated level and we consider a unilateral guarantee from the NOHC to its subsidiaries. The last column in Table 3.3 reports the optimal figures of such an arrangement. When $\rho = 0.5$ the value of the HCM structure is higher than the IC and HG.

Optimal value is 169.43. Capital is held more in the larger subsidiary (31.54) than at the holding company's level (15.29), while debt is issued only on the name of the controlled company. Since equity capital is more evenly distributed between units, the problems we described for HS under consolidated supervision are avoided. Optimal holding's equity value is higher than the solo requirement (4.65), while the subsidiary's one is correspondingly lower. Despite a very high principal (219), the default probability of the subsidiary (23.2%) is just slightly higher than the one of a stand alone (22.3%). Rescue happens 9% of the times, even if the holding expects a remarkably lower level of returns. The figures we obtained seem to reproduce what happens in reality. Non operating holding companies do not issue debt (P_H^* is 0) and thus are perceived as very safe (0.12% 10-year default probability). Very interestingly, such an arrangement outperforms comparable ICs and HGs also when the two units have to meet their capital requirement separately. Value is 168.89 and the capital distribution we described above does not change much, since while the constraint in the subsidiary becomes binding (41.87), the holding still keeps more capital than it is forced to and uses this extra buffer to rescue its subsidiary in case of default (it happens with a probability of 2.7%). The holding optimally levers itself, issuing a principal debt of 16, contrary to what happens with the consolidated constraint. Consequently, it becomes more risky, raising its default probability to 9.5% . The subsidiary, instead, issues a lower nominal amount of debt than its HS counterpart (193), is slightly safer and experiences distress with a probability of 20% in 10 years.

[40]Evidence (Edwards (1998)) shows that NOHC have assets of total size 10 times smaller than their subsidiaries.

3.6.4 Regulatory arbitrage in units subject to different capital requirements

The term *bancassurance* indicates all the kind of relationships between the banking and insurance industries and also the financial conglomeration between these two sectors. Empirical literature highlighted that diversification of products and geographical areas when combining two units can lead to revenue and cost advantages[41]. The difference in regulation between banking, securities and insurance firms, which originates from the very nature of the products and services they offer, along with de-regulation of these segments, gives financial conglomerates possibilities and incentives to shift assets from more to less demanding units in terms of capital requirements. In the current regulatory environment, no diversification benefits are recognized when computing capital requirements when units belong to different sectors[42]. In this section we analyze the purely financial effects of a regulatory asymmetry on capital structure. We consider a simple HS arrangement, made up of two units which differ in the level of confidence of the VaR constraint. We set the level of β to 99% for one entity and to 99.5% for the other one.

The last two columns of Table 3.4 report the optimal configuration when $\beta_1 = 99\%$ and $\beta_2 = 99.5\%$. Cash flow correlation is set to $\rho = 0.5$. The HS is most valuable when there is a mutual guarantee in place between the divisions. When a unilateral guarantee is in place, the optimal configuration of the HS is the one in which the unit that faces the tighter capital requirement - unit 2 in this case - is the guarantor. Firm 1 can be capitalized less and levered up more due to the conditional guarantee. The beneficiary is not extremely leveraged anymore, since it is "solo" supervised. In the unilateral guarantee case, the group issues less debt in the unit that provides support relative to the HG case (106 vs. 107) but levers up the beneficiary more, exploiting the conditional rescue opportunities (rescue happens with a probability of 9.68%). It has the same principal of the SA case, but debt market value is improved by the insurance provided by the guarantee (57.07 vs. 56.33). The value of the HS arrangement under the mutual guarantee is

[41] Chen et al. (2006) empirically study bank and insurance M&As from 1983 to 2004. They highlight the growing trend to conglomeration (2 transactions in 1983, 20 in 2004), they find diversification effects in the bidder bank's beta with the national stock index, but no long-term significant wealth effect on the acquirer. Also, Chen, Li, Liao, Moshirian, and Szablocs (2009) provide evidence in favour of the profitability of bancassurance, highlighting that revenue and cost savings are significant and that the latter are positively related to the size of the operation.

[42] Herring and Carmassi (2008) provide a detailed survey of country-level supervision of the different sectors and analyze possible benefits and disadvantages of creating a unique cross-sectoral regulator as a response to such a problem.

greater than that of an equivalent HG, 169.05 (168.71 when the guarantee is unilateral) vs. 168.49, thanks to the value increase due to coinsurance. Though mutual guarantees are optimal, the HS where unit I is the guarantor of a unilateral guarantee is again preferred to the HG. The value increase is again obtained despite a lower debt capacity (201 vs. 204 in the HG). When can regulatory asymmetries (different βs across units) and size asymmetries are coupled, we can reproduce the phenomenon of capital arbitrage due to asset shifting. Luciano and Nicodano (2010) showed that larger firms should optimally be the guarantors of unilateral guarantees. By means of our model, we can analyze when a smaller firm optimally emerges as the guarantor when subject to a stricter capital requirement.

3.7 Concluding comments

We analyzed optimal capital structure and ex-ante optimal organizational form choices of regulated financial conglomerates. We abstracted from agency problems and informational asymmetries to focus on the trade-off between tax advantaged debt and default costs and coinsurance between units. The presence of prudential regulation in the form of a VaR-type constraint optimally reduces the ex-ante risk-taking incentives of an entrepreneur who maximizes his own initial cash flow. Firm's capital structure choices are obviously influenced by the capital requirements.
When monitored at a consolidated level, HS structures in which conditional guarantees exist hold all their capital in one unit. As in Luciano and Nicodano (2010), we find thinly capitalized units in the optimum, raising both a minority shareholder protection and a political issue. Moreover, these units are highly levered and bear most of the default costs of the group, which are higher than in any other competing organization. Capital is more equally distributed and the group is more sound when units are regulated as "solo" entities. In this latter case, when firms can not freely allocate their equity capital among units, private value drops but still HS can emerge as the private optimal configuration. Moreover, HS structures in this case can bear the lowest level of default costs, which are a deadweight loss to the economy. Hence, our analysis suggests that enforcing conditional guarantees - in a way similar to the one suggested by the Group Support Framework - inside groups could enhance welfare, without altering the optimality of HS structures. We also analyzed the case in which units are subject to different capital requirements and found that unilateral guarantees are optimally issued by the firms which are more strictly regulated. We leave to further work the application of our model to capital arbitrage, exploring the possible combinations of size

and regulatory asymmetry.

3.8 Appendix A - Stand Alone Maximization Problem

Before tackling the problem of the SA optimal solution, we establish some properties for its debt, equity, tax shield and default threshold by proving propositions 3.4.1, 3.4.2 and 3.4.3. Throughout this Appendix, we suppress the index i for notational simplicity, since we refer to a stand alone firm.

Proof of proposition 3.4.1.

Let us first focus on showing part 1 of the proposition. First of all, we prove that $\dfrac{dD_0(P)}{dP}$ is bounded above and below.

$$\frac{dD_0}{dP} = \phi\left[-\alpha\frac{dX^d}{dP}X^d f(X^d) + \tau\frac{dX^Z}{dP}\left(F(X^d) - F(X^Z)\right) + 1 - F(X^d)\right] \tag{3.47}$$

$$\frac{dD_0}{dP} = \frac{-\alpha X^d f(X^d) + \tau\left[F(X^d) - F(X^Z)\right] + 1 - F(X^d)}{\dfrac{1}{\phi} + \dfrac{\alpha\tau}{1-\tau}X^d f(X^d) + \tau(F(X^d) - F(X^Z))} \tag{3.48}$$

This is lower than or equal 1 if its numerator is lower than the denominator. Then:

$$\frac{1}{\phi} + \frac{\alpha\tau}{1-\tau}X^d f(X^d) \geq -\alpha X^d f(X^d) + 1 - F(X^d) \tag{3.49}$$

$$\Longrightarrow \frac{1}{\phi} - 1 + \frac{\alpha\tau}{1-\tau}X^d f(X^d) \geq -\alpha X^d f(X^d) - F(X^d) \tag{3.50}$$

This is true since the l.h.s is non negative $\left(\dfrac{1}{\phi} \geq 1\right)$ while the r.h.s. is negative. The inequality is strict as soon as $r > 0$ $(\phi < 1)$ .This completes the proof of 1 as an upper bound. For the lower bound, we check whether $\dfrac{dD_0}{dP} > -\dfrac{1-\tau}{\tau}$. This would imply:

$$-\alpha\tau X^d f(X^d) + \tau^2\left[F(X^d) - F(X^Z)\right] + \tau - \tau F(X^d) > \tag{3.51}$$

$$-(1-\tau)\left[\frac{1}{\phi} + \frac{\alpha\tau}{1-\tau}X^d f(X^d) + \tau(F(X^d) - F(X^Z))\right] \tag{3.52}$$

$$-\alpha\tau X^d f(X^d) + \tau^2 \left[F(X^d) - F(X^Z)\right] + \tau - \tau F(X^d) > \tag{3.53}$$

$$-\frac{1}{\phi} + \frac{\tau}{\phi} - \alpha\tau X^d f(X^d) - \tau(F(X^d) - F(X^Z)) + \tau^2 \left[F(X^d) - F(X^Z)\right] \tag{3.54}$$

$$\tau - \tau F(X^d) > -\frac{1}{\phi} + \frac{\tau}{\phi} - \tau(F(X^d) - F(X^Z)) \tag{3.55}$$

$$\tau(1 - F(X^d)) > -(1-\tau)\frac{1}{\phi} - \tau(F(X^d) - F(X^Z)) \tag{3.56}$$

Which is always true since the l.h.s. is positive and the r.h.s. negative (even with $\tau = 0$).

Remark 3.8.1 *We can not unequivocally sign the derivative of D_0 with respect to debt.* $\frac{dD_0}{dP}$ *is negative when* $\frac{\alpha}{\tau} \geq \frac{\Pr(X^Z < X < X^d) + \frac{\Pr(X > X^d)}{\tau}}{X^d f(X^d)}$. *Its value when $P = 0$ is*

$$\frac{dD_0(0)}{dP} = \phi \tag{3.57}$$

Before showing concavity, we need to prove some useful bound on the derivatives of X^Z and X^D

Lemma 3.8.1 *X_i^Z and X_i^D are both increasing in P and their derivatives have upper bounds ($\frac{1}{\tau}$ and $\frac{1}{1-\tau}$ respectively)*

Proof.

$$\frac{dX^Z}{dP} = 1 - \frac{dD_0}{dP}. \tag{3.58}$$

$$\frac{dX^Z}{dP} \geq 0 \text{ since } \frac{dD_0}{dP} \leq 1 \text{ and } \frac{dX^Z}{dP} < \frac{1}{\tau} \text{ since } \frac{dD_0}{dP} > -\frac{1-\tau}{\tau} \tag{3.59}$$

$$\frac{dX^d}{dP} = 1 + \frac{\tau}{1-\tau}\frac{dD_0}{dP}. \tag{3.60}$$

$$\text{Hence, } \frac{dX^d}{dP} < 1 + \frac{\tau}{1-\tau} = \frac{1}{1-\tau} \text{ and } \frac{dX^d}{dP} > 0. \tag{3.61}$$

This concludes the proof of our lemma. ■

We can now use this lemma and easily show the concavity of D_0 with respect to P :

$$\frac{d^2D_0}{dP^2} = \frac{-\alpha\left(\frac{dX^d}{dP}\right)^2 f(X^d) - \alpha X^d f(X^d) - \left(1-\tau\frac{dX^Z}{dP}\right)\frac{dX^d}{dP}f(X^d) - \tau\left(\frac{dX^Z}{dP}\right)^2 f(X^Z)}{\frac{1}{\phi} + \frac{\alpha\tau}{1-\tau}X^d f(X^d) + \tau(F(X^d) - F(X^Z))} \tag{3.62}$$

The denominator is positive, hence in order to show concavity we have to prove that the numerator is negative. Using lemma 3.8.1, $\frac{dX^Z}{dP} < \frac{1}{\tau}$ and $\frac{dX^d}{dP} > 0$ imply that all the terms in the numerator are non positive and that at least the third is strictly negative. Hence $\frac{d^2D_0}{dP^2} < 0$.

This completes our proof of the first part of proposition 3.4.1.

Let us now turn to part 2.

The first derivative of E_0 is given by $\frac{dE_0}{dP} = \phi\int_{X^d}^{+\infty} f(x)dx \left(\tau\frac{dX^Z}{dP} - 1\right)$.

From lemma 3.8.1, $\frac{dX^Z}{dP} < \frac{1}{\tau}$ implies that the term in parenthesis is strictly negative. Hence, E_0 is decreasing in P.

The second derivative is

$$\frac{d^2E_0}{dP^2} = \phi\left(1 - F(X^d)\right)\tau\left(-\frac{d^2D_0}{dP^2}\right) + \phi\left(1-\tau\frac{dX^Z}{dP}\right)\frac{dX^d}{dP}f(X^d) \tag{3.63}$$

Since in the first part of the proof we showed concavity of debt with respect to P, $\frac{d^2D_0}{dP^2} < 0$, the first term of this derivative is non-negative, while the second is strictly positive. This implies $\frac{d^2E_0}{dP^2} > 0$ and hence convexity of E_0 with respect to P. This concludes the proof of the proposition. ■

We now establish some conditions under which the objective function $D_0 + E_0$ is quasi-concave. Under these conditions, the Kuhn-Tucker conditions of the program (3.17) are necessary and sufficient for an optimum.

Proposition 3.8.1 ***Conditions for concavity of the objective function.*** *We can provide bounds on the first derivative of* $\frac{dD_0}{dP}$ *that ensure concavity of the objective function.*

Proof.

$$\frac{d^2E_0}{dP^2} = \phi\left(1 - F(X^d)\right)\tau\left(-\frac{d^2D_0}{dP^2}\right) + \phi\left(1-\tau\frac{dX^Z}{dP}\right)\frac{dX^d}{dP}f(X^d) \tag{3.64}$$

$$\frac{d^2E_0}{dP^2}+\frac{d^2D_0}{dP^2}=\frac{d^2D_0}{dP^2}\left(1-\phi\left(1-F(X^d)\right)\tau\right)+\phi\left(1-\tau\frac{dX^Z}{dP}\right)\frac{dX^d}{dP}f(X^d) \tag{3.65}$$

We want to see when $\frac{d^2D_0}{dP^2}\left(1-\phi\left(1-F(X^d)\right)\tau\right)+\phi\left(1-\tau\frac{dX^Z}{dP}\right)\frac{dX^d}{dP}f(X^d)\leq 0.$

This happens if

$$\frac{d^2D_0}{dP^2}\leq-\frac{\phi\left(1-\tau\frac{dX^Z}{dP}\right)\frac{dX^d}{dP}f(X^d)}{\left(1-\phi\left(1-F(X^d)\right)\tau\right)}=\frac{\phi f(X^d)(\tau-1)\left(\frac{dX^d}{dP}\right)^2}{\left(1-\phi\left(1-F(X^d)\right)\tau\right)}. \tag{3.66}$$

(obtained using the fact that $\frac{dX^Z}{dP}=1-\frac{1-\tau}{\tau}\frac{dX^d}{dP}+\frac{1-\tau}{\tau}$) or, since

$$\frac{d^2E_0}{dP^2}+\frac{d^2D_0}{dP^2}=\phi\left[\begin{array}{c}\left(1-F(X^Z)\right)\tau\frac{d^2X^Z}{dP^2}-\alpha\frac{d^2X^d}{dP^2}X^df(X^d)+\\-\alpha\left(\frac{dX^d}{dP}\right)^2f(X^d)-\tau\left(\frac{dX^Z}{dP}\right)^2f(X^Z)\end{array}\right] \tag{3.67}$$

This can be written, using the fact that $\frac{d^2D_0}{dP^2}=\frac{A}{C}\frac{dD_0}{dP}$ (where A is the numerator in (3.48), C the one in (3.62)) as a second order equation in $\frac{dD_0}{dP}$ that implies:

$$\begin{aligned}\frac{-b-\sqrt{b^2-4ac}}{2a}\quad &\leq\quad \frac{dD_0}{dP}\leq\frac{-b+\sqrt{b^2-4ac}}{2a}, && (3.68)\\ & \quad\ \ \textit{where} && (3.69)\\ b\quad &=\quad 1-\phi\tau+\phi\tau F(X^d)+2\phi\tau f(X^d) && (3.70)\\ a\quad &=\quad \frac{\phi\tau^2}{1-\tau}f(X^d) && (3.71)\\ c\quad &=\quad \phi f(X^d)(1-\tau). && (3.72)\end{aligned}$$

■

The KT conditions of the program (3.17) for a SA firm

$$\begin{aligned}&\max_P E_0(P)+D_0(P) && (3.73)\\ s.t.\ P\ &\geq\ 0 && (3.74)\\ E_0(P)\ &\geq\ \mathrm{VaR}_\beta(X)=\inf\{l:P(X>l)\geq 1-\beta\} && (3.75)\end{aligned}$$

are:

1. $(1+\lambda_1)\dfrac{\partial E_0(P)}{\partial P}+\dfrac{\partial D_0(P)}{\partial P}\leq 0$
2. $P\geq 0$
3. $P\left[(1+\lambda_1)\dfrac{\partial E_0(P)}{\partial P}+\dfrac{\partial D_0(P)}{\partial P}\right]=0$
4. $E_0(P)\geq VaR_\beta(X)$
5. $\lambda_1\geq 0$
6. $\lambda_1\left[E_0(P)-VaR_\beta(X)\right]=0$

We can now proceed to prove proposition 3.4.2

Proof of Proposition 3.4.2. The constraint can be non binding only when $\lambda_1=0$. It is easy to show that the l.h.s. of 1 at $P^*=0$ is:

$$\begin{aligned}
\frac{\partial E_0(0)}{\partial P}+\frac{\partial D_0(0)}{\partial P} &= \phi\left[1-F(0)\right]\left[\tau\left(1-\phi(1-F(0)\right)-1\right]+ && (3.76)\\
+\phi\left[1-F(0)\right] &= \phi\tau\left[1-F(0)\right]\left[\phi(1-F(0))\right] && (3.77)\\
\text{which is non negative, since } \phi &\leq 1 \text{ and it is} && (3.78)\\
\text{strictly positive as soon as } r &> 0,\ (1-F(0))<1. && (3.79)
\end{aligned}$$

The KT condition 1 is violated and no optimally unlevered solution exists if the constraint is non binding, unless $\tau=0$ or $F(0)=1$. These are both uninteresting cases, since the first implies no taxes and the second implies no positive probability of positive cash flows.

When $P^*>0$, instead we can characterize the solution through this condition:

$$\alpha\frac{dX^d}{dP}(P^*)X^{d*}f(X^{d*})-\tau\frac{dX^Z}{dP}(P^*)\left(1-F(X^{Z*})\right)=0 \tag{3.80}$$

$$\frac{\alpha}{\tau}=\frac{\dfrac{dX^Z}{dP}(P^*)\left(1-F(X^{Z*})\right)}{X^{d*}\dfrac{dX^d}{dP}(P^*)f(X^{d*})}. \tag{3.81}$$

■

Proof of Proposition 3.4.3. There exists a solution if the KT conditions are not violated. Hence, feasible solutions must meet Condition 1:

$$(1+\lambda_1)(1-F(X^{d*}))\left(\tau\frac{dX^Z}{dP}(P^*)-1\right)-\alpha\frac{dX^d}{dP}(P^*)X^{d*}f(X^{d*})+ \tag{3.82}$$

$$+\tau\frac{dX^Z}{dP}(P^*)\left(F(X^{d*})-F(X^{Z*})\right)+1-F(X^{d*})\leq 0 \tag{3.83}$$

$$\tau\frac{dX^Z}{dP}(P^*)\left(1-F(X^{Z*})\right)+\lambda_1\left(\tau\frac{dX^Z}{dP}(P^*)-1\right)(1-F(X^{d*}))+ \tag{3.84}$$

$$-\alpha\frac{dX^d}{dP}(P^*)X^{d*}f(X^{d*})\leq 0 \tag{3.85}$$

$$\lambda_1\leq\frac{\alpha\frac{dX^d}{dP}(P^*)X^{d*}f(X^{d*})-\tau\frac{dX^Z}{dP}(P^*)\left(1-F(X^{Z*})\right)}{\left(\tau\frac{dX^Z}{dP}(P^*)-1\right)(1-F(X^{d*}))}=f(P^*) \tag{3.86}$$

Notice that this condition must be satisfied with an equality if $P^*>0$.

This condition should be coupled with Condition 5. For Condition 5 to be satisfied, since its denominator is negative $\left(\tau\frac{dX^Z}{dP}(P^*)-1\right)<0$, we need to prove that its numerator is negative too. Hence:

$$\alpha X^{d*}\frac{dX^d}{dP}(P^*)f(X^{d*})-\tau\frac{dX^Z}{dP}\left(1-F(X^{Z*})\right)\leq 0 \tag{3.87}$$

$$\frac{\alpha}{\tau}\leq\frac{\frac{dX^Z}{dP}(P^*)\left(1-F(X^{Z*})\right)}{X^{d*}\frac{dX^d}{dP}(P^*)f(X^{d*})}=g(P^*) \tag{3.88}$$

Hence, the constraint can be binding at the optimum when condition (3.88) holds. In this case, $\lambda_1=f(P^*)$.

As soon as conditions (3.86) and (3.88) hold, one can always find an interval of values of β such that KT condition 4 holds.

When $P^*=0$, condition (3.86) becomes

$$\lambda_1 \leq \frac{-\tau(1-\phi)(1-F(0))}{(\tau(1-\phi)-1)(1-F(0))} \tag{3.89}$$

$$\lambda_1 \leq \frac{\tau(1-\phi)}{1-\tau(1-\phi)} \tag{3.90}$$

while (3.88) is always satisfied since the r.h.s. goes to infinite. Hence, an unlevered solution exists for some β when $0<\lambda_1\leq\frac{\tau(1-\phi)}{1-\tau(1-\phi)}$. This condition is always satisfied for the (uninteresting) case of $\phi\to 0$ (hence, as $r\to\infty$). ■

3.9 Appendix B - HS value maximization problem

Before tackling the proof of proposition 3.4.4, we compute the derivatives of $\Gamma(P_S, P_H)$ with respect to both its variables:

$$\begin{aligned}
\frac{d\Gamma(P_S,P_H)}{dP_H} &= -\phi\alpha\frac{dX_H^D}{dP_H}\left[\int_0^{X_S^Z} xf(x,X_H^D+\frac{P_S}{1-\tau}-\frac{x}{1-\tau})dx+\int_{X_S^Z}^{X_S^D} xf(x,X_H^D+X_S^D-x)dx\right] \\
\frac{d\Gamma(P_S,P_H)}{dP_S} &= \phi\alpha\left[\begin{array}{c} -\frac{1}{1-\tau}\int_0^{X_S^Z} xf(x,X_H^D+\frac{P_S}{1-\tau}-\frac{x}{1-\tau})dx+ \\ -\frac{dX_S^D}{dP_S}\int_{X_S^Z}^{X_S^D} xf(x,X_H^D+X_S^D-x)dx+\int_{X_H^D}^{+\infty}\frac{dX_S^D}{dP_S}X_S^D f(X_S^D,y)dy \end{array}\right]
\end{aligned}$$

Proof of proposition 3.4.4. Without loss of generality, we prove the theorem and derive the conditions when a unilateral guarantee from H to S is in place. A fortiori, a similar result can be obtained when the guarantee is mutual. When the constraint is binding, then $\lambda_1 > 0$.

Then, one of the KT conditions implies (at $P_S^* = 0$)

$$(1+\lambda_1)\frac{dE_{0S}(0)}{dP_S}+\frac{dD_{0S}(0)}{dP_S}+\frac{d\Gamma(0,P_H^*)}{dP_S}+\lambda_1\frac{dE_{0H}(0,P_H^*)}{dP_S}\leq 0$$

The last two terms vanish at $P_S = 0$, as one can see below:

$$\frac{\partial E_{0H}(P_S,P_H)}{\partial P_S}=\phi\left[\begin{array}{c} +\frac{1}{1-\tau}\int_0^{X_S^Z}(P_S-(1-\tau)y-\tau X_S^Z)f(X_H^D+\frac{P_S}{1-\tau}-\frac{y}{1-\tau},y)dy+ \\ +\frac{dX_S^D}{dP_S}\int_{X_S^Z}^{X_S^D}(P_S-(1-\tau)x-\tau X_S^Z)f(X_H^D+X_S^D-y,y)dy+ \\ -\int_{X_H^D+\frac{P_S}{1-\tau}-\frac{y}{1-\tau}}^{+\infty}\left(1-\tau\frac{dX_S^Z}{dP_S}\right)\int_0^{X_S^Z} f(x,y)dydx+ \\ -\int_{X_H^D+X_S^D-y}^{+\infty}\left(1-\tau\frac{dX_S^Z}{dP_S}\right)\int_0^{X_S^Z} f(x,y)dydx \end{array}\right]$$

Hence,

$$(1+\lambda_1)\frac{dE_{0S}(0)}{dP_S}+\frac{dD_{0S}(0)}{dP_S}\leq 0$$

must be satisfied.
Substituting:

$$\begin{aligned}
(1+\lambda_1)\phi\left[1-F(0)\right]\left[\tau\left(1-\phi(1-F(0)\right)-1\right]+\phi\left[1-F(0)\right] &\leq 0 \\
(1+\lambda_1)\left[\tau\left(1-\phi(1-F(0)\right)-1\right]+1 &\leq 0 \\
\lambda_1\left[\tau\left(1-\phi(1-F(0)\right)-1\right]+\tau\left(1-\phi(1-F(0)\right) &\leq 0
\end{aligned}$$

$$\lambda_1 \leq \frac{\tau\left(1-\phi(1-F(0)\right)}{\left[1-\tau\left(1-\phi(1-F(0)\right)\right]}$$

Feasible solutions have $\lambda_1 > 0$. The numerator is always positive, while the denominator is positive when

$$\begin{aligned}[1-\tau\,(1-\phi(1-F(0))] &> 0 \\ \tau &< \frac{1}{1-\phi(1-F(0)}\end{aligned}$$

$$\frac{\partial E_{0H}(P_S,P_H)}{\partial P_H} = \phi\left[\begin{array}{c} \left(\tau\frac{dX_H^Z}{dP_H}-1\right)(1-F(X_H^D))+ \\ +\frac{dX_H^D}{dP_H}\int_0^{X_S^Z}(P_S-\tau X_S^Z-(1-\tau)y)\ f(X_H^D+\frac{P_S}{1-\tau}-\frac{y}{1-\tau},y)dy \\ +\frac{dX_H^D}{dP_H}\int_{X_S^Z}^{X_S^D}(P_S-\tau X_S^Z-(1-\tau)y)\ f(X_H^D+X_S^D-y,y)dy \end{array}\right]$$

As for the other condition, since

$$\frac{\partial E_{0H}(0,P_H)}{\partial P_H} = \phi\left[\left(\tau\frac{dX_H^Z}{dP_H}-1\right)(1-F(X_H^D))\right]$$

then one has

$$-\alpha\frac{dX_H^D}{dP_H}X_H^D f(X_H^D)+\tau\frac{dX_H^Z}{dP_H}(1-F(X_H^Z))(1+\lambda_1) \;\leq\; 0$$

$$\lambda_1 \;\leq\; \frac{\alpha\frac{dX_H^D}{dP_H}X_H^D f(X_H^D)-\tau\frac{dX_H^Z}{dP_H}(1-F(X_H^Z))}{\tau\frac{dX_H^Z}{dP_H}(1-F(X_H^Z))} \tag{3.91}$$

which has feasible solutions only when λ_1 is non negative, leading to a condition on $\frac{\alpha}{\tau}$

$$\begin{aligned}\alpha\frac{dX_H^D}{dP_H}X_H^D f(X_H^D)-\tau\frac{dX_H^Z}{dP_H}(1-F(X_H^Z)) &\geq 0 \\ \frac{\alpha}{\tau}\left(\frac{1}{1-\tau}-\frac{\tau}{1-\tau}\frac{dX_H^Z}{dP_H}\right)X_H^D f(X_H^D)-\frac{dX_H^Z}{dP_H}(1-F(X_H^Z)) &\geq 0\end{aligned}$$

$$\frac{\alpha}{\tau} \;\leq\; \frac{\frac{dX_H^Z}{dP_H}\left(1-F(X_H^Z)\right)(1-\tau)}{\left(1-\tau\frac{dX_H^Z}{dP_H}\right)}$$

part b) Notice that there can't be a solution in which both H and S are unlevered, since from (3.91) $P_H^* = 0$ implies $\lambda_1 \leq -1$ which violates $\lambda_1 \geq 0$.
■

Part III

Hedging Mortality in Stochastic Mortality Models

The last part of my Ph.D. Dissertation studies the hedging problem of life insurance policies, when the mortality rate is stochastic. The field developed recently, adapting well-established techniques widely used in finance to describe the evolution of rates of mortality. The chapter is joint work with my supervisor, prof. Elisa Luciano and Elena Vigna. It studies the hedging problem of life insurance policies, when the mortality and interest rates are stochastic. We focus primarily on stochastic mortality. We represent death arrival as the first jump time of a doubly stochastic process, i.e. a jump process with stochastic intensity. We propose a Delta-Gamma Hedging technique for mortality risk in this context. The risk factor against which to hedge is the difference between the actual mortality intensity in the future and its "forecast" today, the instantaneous forward intensity. We specialize the hedging technique first to the case in which survival intensities are affine, then to Ornstein-Uhlenbeck and Feller processes, providing actuarial justifications for this restriction. We show that, without imposing no arbitrage, we can get equivalent probability measures under which the HJM condition for no arbitrage is satisfied. Last, we extend our results to the presence of both interest rate and mortality risk, when the forward interest rate follows a constant-parameter Hull and White process. We provide a UK calibrated example of Delta and Gamma Hedging of both mortality and interest rate risk.

Chapter 4

Delta and Gamma hedging of mortality and interest rate risk

4.1 Introduction

This paper studies the hedging problem of life insurance policies, when the mortality rate is stochastic. In recent years, the literature has focused on the stochastic modeling of mortality rates, in order to deal with unexpected changes in the longevity of the sample of policyholders of insurance companies. This kind of risk, due to the stochastic nature of death intensities, is referred to as systematic mortality risk[1]. In the present paper we deal with this, as well as with two other sources of risk life policies are subject to: financial risk and non-systematic mortality risk. The former originates from the stochastic nature of interest rates. The latter is connected to the randomness in the occurrence of death in the sample of insured people and disappears in well diversified portfolios.

The problem of hedging life insurance liabilities in the presence of systematic mortality risk has attracted much attention in recent years. It has been addressed either via risk-minimizing and mean-variance indifference hedging strategies, or through the creation of mortality-linked derivatives and securitization. The first approach has been taken by Dahl and Møller (2006) and Barbarin (2008). The second approach was discussed by Dahl (2004) and Cairns, Blake, Dowd, and MacMinn (2006) and has witnessed a lively debate and a number of recent improvements, see f.i. Blake, De Waegenaere, MacMinn, and Nijman (2010) and references therein.

We study Delta and Gamma hedging. This requires choosing a specific change of measure, but has two main advantages with respect to

[1]In this paper we do not distinguish between mortality and longevity risk.

risk-minimizing and mean-variance indifference strategies. On the one side it represents systematic mortality risk in a very intuitive way, namely as the difference between the actual mortality intensity in the future and its "forecast" today. On the other side, Delta and Gamma hedging is easily implementable and adaptable to self-financing constraints. It indeed ends up in solving a linear system of equations. The comparison with securitization works as follows. The Delta and Gamma hedging complements the securitization approach strongly supported by most academics and industry leaders, in two senses. On the one hand, as is known, the change of measure issue on which hedging relies will not be such an issue any more once the insurance market, thanks to securitization and derivatives, becomes liquid. On the other hand, securitization aims at one-to-one hedging or replication, while we push hedging one step further, through local, but less costly, coverage.

Following a well established stream of actuarial literature, we adapt the setting of risk-neutral interest rate modelling to represent stochastic mortality. We represent death arrival as the first jump time of a doubly stochastic process. To enhance analytical tractability, we assume a pure diffusion of the affine type for the spot mortality intensity. Namely, the process has linear affine drift and instantaneous variance-covariance matrix linear in the intensity itself.

In this setting, Cairns, Blake, and Dowd (2006) point out that the HJM no arbitrage condition typical of the financial market can be translated into an equivalent HJM-like condition for forward death intensities. Usually, the respect of the HJM condition on the insurance market is imposed a priori. We show that, for two non-mean reverting processes for the spot intensity, whose appropriateness will be discussed below, there exists an infinity of probability measures – equivalent to the historical one – in which forward death intensities satisfy an HJM condition. No arbitrage holds under any of these measures, even though it is not imposed a priori. These processes belong to the Ornstein Uhlenbeck and the Feller class.

As a consequence, we start by introducing the spot mortality intensities, discuss their soundness as descriptors of the actual – or historical – mortality dynamics, derive the corresponding forward death intensities and tackle the change of measure issue. Among the possible changes, we select the minimal one – which permits to remain in the Ornstein-Uhlenbeck and Feller class – and parameterize it by assuming that the risk premium for mortality risk is constant. By so doing, we can avoid using risk minimizing or mean-variance indifference strategies. We can instead focus on Delta and Gamma hedging. For the sake of simplicity we assume that the market of interest rate bonds is not only arbitrage-free but also complete. First, we consider a pure endowment hedge in the presence of systematic mortality risk only. Then, under

independence of mortality and financial risks, we provide an extension of the hedging strategy to both these risks.
To keep the treatment simple, we build Delta and Gamma coverage on pure endowments, using as hedging tools either pure endowments or zero-coupon survival bonds for mortality risk and zero-coupon-bonds for interest rate risk. Since all these assets can be understood as Arrow-Debreu securities – or building blocks – in the insurance and fixed income market, the Delta and Gamma hedge could be extended to more complex and realistic insurance and finance contracts.
In spite of our restriction to pure endowments, the final calibration of the strategies – which uses UK mortality rates for the male generation born in 1945 and the Hull-White interest rates on the UK market – shows that

1. the unhedged effect of a sudden change on mortality rate is remarkable, especially for long time horizons;

2. the corresponding Deltas and Gammas are quite different if one takes into consideration or ignores the stochastic nature of the death intensity:

3. the hedging strategies are easy to implement and customize to self-financing constraints;

4. Delta and Gamma are bigger for mortality than for financial risk.

The paper is structured as follows: Section 4.2 recalls the doubly stochastic approach to mortality modelling and introduces the two intensity processese considered in the paper. Section 4.3 presents the notion of forward death intensity. Section 4.4 describes the standard financial assumptions on the market for interest rates. Section 4.5 derives the dynamics of forward intensities and survival probabilities, after the appropriate change of measure. Section 4.6 shows that the HJM restriction is satisfied without imposing no arbitrage a priori. In Section 4.7 we discuss the hedging technique for mortality risk. Section 4.8 addresses mortality and financial risk. Section 4.9 presents the application to a UK population sample. Section 4.10 summarizes and concludes.

4.2 Cox modelling of mortality risk

This Section introduces mortality modelling by specifying the so-called spot mortality intensity (mortality intensity for short). Section 4.2.1 describes the general framework, while Section 4.2.2 studies two specific processes which will be considered troughout the paper.

4.2.1 Instantaneous death intensity

Mortality in the actuarial literature has been recently described by means of Cox or doubly stochastic counting processes, as studied by Brémaud (1981). The modelling technique has been drawn from the financial domain and in particular from the reduced form models of the credit risk literature, where the time to default is described as the first stopping time of a Cox process [2]. In the actuarial literature, mortality modelling via Cox processes has been introduced by Milevsky and Promislow (2001) and Dahl (2004). Intuitively, the time to death - analogously to the time to default in finance - is supposed to be a Poisson process with stochastic intensity. The intensity process may be either a pure diffusion or may present jumps. If in addition it is an affine process, then the survival function can be derived in closed form.
Let us introduce a filtered probability space $(\Omega, \mathbf{F}, \mathbb{P})$, equipped with a filtration $\{\mathcal{F}_t : 0 \leq t \leq T\}$ which satisfies the usual properties of right-continuity and completeness. On this space, let us consider a non negative, predictable process λ_x, which represents the mortality intensity of an individual or head belonging to generation x at (calendar) time t. We introduce the following

Assumption 1 The mortality intensity λ_x follows a process of the type:

$$d\lambda_x(t) = a(t, \lambda_x(t))dt + \sigma(t, \lambda_x(t))dW_x(t) + dJ_x(t) \tag{4.1}$$

where J is a pure jump process, W_x is a standard one-dimensional Brownian motion[3] and the regularity properties for ensuring the existence of a strong solution of equation (4.1) are satisfied for any given initial condition $\lambda_x(0) = \lambda_0 > 0$.

The existence of a stochastic mortality intensity generates systematic mortality risk. Given this assumption on the dynamics of the death intensity, let τ be the time to death of an individual of generation x. We define the survival probability from t to $T \geqslant t$, $S_x(t,T)$, as the survival function of the time to death τ under the probability measure $\mathbb{P}$, conditional on the survival up to time t:

$$S_x(t,T) := \mathbb{P}\left(\tau \geqslant T \mid \tau > t\right)$$

It is known since Brémaud (1981) that - under the previous assumption - the survival probability $S_x(t,T)$ can be represented as

$$S_x(t,T) = \mathbb{E}\left[\exp\left(-\int_t^T \lambda_x(s)ds\right) \mid \mathcal{F}_t\right] \tag{4.2}$$

[2]See the seminal paper Lando (1998).

[3]The extension of the mortality intensity definition to a multidimensional Brownian motion is straightforward.

where the expectation is computed under $\mathbb{P}$ and is evidently conditional on $\mathcal{F}_t$. When the evaluation date is zero ($t = 0$), we simply write $S_x(T)$ instead of $S_x(0, T)$.

In this paper, we suppose in addition that

Assumption 2 the drift $a(t, \lambda(t))$, the instantaneous variance-covariance coefficient $\sigma^2(t, \lambda(t))$ and the jump measure η associated with J, which takes values in $\mathbb{R}^+$, have affine dependence on $\lambda(t)$.

Hence, we assume that these coefficients are of the form:

$$a(t, \lambda(t)) = b + c\lambda(t)$$

$$\sigma^2(t, \lambda(t)) = d \cdot \lambda(t)$$

$$\eta(t, \lambda(t)) = l_0 + l_1\lambda(t)$$

where $b, c, d, l_0, l_1 \in \mathbb{R}$. Under this assumption standard results on functionals of affine processes allow us to state that

$$S_x(t, T) = e^{\alpha(T-t)+\beta(T-t)\lambda_x(t)}$$

where α and β solve the following Riccati differential equations (see for instance Duffie and K.Singleton (2000)):

$$\beta'(t) = \beta(t)c + \frac{1}{2}\beta(t)^2 d^2 + l_1 \left[\int_{\mathbb{R}} e^{\beta(t)z} d\nu(z) - 1 \right]$$

$$\alpha'(t) = \beta(t)b + l_0 \left[\int_{\mathbb{R}} e^{\beta(t)z} d\nu(z) - 1 \right]$$

where ν is the distribution function of the jumps of J. The boundary conditions are $\alpha(0) = 0$ and $\beta(0) = 0$.

4.2.2 Ornstein-Uhlenbeck and Feller processes

In this paper we focus on two intensity processes, which belong to the affine class and are purely diffusive. These processes, together with the solutions α and β of the associated Riccati ODEs, are:

— Ornstein-Uhlenbeck (OU) process without mean reversion:

$$d\lambda_x(t) = a\lambda_x(t)dt + \sigma dW_x(t) \tag{4.3}$$

$$\alpha(t) = \frac{\sigma^2}{2a^2}t - \frac{\sigma^2}{a^3}e^{at} + \frac{\sigma^2}{4a^3}e^{2at} + \frac{3\sigma^2}{4a^3} \tag{4.4}$$

$$\beta(t) = \frac{1}{a}(1 - e^{at}) \tag{4.5}$$

— Feller Process (FEL) without mean reversion:

$$
\begin{aligned}
d\lambda_x(t) &= a\lambda_x(t)dt + \sigma\sqrt{\lambda_x(t)}dW_x(t) && (4.6)\\
\alpha(t) &= 0 && (4.7)\\
\beta(t) &= \frac{1-e^{bt}}{c+de^{bt}} && (4.8)
\end{aligned}
$$

with $b = -\sqrt{a^2 + 2\sigma^2}, c = \frac{b+a}{2}, d = \frac{b-a}{2}$. Here, we assume $a > 0, \sigma \geq 0$.

A process, in order to describe human survivorship realistically, has to be "biologically reasonable", i.e. it has to satisfy two technical features: the intensity must never be negative and the survival function has to be decreasing in time T.
In the OU case, λ can indeed turn negative, with positive probability:

$$
u = \mathbb{P}(\lambda(t) \leqslant 0) = \phi\left(-\frac{\lambda(0)e^{at}}{\sigma\sqrt{\frac{e^{2at}-1}{2a}}}\right)
$$

where ϕ is the distribution function of the standard normal. The survival function is always decreasing when time T is below a certain level T^*:

$$
T < T^* = \frac{1}{a}ln\left[1 + \frac{a^2\lambda(0)}{\sigma^2}\left(1 + \sqrt{1 + \frac{2\sigma^2}{a^2\lambda(0)}}\right)\right] \tag{4.9}
$$

In practical applications (section 4.9) we verify that the probability u is negligible and that the length of the time horizon we consider (the duration of a human life) never exceeds T^*.
For the FEL process, instead, the intensity can never turn negative and the survival function is guaranteed to be decreasing in T if and only if the following condition holds:

$$
e^{bt}(\sigma^2 + 2d^2) > \sigma^2 - 2dc. \tag{4.10}
$$

We verify this condition, which is satisfied whenever $\sigma^2 - 2dc < 0$, for our calibrated parameters (see section 4.9).

In spite of the technical restrictions, Luciano and Vigna (2008) and Luciano, Spreeuw, and Vigna (2008) suggest the appropriateness of these processes for describing the intensity of human mortality. In fact, they show that these models meet all but one of the criteria - motivated by Cairns, Blake, and Dowd (2006) - that a good mortality model should meet:

1. the model should be consistent with historical data: the calibrations of Luciano and Vigna (2008) show that the models meet this criterium;

2. the force of mortality should keep positive: the first model does not meet this criterium; however, the probability of negative values of the intensity is shown to be negligible for practical applications;

3. long-term future dynamics of the model should be biologically reasonable: the models meet this criterium, as the calibrated parameters satisfy conditions (4.9) and (4.10) above;

4. long-term deviations in mortality improvements should not be mean-reverting to a pre-determined target, even if the target is time-dependent: the models meet this criterium by construction;

5. the model should be comprehensive enough to deal appropriately with pricing valuation and hedging problem: these models meet this criterium, since it is straightforward to extend them in order to deal with pricing, valuation and hedging problems; this is indeed the scope of the present paper;

6. it should be possible to value mortality linked derivatives using analytical methods or fast numerical methods: these models meet this criterium, as they produce survival probabilities in closed form and with a very small number of parameters.

Cairns, Blake, and Dowd (2006) add that no one of the previous criteria dominates the others. Consistently with their view, we claim the validity of the proposed models, which meet five criteria out of six. The violation of the second criterium above in the OU case is the price paid in order to have a simple and parsimonious model. Notice though that this is only a theoretical limit of the model, as a negative force of mortality has a negligible probability of occurring in practical applications. In addition, the fact that survival functions are given in closed form and depend on a very small number of parameters simplifies the calibration procedure enormously. Last but not least, these two processes are natural stochastic generalizations of the Gompertz model for the force of mortality and, thus, they are easy to interpret in the light of the traditional actuarial practice.
These processes (and especially the first one, the OU) turn out to be significantly suitable for the points 5 and 6 above. In fact, in Sections 6, 7 and 8 we will show that the Delta and Gamma OU-coefficients can be expressed in a very simple closed form. Thus, the Delta-Gamma Hedging technique – widely used in the financial context to hedge purely financial assets – turns

out to be remarkably easy to apply. This feature renders quite applicable this hedging technique also in the actuarial-financial context. The Delta and Gamma FEL coefficients are more complicated to find, but the technique is still applicable.

4.3 Forward death intensities

This Section aims at shifting from mortality intensities to their forward counterparts, both for the general affine case and for the OU and FEL processes. The notion of forward instantaneous intensity for counting processes representing firm defaults has been introduced by Duffie (1998) and Duffie and Singleton (1999), following a discrete-time definition in Litterman and Iben (1991). Stochastic modelling of this quantity has been extensively studied in the financial domain. In the credit risk domain indeed the notion of forward intensity is very helpful, since it allows to determine the change of measure or the intensity dynamics useful for pricing and hedging defaultable bonds (the characterization is obtained under a no arbitrage assumption for the financial market and is unique when the market is also complete).
Suppose that arbitrages are ruled out, that the recovery rate is null and $\lambda_x(t)$ in (4.1) represents the default intensity of a firm whose debt is traded in a complete market. Then, we would have the following HJM restriction under the (unique) risk-neutral measure corresponding to $\mathbb{P}$:

$$a(t, \lambda_x(t)) = \sigma(t, \lambda_x(t)) \int_0^t \sigma(u, \lambda_x(u))du \tag{4.11}$$

In the actuarial domain, forward death intensities have already been introduced by Dahl (2004) and Cairns, Blake, and Dowd (2006), paralleling the financial definition. In section 4.5 we prove that, even though the restriction (4.11) can be violated by death intensities in general, it holds true for the OU and FEL intensity processes, even without imposing no arbitrage, but simply restricting the measure change so that the intensity remains OU or FEL under the new measures.

Let us start from the forward death rate over the period $(t, t+\Delta t)$, evaluated at time zero, as the ratio between the conditional probability of death between t and $t+\Delta t$ and the time span Δt, for a head belonging to generation x, conditional on the event of survival until time t:

$$\frac{1}{\Delta t}\left(\frac{S_x(t) - S_x(t+\Delta t)}{S_x(t)}\right)$$

Let us consider its instantaneous version, which we denote as $f_x(0,t)$. We refer to it as to forward death intensity. It is evident from its definition that - if it exists - the forward death intensity is the logarithmic derivative of the (unconditional) survival probability, as implied by the process λ:

$$f_x(0,t) := \lim_{\Delta t \to 0} \frac{1}{\Delta t}\left(1 - \frac{S_x(t+\Delta t)}{S_x(t)}\right) = -\frac{\partial}{\partial t}\ln\left(S_x(t)\right)$$

The similarity of this definition with the force of mortality is quite strong.[4] Similarly, one can define the forward death intensity for the tenor T, as evaluated at time $t < T$, starting from the survival probability $S_x(t,T)$:

$$f_x(t,T) = -\frac{\partial}{\partial T}\ln\left(S_x(t,T)\right) \tag{4.12}$$

The forward death intensity $f_x(t,T)$ represents the intensity of mortality which will apply instantaneously at time $T > t$, implied by the knowledge of the process λ up to t (or under the filtration $\mathcal{F}_t$). This explains the dependence of f_x on the current date t as well as on the future one, T. It can be interpreted as the "best forecast" of the actual mortality intensity, since it coincides with the latter when $T = t$:

$$f_x(t,t) = \lambda_x(t)$$

Please notice also that the forward death intensity definition, and consequently its expression for the affine case, is analogous to the one of forward instantaneous interest rates, the latter being defined starting from discount factors rather than survival probabilities. As in the case of forward instantaneous interest rates, it can be shown that forward intensities, for given t, can be increasing, decreasing or humped functions of the application date T. It follows from the above definition that the survival probabilities from t to $T > t$ can be written as integrals of (deterministic) forward death probabilities:

$$S_x(t,T) = \exp\left(-\int_t^T f_x(t,s)ds\right) \tag{4.13}$$

and not only as expectations wrt the intensity process λ , as in (4.2) above.[5]

[4] The two concepts coincide when the diffusion coefficient of the intensity process is null.

[5] Notice that, at any initial time t, forward death intensities can be interpreted as the (inhomogeneous) Poisson arrival rates implied in the current Cox process. Indeed, it is quite natural, especially if one wants a description of survivorship without the mathematical complexity of Cox processes, to try to describe mortality via the equivalent survival

Let us turn now to the affine case. As it can be easily shown from (4.13), when λ is an affine process the initial forward intensity depends on the functions α and β:

$$f_x(0,t) = -\alpha'(t) - \beta'(t)\lambda_x(0) = -\alpha'(t) - \beta'(t) f_x(0,0) \tag{4.14}$$

and at any time $t \geq T \geq 0$:

$$f_x(t,T) = -\alpha'(T-t) - \beta'(T-t)\lambda_x(t) = -\alpha'(T-t) - \beta'(T-t) f_x(t,t)$$

For the processes defined by equations (4.3) and (4.6), the instantaneous forward intensities can be computed as:

$$OU \qquad f_x(t,T) = \lambda_x(t)e^{a(T-t)} - \frac{\sigma^2}{2a^2}(e^{a(T-t)} - 1)^2 \tag{4.15}$$

$$FEL \qquad f_x(t,T) = \frac{4\lambda_x(t)b^2 e^{b(T-t)}}{[(a+b)+(b-a)e^{b(T-t)}]^2}$$

4.4 Financial risk

In order to introduce a valuation framework for insurance policies, we need to provide provide a description of the financial environment. In addition to mortality risk, we assume the existence of a financial risk, in the sense that the interest rate is described by a stochastic process. While in the mortality domain we started from (spot) intensities – for which we were able to motivate specific modelling choices – and then we went to their forward counterpart, here we follow a well established bulk of literature – starting from Heath, Jarrow, and Morton (1992) – and model directly the instantaneous forward rate $F(t,T)$, i.e. the date-t rate which applies instantaneously at T .

Assumption 3 The process for the forward interest rate $F(t,T)$, defined on the probability space $(\Omega, \mathbf{F}, \mathbb{P})$, is:

$$dF(t,T) = A(t,T)dt + \Sigma(t,T)dW_F(t) \tag{4.16}$$

probability in a simpler (inhomogeneous) Poisson model. Once a λ process has been fixed, and therefore survival probabilities have been computed, according to (4.2), one can wonder: what would be the intensity of an inhomogeneous Poisson death arrival process, that would produce the same survival probabilities? Recalling that in the Poisson case survival probabilities are of the type (4.13), one can interpret – and use – $f(t,T)$ exactly as the survival intensity of an (inhomogeneous) Poisson model equivalent to the given, Cox one.

where the real functions $A(t,T)$ and $\Sigma(t,T)$ satisfy the usual assumptions for the existence of a strong solution to (4.16), and W_F is a univariate Brownian motion[6] independent of W_x for all x.

The independence between the Brownian motions means, loosely speaking, independence between mortality and financial risk.[7]
Let us also denote as $\{\mathcal{H}_t : 0 \leq t \leq T\}$ the filtration generated by the interest rate process. As a particular subcase of the forward rate, obtained when $t = T$, one obtains the short rate process, which we will denote as $r(t)$:

$$F(t,t) := r(t) \tag{4.17}$$

It is known that, when the market is assumed to admit no arbitrages and be complete, there is a unique martingale measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ - which we will characterize in the next section - under which the zero-coupon-bond price for the maturity T, evaluated at time t, $B(t,T)$, is

$$B(t,T) = \exp\left(-\int_t^T F(t,u)du\right) = \mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_t^T r(u)du\right)\right] \tag{4.18}$$

We will provide a specific choice for the forward interest rate only at a later stage. We will have no need to motivate it, since it corresponds to a very popular model in Finance, the one-factor Hull and White (Hull and White (1990)).

4.5 Change of measure and insurance valuations

This section discusses the change of measure that allows us to compute the prices of policies subject to mortality risk in a fashion analogous to (4.18). First, we define the process of death occurrence inside the sample of insured people of interest. As in Dahl and Møller (2006), we represent it as follows. Let $\tau_1, \tau_2, ...\tau_N$ be the lifetimes of the N insured in the cohort x, assumed to be i.i.d. with distribution function $S_x(t,T)$ in (4.2). Let $M(x,t)$ be the

[6] We assume a single Brownian motion for the forward rate dynamics, since we reduced the discussion of mortality risk to a single risk source too: however, the extension to a multidimensional Brownian motion is immediate.

[7] This assumption is common in the literature and seems to be intuitively appropriate. See Miltersen and Persson (2006) for a setting in which mortality and financial risks can be correlated.

(pure jump) process which counts the number of deaths in such an insurance portfolio:

$$M(x,t) := \sum_{i=1}^{N} 1_{\{\tau_i \leq t\}}$$

where 1 is the indicator function. We define a filtration on $(\Omega, \mathbf{F}, \mathbb{P})$ whose σ -algebras $\{\mathcal{G}_t : 0 \leq t \leq T\}$ are generated by $\mathcal{F}_t$ and $\{M(x,s) : 0 \leq s \leq t\}$. This filtration intuitively collects the information on both the past mortality intensity and on actual death occurrence in the portfolio. Let us consider, on the probability space $(\Omega, \mathbf{F}, \mathbb{P})$, the sigma algebras $\mathcal{I}_t := \mathcal{G}_t \vee \mathcal{H}_t$ generated by unions of the type $\mathcal{G}_t \cup \mathcal{H}_t$, where the σ-algebra $\mathcal{G}_t$ collects information on the mortality intensity and actual death process, while $\mathcal{H}_t$, which is independent of $\mathcal{G}_t$, reflects information on the financial market, namely on the forward rate process The filtration $\{\mathcal{I}_t : 0 \leq t \leq T\}$ therefore represents all the available information on both financial and mortality risk. In order to perform insurance policies evaluations in $(\Omega, \mathbf{F}, \mathbb{P})$, equipped with such a filtration, we need to characterize at least one equivalent measure. This can be done using a version of Girsanov's theorem, as in Jacod and Shiryaev (1987)[8]:

Theorem 4.5.1 *Let the bi-dimensional process* $\theta(t) := [\theta_x(t) \quad \theta_F(t)]$ *and the univariate, positive one* $\varphi(t)$ *be predictable, with*

$$\begin{aligned} \int_0^T \theta_x^2(t)dt &< \infty, \\ \int_0^T \theta_F^2(t)dt &< \infty, \\ \int_0^T |\varphi(t)| \, \lambda_x(t)dt &< \infty \end{aligned}$$

Define the likelihood process $L(t)$ *by*

$$\begin{cases} L(0) = 1 \\ \frac{dL(t)}{L(t-)} = \theta_x(t)dW_x(t) + \theta_F(t)dW_F(t) + (\varphi(t) - 1)\, dM(x,t) \end{cases}$$

and assume $\mathbb{E}^{\mathbb{P}}[L(t)] = 1, t \leq T$. *Then there exists a probability measure* $\mathbb{Q}$ *equivalent to* $\mathbb{P}$, *such that the restrictions of* $\mathbb{P}$ *and* $\mathbb{Q}$ *to* $\mathcal{I}_t$, $\mathbb{P}_t := \mathbb{P} \mid \mathcal{I}_t$, $\mathbb{Q}_t := \mathbb{Q} \mid \mathcal{I}_t$, *have Radon-Nykodim derivative* $L(t)$:

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = L(t)$$

[8] See also Dahl and Møller (2006) for an application to a similar actuarial setting.

The mortality indicator process has intensity $\varphi(t)\lambda_x(t)$ *under* $\mathbb{Q}$ *and*

$$\begin{aligned} dW'_x &: = dW_x - \theta_x(t)dt \\ dW'_F &: = dW_F - \theta_F(t)dt \end{aligned}$$

define $\mathbb{Q}-$*Brownian motions. All the probability measures equivalent to* $\mathbb{P}$ *can be characterized this way.*

Actually, the previous theorem characterizes an infinity of equivalent measures, depending on the choices of the processes $\theta_x(t), \theta_F(t)$ and $\varphi(t)$. These processes represent the prices - or premia - given to the three different sources of risk we model.

The first source of risk, the systematic mortality one, is represented by $\theta_x(t)$. This source of risk is not diversifiable, since it originates from the randomness of death intensity. We have no standard choices to apply in the choice of $\theta_x(t)$, see for instance the extensive discussion in Biffis (2005) and Cairns, Blake, Dowd, and MacMinn (2006). For the sake of analytical tractability, as in Dahl and Møller (2006), we restrict it so that the risk-neutral intensity is still affine. Therefore, we substitute Assumptions 1 and 2 with the following

Assumption 4 The intensity process under $\mathbb{P}$ is purely diffusive and affine. The systematic mortality risk premium is such as to leave it affine under $\mathbb{Q}$:

$$\theta_x(t) := \frac{p(t) + q(t)\lambda_x(t)}{\sigma(t, \lambda_x(t))}$$

with $p(t)$ and $q(t)$ continuous functions of time.

Indeed, with such a risk premium, the intensity process under $\mathbb{Q}$ is

$$d\lambda_x(t) = [a(t, \lambda_x(t)) + p(t) + q(t)\lambda_x(t)]\, dt + \sigma(t, \lambda_x(t))dW'_x. \tag{4.19}$$

which is still affine. This choice boils down to selecting the so-called minimal martingale measure. It can be questioned – as any other choice – but proves to be very helpful for hedging.[9] For the OU and FEL processes we choose the functions $p = 0$ and q constant, so that we have the same type of process under $\mathbb{P}$ and $\mathbb{Q}$, with the coefficient a in equations (4.3) and (4.6) replaced by $a' := a + q$.

The second source of risk, the financial one, originates from the stochastic nature of interest rates. The process $\theta_F(t)$ represents the so called premium

[9] Its calibration will be straightforward, as soon as the market for mortality derivatives becomes liquid enough.

for financial risk. Assume that the financial market is complete. The only choice consistent with no arbitrage is

$$\theta_F(t) := -A(t,T)\Sigma^{-1}(t,T) + \int_t^T \Sigma(t,u)du$$

Under this premium indeed the drift coefficient of the forward dynamics $A'(t,T)$ is tied to the diffusion by an HJM relationship:

$$A'(t,T) = \Sigma(t,T)\int_t^T \Sigma(t,u)du \tag{4.20}$$

It follows that, under the measure $\mathbb{Q}$,

$$dF(t,T) = \left[\Sigma(t,T)\int_t^T \Sigma(t,u)du\right]dt + \Sigma(t,T)dW_F'(t) \tag{4.21}$$

The time-t values of the forward and short rate are respectively (see f.i. Shreve (2004)):

$$\begin{aligned} F(t,T) &= F(0,T) + \int_0^t \Sigma(s,T)\int_s^T \Sigma(s,m)dmds + \int_0^t \Sigma(u,T)dW_F'(u) \\ r(t) &= F(0,t) + \int_0^t \Sigma(s,T)\int_s^t \Sigma(s,m)dmds + \int_0^t \Sigma(u,t)dW_F'(u) \end{aligned} \tag{4.22}$$

where

$$\begin{aligned} dW_F' &= dW_F - \theta_F(t)dt \text{ and} \\ \theta_F(t) &= -A(t,T)\Sigma^{-1}(t,T) + \int_t^T \Sigma(t,u)du. \end{aligned}$$

The third source of risk is the non systematic mortality one, arising from the randomness of death occurrence inside the portfolio of insured people. In the presence of well diversified insurance portfolios, insurance companies are uninterested in hedging this idiosyncratic component of mortality risk, since the law of large numbers is expected to apply. Hence, we assume that the market gives no value to it and we make the following assumption for $\varphi(t)$, which represents the premium for idiosyncratic mortality risk:

Assumption 5 $\varphi(t) = 0$ for every t

The fair premium and the reserves of life insurance policies can be computed as expected values under the measure $\mathbb{Q}$.
Consider the case of a pure endowment contract[10] starting at time 0 and paying one unit of account if the head x is alive at time T. The fair premium or price of such an insurance policy, given the independence between the financial and the actuarial risk, is:

$$P(0,T) = S_x(T)B(0,T) = e^{\alpha(T)+\beta(T)\lambda_x(0)} E_{\mathbb{Q}}\left[-\exp\left(\int_0^T r(u)du\right)\right]$$

The value of the same policy at any future date t is:

$$\begin{aligned} P(t,T) &= S_x(t,T)B(t,T) \\ &= E_{\mathbb{Q}}\left[\exp\left(-\int_t^T \lambda(s)ds\right)\right] E_{\mathbb{Q}}\left[-\exp\left(\int_0^T r(u)du\right)\right] \quad (4.23) \end{aligned}$$

Hence, we can define a "term structure of pure endowment contracts". The last expression, net of the initial premium, is also the time t reserve for the policy, which the insurance company will be interested in hedging. Notice that we did not impose no arbitrage on the market for these instruments. Once the change of measure has been performed, we can write $P(0,T)$ in terms of the instantaneous forward probability and interest rate (f and F respectively):

$$P(t,T) = \exp\left(-\int_t^T \left[f_x(t,u) + F(t,u)\right] du\right)$$

4.6 HJM restriction on forward death intensities

In this section we show that, if the risk premium for mortality is constant, then the OU and FEL processes for mortality intensity satisfy an HJM-like restriction on the drift and diffusion. This is important, since proving that the HJM condition holds is equivalent to showing that no arbitrage holds, without having assumed it to start with. We keep the head x fixed, and in the notation we drop the dependence on x.

Forward death intensities, being defined as log derivatives of survival probabilities, follow a stochastic process. This process can be derived starting

[10] We do this recognizing that more complex policies or annuities can be decomposed into these basic contracts.

from the one of the survival probabilities themselves, recalling that the process λ is given by (4.19). Under Assumption 4, Ito's lemma implies that the functional S follows the process:

$$dS(t,T) = S(t,T)m(t,T)dt + S(t,T)n(t,T)dW'(t)$$

where

$$\begin{aligned} m(t,T) &= \frac{1}{S}\left[\frac{\partial S}{\partial t} + \frac{\partial S}{\partial \lambda}\left[a(t,\lambda) + p(t) + q(t)\lambda(t)\right] + \frac{1}{2}\frac{\partial^2 S}{\partial \lambda^2}\sigma^2(t,\lambda)\right] \\ n(t,T) &= \frac{1}{S}\frac{\partial S}{\partial \lambda}\sigma(t,\lambda) \end{aligned}$$

The forward death intensity $f(t,T)$, defined as the logarithmic derivative of $S(t,T)$, can be shown to follow the dynamics:

$$df(t,T) = v(t,T)dt + w(t,T)dW'(t) \tag{4.24}$$

where the drift and diffusion coefficients are:

$$v(t,T) = \frac{\partial n(t,T)}{\partial T}n(t,T) - \frac{\partial m(t,T)}{\partial T} \tag{4.25}$$

$$w(t,T) = -\frac{\partial n(t,T)}{\partial T} \tag{4.26}$$

Since – according to the Assumption 4 – the intensity process is of the affine class, the drift and diffusion of the survival probabilities are

$$\begin{aligned} m(t,T) &= -\alpha'(T-t) - \beta'(T-t)\lambda(t) + \\ &+ \left[a(t,\lambda) + p(t) + q(t)\lambda(t)\right]\beta(T-t) + \frac{1}{2}\sigma^2(t,\lambda)\beta^2(T-t) \\ n(t,T) &= \sigma(t,\lambda)\beta(T-t) \end{aligned}$$

Given that, one can easily derive the forward intensity process coefficients:

$$\begin{aligned} v(t,T) &= \alpha''(T-t) + \beta''(T-t)\lambda(t) - \left[a(t,\lambda) + p(t) + q(t)\lambda\right]\beta'(T-t) \\ w(t,T) &= -\sigma(t,\lambda)\beta'(T-t) \end{aligned}$$

In general, the forward dynamics then depends on the drift and diffusion coefficients of the mortality intensity and on the properties of the solutions of the Riccati equations. One can wonder whether - starting from a mortality intensity process - an HJM-like condition, which works on the forward survival intensities,

$$v(t,T) = w(t,T)\int_t^T w(t,s)ds \tag{4.27}$$

is satisfied. We provide the following:

Theorem 4.6.1 *Let λ be a purely diffusive process which satisfies Assumption 4. Then, the HJM condition (4.27) is satisfied if and only if:*

$$\frac{\partial m(t,T)}{\partial T} = n(t,t)\frac{\partial n(t,T)}{\partial T}.$$

In particular, this condition is satisfied in the cases of the Ornstein-Uhlenbeck process (4.3) and of the Feller process (4.6) with $p = 0$ and q constant.

Proof.

Using (4.26), we get the r.h.s. of the HJM condition (4.27):

$$w(t,T)\int_t^T w(t,s)ds = \frac{\partial n(t,T)}{\partial T}(n(t,T) - n(t,t)).$$

Hence, plugging (4.25) into the HJM condition (4.27) we get [11]:

$$\frac{\partial m(t,T)}{\partial T} = n(t,t)\frac{\partial n(t,T)}{\partial T}.$$

As for the second part, if the intensity follows an OU process, the forward probability f satisfies the HJM condition (4.27). This result is a straightforward consequence of the fact that – with $p = 0$ and q constant – the functions α and β of the OU process satisfy the system of ODEs':

$$\begin{cases} \beta'(t) = -1 + a'\beta(t) \\ \alpha'(t) = \frac{1}{2}\sigma^2\beta^2(t) \end{cases} \tag{4.28}$$

with the boundary conditions $\alpha(0) = 0$ and $\beta(0)$=0. In fact,

$$\begin{aligned} v(t,T) &= \alpha''(T-t) + \beta''(T-t)\lambda(t) - \beta'(T-t)a'\lambda(t) \\ &= \sigma^2\beta(T-t)\beta'(T-t) \\ w(t,T) &= -\sigma\beta'(T-t). \end{aligned}$$

and property (4.27) is satisfied.

Consider now the Feller process (4.6) and its well-known solution to the Riccati ODE:

$$\begin{cases} \alpha'(t) = 0 \\ \beta'(t) = -1 + a'\beta(t) + \frac{1}{2}\sigma^2\beta^2(t) \end{cases} \tag{4.29}$$

Again, we can easily show that condition (4.27) is satisfied, since

$$v(t,T) = \beta''(T-t)\lambda(t) - a'\lambda(t)\beta'(T-t) = \sigma^2\beta(T-t)\beta'(T-t)\lambda(t)$$

[11]Notice that a similar condition on the drift and diffusion of spot interest rates is in Shreve (2004).

$$w(t,T) = -\sigma(t,\lambda)\beta'(T-t) = -\sigma\sqrt{\lambda(t)}\beta'(T-t).$$

The HJM condition is a characterizing feature of some models for interest rates such as the Vasicek (1977), Hull and White (1990), the CIR (Cox, Ingersoll Jr, and Ross (1985)). It is well known that the HJM condition (4.27), applied to the coefficients of the interest rate process, as in (4.20), is equivalent to the absence of arbitrage. In our case, since we showed that - under Assumption 4 - the OU and FEL processes satisfy the HJM condition, arbitrage is ruled out without being imposed. Please notice that the dynamics of the forward intensity under for the OU case $\mathbb{Q}$ is

$$df(t,T) = \frac{\sigma^2}{a'} e^{a'(T-t)} \left(e^{a'(T-t)} - 1 \right) dt + \sigma e^{a'(T-t)} dW'(t). \tag{4.30}$$

It reminds of the Hull and White dynamics for forward interest rates, when the parameters are constant.

4.7 Mortality risk hedging

In order to study the hedging problem of a portfolio of pure endowment contracts, we assume first that the interest rate is deterministic and, without loss of generality, equal to zero. This allows us to focus in this Section on the hedging of systematic mortality risk only. At a later stage, we will introduce again financial risk (section 4.8) and study the problem of hedging both mortality and financial risk simultaneously.
Once the risk-neutral measure $\mathbb{Q}$ has been defined, in order to introduce an hedging technique for systematic mortality risk we need to derive the dynamics of the reserve, which represents the value of the policy for the issuer (assuming that the unique premium has already been paid). We do this, for the sake of simplicity, assuming an OU behavior for the intensity.

4.7.1 Dynamics and sensitivity of the reserve

Affine intensity

Let us integrate (4.24), to obtain the forward death probability:

$$f(t,T) \quad = \quad f(0,T) + \int_0^t [v(u,T)du + w(u,T)dW'(u)] \tag{4.31}$$

Substituting it into the survival probability (4.13) and recalling that we write $S(u)$ for $S(0,u)$, we obtain an expression for the future survival probability

$S(t,T)$ in terms of the time-zero ones:

$$S(t,T) \quad = \quad \frac{S(T)}{S(t)}\left[\exp - \int_t^T \int_0^z \left[v(u,T)du + w(u,T)dW'(u)\right] dz\right]$$

Considering the expressions for v and w under Assumption 4, we have:

$$\begin{aligned} S(t,T) \quad &= \quad \frac{S(T)}{S(t)}[\exp - \int_t^T \int_0^z \{\alpha''(T-u) + \beta''(T-u)\lambda(u) + \\ &- \quad \beta'(T-u)\left[a(u,\lambda) + p(u) + q(u)\lambda(u)\right]]du + \\ &- \quad \beta'(T-u)\sigma(u,\lambda)dW'(u)]\}dz]. \end{aligned}$$

OU and FEL intensities

We focus now on the OU intensity case. We derive the expression for the forward survival probabilities integrating the dynamics (4.30):

$$\begin{aligned} f(t,T) \quad &= \quad f(0,T) - \frac{\sigma^2}{2a'^2}\{e^{2a'T}\left[e^{-2a't} - 1\right] - 2e^{a'T}\left[e^{-a't} - 1\right]\} \\ &+ \quad \sigma \int_0^t e^{a'(T-s)} dW(s) \end{aligned}$$

Hence, the reserve can be written simply as

$$P(t,T) = S(t,T) = \frac{S(T)}{S(t)} \exp\left[-X(t,T)I(t) - Y(t,T)\right] \tag{4.32}$$

where[12]

$$X(t,T) = \frac{\exp(a'(T-t)) - 1}{a'}$$

$$Y(t,T) = -\sigma^2[1 - e^{2a'(T-t)}]X(t,T)^2/(4a')$$

$$I(t) := \lambda(t) - f(0,t)$$

We have therefore provided an expression for the future survival probabilities - and reserves - in terms of deterministic quantities (X, Y) and of a stochastic term $I(t)$, defined as the difference between the actual mortality intensity at time t and its forecast today $f(0,t)$. $I(t)$, therefore, represents the systematic mortality risk factor. Let us notice that, as in the corresponding bond expressions of HJM, the risk factor is unique for all the survival probabilities from t onwards, no matter which horizon $T - t$ they cover. Moreover, as Cairns, Blake, and Dowd (2006) point out, if we extend our

[12]Notice that $-X(t,T) = \beta$ as soon as $a = a'$.

framework across generations and model the risk factor as an n dimensional Brownian motion, we obtain that the HJM condition is satisfied for each cohort. Applying Ito's lemma to the survival probabilities, considered as functions of time and the risk factor, we have:

$$dPdS \simeq \frac{\partial S}{\partial t}dt + \frac{\partial S}{\partial I}dI + \frac{1}{2}\frac{\partial^2 S}{\partial I^2}(dI)^2$$

It follows that the hedging coefficients for mortality risk are

$$\frac{\partial S}{\partial I} = -S(t,T)X(t,T) \leq 0 \tag{4.33}$$

$$\frac{\partial^2 S}{\partial I^2} = S(t,T)X^2(t,T) \geq 0 \tag{4.34}$$

or, for given t,

$$\frac{dP(t,T)}{P(t,T)} \simeq -X(t,T)dI + \frac{1}{2}X(t,T)^2(dI)^2$$

We denote (4.33) as Delta(Δ^M) and (4.34) Gamma (Γ^M), where the superscript M indicates that the coefficient refers to mortality risk. These factors allow us to hedge mortality risk up to first and second order effects. They are the analogous of the duration and convexity terms in classical financial hedging of zero-coupon-bonds, and they actually collapse into them when $\sigma(t,\lambda) = \sigma = 0$. In this case, in which mortality has no systematic risk component, we have:

$$Y(t,T) = 0$$

Hence, Delta and Gamma are functions of a' only, as in the deterministic case. We have

$$\Delta^{\sigma=0} = \frac{S(T)}{S(t)}X(t,T)$$

$$\Gamma^{\sigma=0} = \frac{S(T)}{S(t)}X^2(t,T)$$

It is straightforward to compute the sensitivity of any pure endowment policy portfolio with respect to mortality risk (evidently, this must be done for each generation separately). If the portfolio, valued Π, is made up of n_i policies with maturity T_i, $i = 1,..n$, each one with value $S(t,T_i)$, we have

$$d\Pi = \sum n_i dS(t,T_i) =$$

$$\sum_{i=1}^{n} n_i \frac{\partial S}{\partial t}dt + \sum_{i=1}^{n} n_i \frac{\partial S}{\partial I}dI + \frac{1}{2}\sum_{i=1}^{n} n_i^2 \frac{\partial^2 S}{\partial I^2}(dI)^2$$

4.7.2 Hedging

In order to hedge the reserve we have derived in the previous section we assume that the insurer can use either other pure endowments – with different maturities – or zero-coupon longevity bonds on the same generation [13]. Since we did not price idiosyncratic mortality risk, the price/value of a zero-coupon longevity bond is indeed equal to the pure endowment one. The difference, from the standpoint of an insurance company, is that it can sell endowments – or reduce its exposure through reinsurance – and buy longevity bonds, while, at least in principle, it cannot do the converse[14]. We could use a number of other instruments to cover the initial pure endowment, starting from life assurances or death bonds, which pay the benefit in case of death of the insured individual. We restrict the attention to pure endowments and longevity bonds for the sake of simplicity. Let us recall also that – together with the life assurance and death bonds – they represent the Arrow Debreu securities of the insurance market. Once hedging is provided for them, it can be extended to every more complicated instrument.
Suppose for instance that, in order to hedge n endowments with maturity T, it is possible to choose the number of endowments/longevity bonds with maturity T_1 and T_2: call them n_1 and n_2. The value of a portfolio made up of the three assets is

$$\Pi(t) = nS(t,T) + n_1 S(t,T_1) + n_2 S(t,T_2).$$

Its Delta and Gamma are respectively

$$\begin{aligned} \Delta_\Pi^M(t) = & \quad n\frac{\partial S}{\partial I}(t,T) + n_1\frac{\partial S}{\partial I}(t,T_1) + n_2\frac{\partial S}{\partial I}(t,T_2) \\ \Gamma_\Pi^M(t) = & \quad n\frac{\partial^2 S}{\partial I^2}(t,T) + n_1\frac{\partial^2 S}{\partial I^2}(t,T_1) + n_2\frac{\partial^2 S}{\partial I^2}(t,T_2) \end{aligned}$$

We can set these Delta and Gamma coefficients to zero (or some other precise value) by adjusting the quantities n_1 and n_2. One can easily solve the system of two equations in two unknowns and obtain hedged portfolios:

$$\begin{cases} \Delta_\Pi^M = 0 \\ \Gamma_\Pi^M = 0 \end{cases}$$

[13] If there is no longevity bond for a specific generation, basis risk arises: see for instance Cairns, Blake, Dowd, and MacMinn (2006).

[14] Reinsurance companies have less constraints in this respect. For instance, they can swap pure endowments or issue longevity bonds: see for instance Cowley and Cummins (2005).

Any negative solution for n_i has to be interpreted as an endowment sale, since this leaves the insurer exposed to a liability equal to n times the policy fair price. Any positive solution for n_i has to be interpreted as a longevity bond purchase. The cost of setting up the covered portfolio – which is represented by $\Pi(t)$ – can be paid using the pure endowment premium received by the policyholder. Alternatively, the problem can be extended so as to make the hedged portfolio self-financing. Self-financing can be guaranteed by endogenizing n and solving simultaneously the equations $\Pi = 0, \Delta_\Pi^M = 0$ and $\Gamma_\Pi^M = 0$ for n, n_1, n_2. As an alternative, if n is fixed, a third pure endowment/bond with maturity T_3 can be issued or purchased, so that the portfolio made up of $S(t,T), S(t,T_1), S(t,T_2)$ and $S(t,T_3)$ is self-financing and Delta and Gamma hedged. Our application in section 4.9 will cover both the non self-financing and self-financing possibilities.

4.8 Mortality and financial risk hedging

Let us consider now the case in which both mortality and financial risk exist. Again we develop the technique assuming a OU intensity. We also select a constant-parameter Hull and White model for the interest rate under the risk-neutral measure:

$$\Sigma(t,T) = \Sigma \exp(-g(T-t)) \tag{4.35}$$

with $\Sigma, g \in \mathbb{R}^+$. Substituting in (4.21) indeed we have

$$r(t) = F(0,t) + \frac{1}{2}\frac{\Sigma^2}{g^2}(1-e^{-gt})^2 + \Sigma \int_0^t e^{-g(t-s)} dW`_F(s).$$

which allows us to derive an expression for $B(t,T)$ analogous to (4.32):

$$B(t,T) = \frac{B(0,T)}{B(0,t)} \exp\left[-\bar{X}(t,T)K(t) - \bar{Y}(t,T)\right]$$

where

$$\bar{X}(t,T) := \frac{1-\exp(-g(T-t))}{g}$$

$$\bar{Y}(t,T) := \frac{\Sigma^2}{4g}\left[1-\exp(-2gt)\right]\bar{X}^2(t,T)$$

where K is the financial risk factor, measured by the difference between the short and forward rate:

$$K(t) := r(t) - F(0,t)$$

The pure endowment reserve at time t, according to (4.5) above, is

$$P(t,T) = \exp\left(-\int_t^T \left[f(t,u)+F(t,u)\right]du\right) = S(t,T)B(t,T)$$

Given the independence stated in Assumption 3, we can apply Ito's lemma and obtain the dynamics of the reserve $P(t,T)$ as

$$\begin{aligned} dP &= BdS + PdB \simeq B\left[\frac{\partial S}{\partial t}dt + \frac{\partial S}{\partial I}dI + \frac{1}{2}\frac{\partial^2 S}{\partial I^2}(dI)^2\right] + \\ &+ S\left[\frac{\partial B}{\partial t}dt + \frac{\partial B}{\partial K}dK + \frac{1}{2}\frac{\partial^2 B}{\partial K^2}(dK)^2\right] \end{aligned}$$

where

$$\frac{\partial B(t,T)}{\partial K} = -B(t,T)\bar{X}(t,T) \leq 0$$
$$\frac{\partial^2 B(t,T)}{\partial K^2} = B(t,T)\bar{X}^2(t,T) \geq 0$$

It follows that, for given t,

$$\frac{dP(t,T)}{P(t,T)} \simeq -X(t,T)dI + \frac{1}{2}X(t,T)^2(dI)^2 - \bar{X}^2(t,T)dK + \frac{1}{2}\bar{X}^2(t,T)(dK)^2$$

Hedging of the reserve is again possible by a proper selection of pure endowment/longevity bond contracts with different maturities and/or zero-coupon-bonds with different maturities. Here we consider the case in which the hedge against mortality and financial risk is obtained either issuing (purchasing) pure endowments (longevity bonds) or using also bonds.
Consider first using mortality linked contracts only. We can see that Delta and Gamma hedging of both the mortality and financial risk of n endowments with maturity T can be obtained via a mix of n_1, n_2, n_3, n_4 endowments/longevity bonds with maturities ranging from T_1 to T_4, by solving simultaneously the following hedging equations:

$$\begin{cases} \Delta_\Pi^M = 0 \\ \Gamma_\Pi^M = 0 \\ \Delta_\Pi^F = 0 \\ \Gamma_\Pi^F = 0 \end{cases} \tag{4.36}$$

This indeed means solving the system of equations

$$\begin{cases} \Delta_\Pi^M = nBSX + n_1B_1S_1X_1 + n_2B_2S_2X_2 + n_3B_3S_3X_3 + n_4B_4S_4X_4 = 0 \\ \Gamma_\Pi^M = nBSX^2 + n_1B_1S_1X_1^2 + n_2B_2S_2X_2^2 + n_3B_3S_3X_3^2 + n_4B_4S_4X_4^2 = 0 \\ \Delta_\Pi^F = nBS\bar{X} + n_1B_1S_1\bar{X}_1 + n_2B_2S_2\bar{X}_2 + n_3B_3S_3\bar{X}_3 + n_4B_4S_4\bar{X}_4 = 0 \\ \Gamma_\Pi^F = nBS\bar{X}^2 + n_1B_1S_1\bar{X}_1{}^2 + n_2B_2S_2\bar{X}_2{}^2 + n_3B_3S_3\bar{X}_3{}^2 + n_4B_4S_4\bar{X}_4{}^2 = 0 \end{cases} \tag{4.37}$$

where B denotes $B(t,T)$ and $B_i, X_i, \bar{X}_i$ denote $B(t,T_i), X(t,T_i), \bar{X}(t,T_i)$ for $i = 1,...,4$.

Consider now using both mortality-linked contracts and zero-coupon-bonds. In this case, the hedging equations (4.36) become:

$$\begin{cases} \Delta_{\Pi}^{M} = nBSX + n_1 B_1 S_1 X_1 + n_2 B_2 S_2 X_2 = 0 \\ \Gamma_{\Pi}^{M} = nBSX^2 + n_1 B_1 S_1 X_1^2 + n_2 B_2 S_2 X_2^2 = 0 \\ \Delta_{\Pi}^{F} = nBS\bar{X} + n_1 B_1 S_1 \bar{X}_1 + n_2 B_2 S_2 \bar{X}_2 + n_3 B_3 \bar{X}_3 + n_4 B_4 \bar{X}_4 = 0 \\ \Gamma_{\Pi}^{F} = nBS\bar{X}^2 + n_1 B_1 S_1 \bar{X}_1{}^2 + n_2 B_2 S_2 \bar{X}_2{}^2 + n_3 B_3 \bar{X}_3{}^2 + n_4 B_4 \bar{X}_4{}^2 = 0 \end{cases} \quad (4.38)$$

These equations can be solved either all together or sequentially (the first 2 with respect to n_1, n_2, the others with respect to n_3 and n_4), covering mortality risk at the first step and financial risk at the second step.
Both problems outlined in (4.37) and (4.38) can be extended to self-financing considerations. In both cases the value of the hedged portfolio is given by

$$\Pi(t) = nBS + n_1 B_1 S_1 + n_2 B_2 S_2 + n_3 B_3 S_3 + n_4 B_4 S_4$$

It is self-financing if $\Pi(0) = 0$ or if an additional contract is inserted, so that the enlarged portfolio value is null. In our applications we will explore both possibilities.

4.9 Application to a UK sample

In this Section, we present an application of our hedging model to a sample of UK insured people. We exploit our minimal change of measure, which preserves the biological and historically reasonable behaviour of the intensity. We also assume that $a' = a$, i.e. that the risk premium on mortality risk is null. This assumption could be easily removed by calibrating the model parameters to actual insurance products, most likely derivatives. We take the view that their market is not liquid enough to permit such calibration (see also Biffis (2005), Cairns, Blake, Dowd, and MacMinn (2006), to mention a few). We therefore calibrate the mortality parameters to historical data (the IML tables, that are projected tables for English annuitants). We assume also - at first - that the interest rate is constant and, without loss of generality, null - as in Section 4.7. We derive a "term structure of pure endowments" and the values of coefficients Delta and Gamma of the contracts. Afterwards, we introduce also a stochastic interest rate.

4.9.1 Mortality risk hedging

We keep the head fixed, considering contracts written on the lives of male individuals who were 65 years old on 31/12/2010. Hence, we set $t = 0$ and we calibrate our parameters $a_{65}, \sigma_{65}, \lambda_{65}(0) = -\ln p_{65}$ from our data set, considering the generation of individuals that were born in 1945. For this generation, a is calibrated to 10.94 %, while σ is 0.07 %. $\lambda_{65}(0)$ is instead 0.885%.[15]

First of all, we analyze the effect of a shock of one standard deviation on the Wiener driving the intensity process. Figure 4.1 shows graphically the impact of an upward and downward shock of one standard deviation on the forward intensity at $t = 1$ for different time horizons T. The forward mortality structure is derived from (4.3) using (4.15). The Figure clearly highlights that the effect becomes more and more evident – the trumpet opens up – as soon as the time horizon of the forward mortality becomes longer. Please notice that the behaviour is – as it should, from the economic point of view – opposite to the one of the corresponding Hull-White interest rates. In the rates case indeed the trumpet is reversed, since short-term forward rates are affected more than longer ones.

The following Table 4.1 reports the "term structure of pure endowment contracts" and compares the Delta and Gamma coefficients associated with contracts of different maturity in the stochastic case with the deterministic ones.

It appears clearly from the previous Table that the model gives hedging coefficients for mortality-linked contracts which are quite remarkably different from the deterministic ones for long maturities. For instance, the Δ^M and Γ^M hedging coefficients for a contract with maturity 30 years are respectively 6% smaller and larger than their deterministic counterparts. Contracts with long maturities are clearly very interesting from an insurer's point of view and hence their proper hedging is important.

As an example, imagine that an insurer has issued a pure endowment contract with maturity 15 years. Suppose that he wants to Delta-Gamma hedge this position using as cover instruments mortality-linked contracts with maturity 10 and 20 years. At a cost of 0.37, the insurer can instantaneously Delta-Gamma hedge its portfolio, by purchasing, respectively, 1.11 and 0.26 zero-coupon longevity bonds on these maturities. Having at disposal also the possibility of using contracts with a maturity of 30 years on the same population of individuals, a self-financing Delta-Gamma hedging strategy can

[15] We refer the reader to Luciano and Vigna (2008) for a full description of the data set and the calibration procedure.

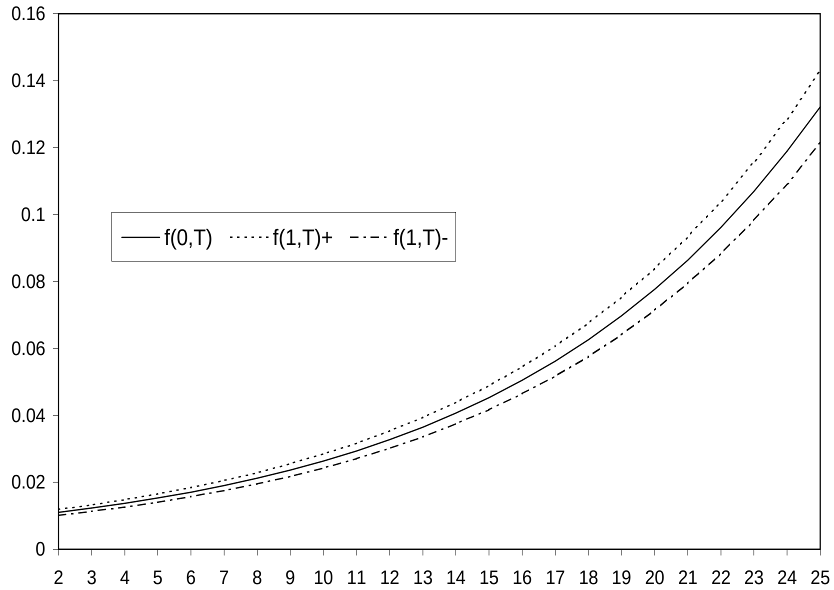

Figure 4.1: This figure shows the effect on the forward death intensity $f(1,T)$ of a shock equal to one standard deviation as a function of T. The central – solid – line represents the initial forward mortality intensity curve $f(0,T)$

Table 4.1: Stochastic vs. deterministic hedging coefficients

	Stochastic hedge			Deterministic hedge	
Maturity	$S(t,T)$	Δ^M	Γ^M	$\Delta^{\sigma=0}$	$\Gamma^{\sigma=0}$
1	0.99069	-1.04691	1.10633	-1.04691	1.10633
2	0.98041	-2.19187	4.90030	-2.19187	4.90030
5	0.94282	-6.27449	41.75698	-6.27439	41.75633
7	0.91116	-9.58396	100.80807	-9.58347	100.80284
10	0.85174	-15.46366	280.74803	-15.46053	280.69129
12	0.80306	-19.94108	495.16678	-19.93255	494.95501
15	0.71505	-27.19228	1034.08392	-27.16108	1032.89754
18	0.60899	-34.31821	1933.91002	-34.22325	1928.55907
20	0.52957	-38.32543	2773.64051	-38.14219	2760.37929
25	0.31713	-41.77104	5501.91988	-41.05700	5407.86868
27	0.23633	-39.27090	6525.53620	-38.18393	6344.91753
30	0.13319	-31.20142	7309.51024	-29.46466	6902.64225
35	0.03144	-12.93603	5322.98669	-10.78469	4437.74408

be implemented by purchasing 0.48 and 0.60 longevity bonds with maturity respectively 10 and 20 years, and issuing 0.10 pure endowments with maturity 30 years.

4.9.2 Mortality and financial risk hedging

The same procedure, as shown in Section 4.8, can be followed to hedge simultaneously the risks deriving from both stochastic mortality intensities and interest rates. Notice that, if we consider that the interest rate is stochastic (or at least different from zero), prices of pure endowment contracts no longer coincide with survival probabilities. Nonetheless, their Δ^M and Γ^M, the factors associated to mortality risk, remain unchanged when we introduce financial risk (see Section 4.8). Once one has estimated the coefficients underlying the interest rate process, we can easily derive the values of Δ^F and Γ^F, the factors associated to the financial risk, and the prices $P(t,T)$ of pure endowment/longevity bond contracts.
We calibrate our constant-parameter Hull and White model for forward interest rates to the observed zero-coupon UK government bonds at 31/12/2010.[16]
Table 4.2 shows prices and financial risk hedging factors of pure endowment contracts subject to both financial and mortality risks. Please notice that

[16]The parameter g is 2.72%, while the diffusion parameter Σ is calibrated to 0.65 %

Table 4.2: Hedging coefficients for stochastic financial risk

Maturity	$P(t,T)$	Δ^F	Γ^F
1	0.98395	-0.9798	0.9666
2	0.96214	-1.9103	3.7185
5	0.86696	-4.2988	20.0963
7	0.78430	-5.4865	34.9707
10	0.64372	-6.6170	57.9341
12	0.54597	-6.9606	71.2657
15	0.40404	-6.9596	85.7216
20	0.20649	-6.0149	92.7836
25	0.07972	-4.5599	82.7129
27	0.04902	-3.9667	75.8645
30	0.02037	-3.1366	64.3246
35	0.00278	-1.9995	45.1377

the absolute values of the factors related to the financial market are smaller than the ones related to the mortality risk.

These factors Δ^F and Γ^F, together with their mortality risk counterparts, Δ^M and Γ^M, allow us to hedge pure endowment contracts from both financial and mortality risk by setting up a portfolio - even self-financing - which instantaneously presents null values of all the Delta and Gamma factors.
As an example, consider again the hedging of a pure endowment with maturity 15 years. In order to Delta-Gamma hedge against both risks, we need to use four instruments (five if we want to self-finance the strategy). We can either use four pure endowments/longevity bonds written on the lives of the 65 year-old individuals or two mortality-linked contracts and two zero-coupon-bonds. In the first case, imagine to use contracts with maturity 10, 20, 25 and 30 years. The hedging strategy consists then in purchasing 0.35 longevity bonds with maturity 10 years, 1.27 with maturity 20 years and 0.30 with maturity 30 years, while issuing 0.87 pure endowment policies with maturity 25 years. In the second case, imagine the hedging instruments are mortality contracts with maturities 10 and 20 years and two zero-coupon-bonds with maturities 5 and 20 years. The strategy consists in purchasing 1.11 longevity bonds with maturity 10 years and 0.26 with maturity 20 years and in taking a short position on 0.60 zero-coupon-bonds with maturity 5 years and a long one on 0.10 zero-coupon-bonds with maturity 20 years. A self-financing hedge can be easily obtained by adding an instrument to the portfolio. For example, such a self-financing hedge can be obtained by purchasing 0.41 longevity bonds with maturity 10 years, 0.98 with maturity 20

years and 0.22 with maturity 35 years and issuing 0.38 pure endowments with maturity 25 years and 0.13 with maturity 30 years.

4.10 Summary and conclusions

This paper develops a Delta and Gamma hedging framework for mortality and interest rate risk.
We have shown that, consistently with the interest rate market, when the spot intensity of stochastic mortality follows an OU or FEL process, an HJM condition on its drift holds for every constant risk premium, without assuming no arbitrage. Hence, we have shown that it is possible to hedge systematic mortality risk in a way which is identical to the Delta and Gamma hedging approach in the HJM framework for interest rates. Delta and Gamma are very easy to compute, at least in the OU case. Similarly, the hedging quantities are easily obtained as solutions to linear systems of equations. Hence, this hedging model can be very attractive for practical applications.
Adding financial risk is a straightforward extension in terms of insurance pricing, if the bond market is assumed to be without arbitrages (and complete, so that the financial change of measure is unique). Delta and Gamma hedging is straightforward too if - as in the examples - the risk-neutral dynamics of the forward interest rate is constant-parameter Hull and White. Our application shows that the unhedged effect of a sudden change on the mortality rate is remarkable and the stochastic and deterministic Deltas and Gammas are quite different, especially for long time horizons. Last but not least, calibrated Deltas and Gammas are bigger for mortality than for financial risk. The Delta and Gamma computation can be performed in the presence of FEL stochastic mortality. The whole hedging technique can also be extended using the same change of measure to the case of a CIR mortality intensity, reverting to a function of time.

4.11 Appendix: Cox Processes

Throughout the paper, we explored models in which the intensity of a Poisson process is stochastic. Hence, stochasticity of the process is two-fold. On one side, the jump component η is random, on the other its probability is also stochastic. For this reason, these processes we studied are called *doubly stochastic* or *Cox processes*. In this Appendix we briefly review their definition and properties.
We assume that the process underlying the intensity $\lambda(t)$ is adapted to the

filtration $\mathcal{F}_t$, right continuous and independent of the jump process η. Obviously, conditional on the filtration $\mathcal{F}^\lambda$ - and hence on $\lambda(t)$ - we still have a Poisson process. We define the cumulated intensity or hazard process $\Lambda(t) = \int_0^T \lambda_t dt$.
The first jump time can be represented as $\tau = \Lambda^{-1}(\eta)$. Hence, we can state the following property:

$$\mathbb{P}\{\tau \in [t, t+dt] | \tau \geqslant t, \mathcal{F}_t\} = \lambda_t dt$$

Hence, the probability of a first jump in the interval dt, given the information set at t, is equal to $\lambda_t dt$, i.e. the product of the intensity at time t and the length of the interval. Moreover, it is possible to show the following, which in the paper allowed us to compute the survival probabilities:

$$\begin{aligned}
\mathbb{P}\{\tau \geqslant s\} &= \mathbb{P}\{\Lambda(\tau) \geqslant \Lambda(s)\} = \mathbb{P}\{\eta \geqslant \int_0^s \lambda(u)du\} = \\
&= \mathbb{E}\left[\mathbb{P}\{\eta \geqslant \int_0^s \lambda(u)du | \mathcal{F}^\lambda\}\right] = \mathbb{E}\left[e^{-\int_0^s \lambda(u)du}\right]
\end{aligned}$$

Bibliography

Barbarin, J. (2008). Heath-jarrow-morton modelling of longevity bonds and the risk minimization of life insurance portfolios. *Insurance: Mathematics and Economics 43*(1), 41–55.

Biffis, E. (2005). Affine processes for dynamic mortality and actuarial valuations. *Insurance: Mathematics and Economics 37*(3), 443–468.

Blake, D., A. De Waegenaere, R. MacMinn, and T. Nijman (2010). Longevity risk and capital markets: The 2008-2009 update. *Insurance: Mathematics and Economics 46*, 135–138.

Boot, A., S. Greenbaum, and A. Thakor (1993). Reputation and discretion in financial contracting. *The American Economic Review 83*(5), 1165–1183.

Bornhuetter, R.L., F. R. (1972). The actuary and ibnr. *Proc. Cas. Act. Soc.*, 181–195.

Brémaud, P. (1981). *Point processes and queues. Martingale dynamics.* Springer Series in Statistics, New York (NY).

Cairns, A., D. Blake, and K. Dowd (2006). Pricing death: Frameworks for the valuation and securitization of mortality risk. *Astin Bulletin 36*(1), 79–120.

Cairns, A., D. Blake, K. Dowd, and R. MacMinn (2006). Longevity bonds: financial engineering, valuation, and hedging. *The Journal of Risk and Insurance 73*(4), 647–672.

Chen, Z., D. Li, L. Liao, F. Moshirian, and C. Szablocs (2009). Expansion and consolidation of bancassurance in the 21st century. *Journal of International Financial Markets, Institutions and Money 19*(4), 633–644.

Cherubini, U., E. Luciano, and W. Vecchiato (2004). *Copula methods in finance.* John Wiley & Sons, New York (NY).

Cowles, M. and B. Carlin (1996). Markov Chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association 91*(434).

Cowley, A. and J. Cummins (2005). Securitization of life insurance assets and liabilities. *The Journal of Risk and Insurance 72*(2), 193–226.

Cox, J., J. Ingersoll Jr, and S. Ross (1985). A theory of the term structure of interest rates. *Econometrica 53*(2), 385–407.

Dahl, M. (2004). Stochastic mortality in life insurance: market reserves and mortality-linked insurance contracts. *Insurance: Mathematics and Economics 35*(1), 113–136.

Dahl, M. and T. Møller (2006). Valuation and hedging of life insurance liabilities with systematic mortality risk. *Insurance: Mathematics and Economics 39*(2), 193–217.

Dalla Valle, L. (2009). Bayesian copulae distributions, with application to operational risk management. *Methodology and Computing in Applied Probability 11*(1), 95–115.

De Alba, E. (2002). Bayesian estimation of outstanding claim reserves. *North American Actuarial Journal 6*(4), 1–20.

De Alba, E. and L. Nieto-Barajas (2008). Claims reserving: A correlated bayesian model. *Insurance: Mathematics and Economics 43*(3), 368–376.

De Jong, P. (2009). Modeling dependence between loss triangles using copulas. *Working Paper, Sidney NSWA*.

DeGroot, M. (2004). *Optimal statistical decisions*. John Wiley / Sons, Hoboken (NJ).

Dell'Ariccia, G. and R. Marquez (2010). Risk and the corporate structure of banks. *The Journal of Finance 65*(3), 1075–1096.

Dierick, F. (2004). The supervision of mixed financial services groups in Europe.

Duffie, D. (1998). Defaultable term structure models with fractional recovery of par. *Working paper, Stanford University*.

Duffie, D. and K. Singleton (1999). Modeling term structures of defaultable bonds. *Review of Financial Studies 12*(4), 687.

Duffie, D., J. P. and K.Singleton (2000). Trasnform Analysis and Asset Pricing for Affine Jump-Diffusions. *Econometrica 68*(6), 1343–1376.

Edwards, P. (1998). Managing Risk and Capital in Financial Conglomerates". In *Risk and Capital Management Conference of Australian Prudential Regulation Authority, conference paper.*

England, P. and R. Verrall (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal 8*(3), 443–518.

Flannery, M., J. Houston, and S. Venkataraman (1993). Financing multiple investment projects. *Financial Management 22*(2), 161–172.

Freixas, X., G. Lóránth, and A. Morrison (2007). Regulating financial conglomerates. *Journal of Financial Intermediation 16*(4), 479–514.

Gelman, A. and D. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical science 7*(4), 457–472.

Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika 74*(3), 549.

Genest, C. and L. Rivest (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association 88*(423), 1034–1043.

Gilks, W., S. Richardson, and D. Spiegelhalter (1996). Introducing markov chain monte carlo. In S. D. Gilks WR, Richardson S (Ed.), *Markov Chain Monte Carlo in Practice*, pp. 1–19. Chapman and Hall, London.

Gropp, R. and F. Heider (2010). The determinants of bank capital structure. *Review of Finance, forthcoming.*

Harding, J., X. Liang, and S. Ross. The Optimal Capital Structure of Banks: Balancing Deposit In-surance, Capital Requirements and Tax-Advantaged Debt. *Working paper, University of Connecticut.*

Harr, T. and T. Rønde (2003). Branch or subsidiary? Capital regulation of multinational banks. *Mimeo, University of Copenaghen.*

Hastings, W. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika 57*(1), 97.

Heath, D., R. Jarrow, and A. Morton (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica 60*(1), 77–105.

Hellwig, M. (2009). Systemic risk in the financial sector: an analysis of the subprime-mortgage financial crisis. *De Economist 157*(2), 129–207.

Herring, R. and J. Carmassi (2008). The structure of cross-sector financial supervision. *Financial Markets, Institutions & Instruments 17*(1), 51–76.

Huang, J. and M. Huang (2002). How Much of Corporate-Treasury Yield Spread Is Due to Credit Risk? *Working Paper, Penn State University*.

Hull, J. and A. White (1990). Pricing interest-rate-derivative securities. *Review of financial studies 3*(4), 573.

Jacod, J. and A. Shiryaev (1987). *Limit theorems for stochastic processes.* Springer, Berlin.

Jouini, M. and R. Clemen (1996). Copula models for aggregating expert opinions. *Operations Research 44*(3), 444–457.

Kahn, C. and A. Winton (2004). Moral hazard and optimal subsidiary structure for financial institutions. *The Journal of Finance 59*(6), 2531–2575.

Kretzschmar, G., A. McNeil, and A. Kirchner (2010). Integrated models of capital adequacy-Why banks are undercapitalised. *Journal of Banking & Finance 34*(12), 2838–2850.

Kuritzkes, A., T. Schuermann, and S. Weiner (2003). Risk measurement, risk management, and capital adequacy in financial conglomerates. *Brookings-Wharton Papers on Financial Services 2003*(1), 141–193.

Laeven, L. and R. Levine (2007). Is there a diversification discount in financial conglomerates? *Journal of Financial Economics 85*(2), 331–367.

Lando, D. (1998). On cox processes and credit risky securities. *Review of Derivatives research 2*(2), 99–120.

Lee, A. (1993). Generating random binary deviates having fixed marginal distributions and specified degrees of association. *The American Statistician 47*(3).

Leland, H. (2007). Financial synergies and the optimal scope of the firm: Implications for mergers, spinoffs, and structured finance. *The Journal of Finance 62*(2), 765–807.

Lewellen, W. (1971). A pure financial rationale for the conglomerate merger. *Journal of Finance 26*(2), 521–537.

Li, J. (2006). Modelling dependency between different lines of business with copulas. *Working Paper, University of Melbourne*.

Litterman, R. and T. Iben (1991). Corporate bond valuation and the term structure of credit spreads. *The journal of portfolio management 17*(3), 52–64.

Luciano, E. and G. Nicodano (2010). Intercorporate guarantees, leverage and taxes. *Unpublished, University of Torino*.

Luciano, E., J. Spreeuw, and E. Vigna (2008). Modelling stochastic mortality for dependent lives. *Insurance: Mathematics and Economics 43*(2), 234–244.

Luciano, E. and E. Vigna (2008). Mortality risk via affine stochastic intensities: calibration and empirical relevance. *Belgian Actuarial Bulletin 8*(1), 5–16.

Macey, J. (1999). Business of Banking: Before and after Gramm-Leach-Bliley, The. *Journal of Corporation Law 25*(4), 691–719.

Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *Astin Bulletin 23*(2), 213–225.

Merz, M. and M. Wuthrich (2008). *Stochastic claims reserving methods in insurance*. Wiley.

Milevsky, M. and D. Promislow (2001). Mortality derivatives and the option to annuitise. *Insurance: Mathematics and Economics 29*(3), 299–318.

Miltersen, K. and S. Persson (2006). Is mortality dead? Stochastic forward force of mortality rate determined by no arbitrage. *Working paper, Norwegian School of Economics and Business Administration*.

Nelsen, R. (2006). *An introduction to copulas*. Springer Verlag.

Ntzoufras, I. and P. Dellaportas (2002). Bayesian modeling of outstanding liabilities incorporating claim count uncertainty. *North American Actuarial Journal 6*(1), 113–136.

Robert, C. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Verlag, New York (NY).

Roberts, G. (1996). Markov chain concepts related to sampling algorithms. In S. D. Gilks WR, Richardson S (Ed.), *Markov Chain Monte Carlo in Practice*. Chapman and Hall.

Roberts, G. and J. Rosenthal (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science 16*(4), 351–367.

Schaefer, S. and I. Strebulaev (2008). Structural models of credit risk are useful: Evidence from hedge ratios on corporate bonds. *Journal of Financial Economics 90*(1), 1–19.

Schmid, M. and I. Walter (2009). Do financial conglomerates create or destroy economic value? *Journal of Financial Intermediation 18*(2), 193–216.

Schweizer, B. and A. Sklar (1983). *Probabilistic metric spaces.* North-Holland, Amsterdam.

Shreve, S. (2004). *Stochastic calculus for finance: Continuous-time models.* Springer Verlag, New York, NY.

Strebulaev, I. and B. Yang (2006). The mystery of zero-leverage firms. *Working Paper, Stanford University*.

Tang, A. and E. Valdez (2005). Economic capital and the aggregation of risks using copulas. *School of Actuarial Studies and University of New South Wales*.

Van Lelyveld, I. and K. Knot (2009). Do financial conglomerates create or destroy value? evidence for the eu. *Journal of Banking & Finance 33*(12), 2312–2321.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of financial economics 5*(2), 177–188.

Vennet, R. (2002). Cost and profit efficiency of financial conglomerates and universal banks in europe. *Journal of Money, Credit, and Banking 34*(1), 254–282.

Zhao, X. and X. Zhou (2010). Applying copula models to individual claim loss reserving methods. *Insurance: Mathematics and Economics 46*(2), 290–299.