Forward ordinal probability models for point-in-time probability of default term structure: methodologies and implementations for IFRS 9 expected credit loss estimation and CCAR stress testing
Bill Huajian Yang

Bayesian analysis in an aggregate loss model: validation of the structure functions
A. Hernández-Bastida, J. M. Pérez-Sánchez and M. P. Fernández-Sánchez

On the correlation and parametric approaches to calculation of credit value adjustment
Tao Pang, Wei Chen and Le Li

The use of the triangular approximation for some complicated risk measurement calculations
Nick Georgiopoulos
Risk.net in numbers

200+
New articles & technical papers (each month)

140,000
Users (each month)

370,000
Page views (each month)

58,000+
articles stretching back 20 years

6,500+
on Asset Management

6,900+
on Commodities

£21,000+
on Derivatives

19,400+
on Regulation

19,600+
on Risk Management

See what you’re missing

Visit the world’s leading source of exclusive in-depth news & analysis on risk management, derivatives and complex finance now.

To subscribe to a Risk Journal visit subscriptions.risk.net/journals or email info@risk.net
The Journal of Risk Model Validation
EDITORIAL BOARD

Editor-in-Chief
STEVE SATCHELL  University of Cambridge

Associate Editors

MOHAN BHATIA  Wipro Ltd
STEFAN BLOCHWITZ  Deutsche Bundesbank
J. L. BREEDEN  Prescient Models LLC
WEI CHEN  SAS Institute Inc.
GEORGE CHRISTODOULAKIS  Manchester Business School
MARCELO CRUZ
KLAUS DUellmann  European Central Bank
DOUGLAS Dwyer  Moody’s Analytics
CHRISTOPHER C. FINGER  Federal Reserve Board
DAVID LI  Mitsubishi Bank

CHRISTIAN MEYER  DZ BANK AG
PETER MIU  McMaster University
BOGIE OZDEMIR  Sun Life Financial Group
PETER QUELL  DZ BANK AG
DANIEL RÖSCH  University of Hannover
HARALD SCHEULE  University of Technology, Sydney
ROGER STEIN  Moody’s Investors Service
ALLAN TIMMERMANN  University of California, San Diego
BILL HUAJIAN YANG  Royal Bank of Canada

SUBSCRIPTIONS

The Journal of Risk Model Validation (Print ISSN 1753-9579 | Online ISSN 1753-9587) is published quarterly by Infopro Digital Risk Limited, Haymarket House, 28–29 Haymarket, London SW1Y 4RX, UK. Subscriptions are available on an annual basis, and the rates are set out in the table below.

Subscription packages are available on an annual basis. To find out about the different options, including our exclusive academic rates starting from £100, visit subscriptions.risk.net/journals-print or contact ben.anderton@infopro-digital.com.

All prices include postage. All subscription orders, single/back issues orders, and changes of address should be sent to:

UK & Europe Office: Infopro Digital, Haymarket House, 28-29 Haymarket, London SW1Y 4RX, UK. Tel: +44 1858 438 800

US & Canada Office: Infopro Digital, 55 Broad Street, Floor 22, New York, NY, 10005, USA. Tel: +1 212 776 8075

Asia & Pacific Office: Infopro Digital, Unit 1704–05, Berkshire House, Taikoo Place, 25, Westlands Road, Quarry Bay, Hong Kong. Tel +852 3411 4828

Website: risk.net/journals
E-mail: info@risk.net

To subscribe to a Risk Journal visit subscriptions.risk.net/journals or email info@risk.net
The Journal of Risk Model Validation

GENERAL SUBMISSION GUIDELINES

Manuscripts and research papers submitted for consideration must be original work that is not simultaneously under review for publication in another journal or other publication outlets. All articles submitted for consideration should follow strict academic standards in both theoretical content and empirical results. Articles should be of interest to a broad audience of sophisticated practitioners and academics.

Submitted papers should follow *Webster’s New Collegiate Dictionary* for spelling, and *The Chicago Manual of Style* for punctuation and other points of style, apart from a few minor exceptions that can be found at www.risk.net/journals. Papers should be submitted electronically via email to: journals@infopro-digital.com. Please clearly indicate which journal you are submitting to.

You must submit two versions of your paper; a single \TeX version and a PDF file. \TeX files need to have an explicitly coded bibliography included. All files must be clearly named and saved by author name and date of submission. All figures and tables must be included in the main PDF document and also submitted as separate editable files and be clearly numbered.

All papers should include a title page as a separate document, and the full names, affiliations and email addresses of all authors should be included. A concise and factual abstract of between 150 and 200 words is required and it should be included in the main document. Five or six keywords should be included after the abstract. Submitted papers must also include an Acknowledgements section and a Declaration of Interest section. Authors should declare any funding for the article or conflicts of interest. Citations in the text must be written as (John 1999; Paul 2003; Peter and Paul 2000) or (John et al 1993; Peter 2000).

The number of figures and tables included in a paper should be kept to a minimum. Figures and tables must be included in the main PDF document and also submitted as separate individual editable files. Figures will appear in color online, but will be printed in black and white. Footnotes should be used sparingly. If footnotes are required then these should be included at the end of the page and should be no more than two sentences. Appendixes will be published online as supplementary material.

Before submitting a paper, authors should consult the full author guidelines at: http://www.risk.net/static/risk-journals-submission-guidelines

Queries may also be sent to:

The Journal of Risk Model Validation, Infopro Digital, Haymarket House, 28–29 Haymarket, London SW1Y 4RX, UK
Tel: +44 1858 438 800 (UK/ EU), +1 212 776 8075 (USA), +852 3411 4828 (Asia)
E-mail: Journals@infopro-digital.com

To subscribe to a Risk Journal visit subscriptions.risk.net/journals or email info@risk.net
As monetary institutions rely heavily on economic and financial models for a wide array of applications, model validation has become progressively inventive within the field of risk. *The Journal of Risk Model Validation* focuses on the implementation and validation of risk models, and it aims to provide a greater understanding of key issues including the empirical evaluation of existing models, pitfalls in model validation and the development of new methods. We also publish papers on backtesting. Our main field of application is in credit risk modeling but we are happy to consider any issues of risk model validation for any financial asset class.

*The Journal of Risk Model Validation* considers submissions in the form of research papers on, but not limited to, the following topics.

- Empirical model evaluation studies.
- Backtesting studies.
- Stress-testing studies.
- New methods of model validation/backtesting/stress testing.
- Best practices in model development, deployment, production and maintenance.
- Pitfalls in model validation techniques (all types of risk, forecasting, pricing and rating).

To subscribe to a Risk Journal visit subscriptions.risk.net/journals or email info@risk.net
CONTENTS

Letter from the Editor-in-Chief vii

RESEARCH PAPERS

Forward ordinal probability models for point-in-time probability of default term structure: methodologies and implementations for IFRS 9 expected credit loss estimation and CCAR stress testing 1
Bill Huajian Yang

Bayesian analysis in an aggregate loss model: validation of the structure functions 19
A. Hernández-Bastida, J. M. Pérez-Sánchez and M. P. Fernández-Sánchez

On the correlation and parametric approaches to calculation of credit value adjustment 49
Tao Pang, Wei Chen and Le Li

The use of the triangular approximation for some complicated risk measurement calculations 69
Nick Georgiopoulos
LETTER FROM THE EDITOR-IN-CHIEF

Steve Satchell
Trinity College, University of Cambridge

This issue of The Journal of Risk Model Validation has a mildly actuarial flavor, as two of our four papers deal with collective risk models. This is a term used in the insurance literature when the variable of interest is total claims, which is a random sum of individual claims. Clearly, such a structure fits neatly into a framework for aggregate counterparty default loss, for example, as well as other ideas familiar to our readers.

The issue’s first paper, “Forward ordinal probability models for point-in-time probability of default term structure: methodologies and implementations for IFRS 9 expected credit loss estimation and CCAR stress testing” by Bill Huajian Yang, addresses some of the issues that arise in models that involve ranked data and what the author refers to as sensitivities. Using rank-specific sensitivities as basic building blocks in modeling allows one to directly implement matters such as rating changes/upgrades, among others. An estimation approach is outlined, certain calculations that arise from International Financial Reporting Standard 9 (IFRS 9) are discussed, and an example illustrating the superiority of this approach to the Merton model is included.

“Bayesian analysis in an aggregate loss model: validation of the structure functions” by A. Hernández-Bastida, J. M. Pérez-Sánchez and M. P. Fernández-Sánchez is the second paper in this issue. It deals with the empirical evaluation of a collective risk model using a Bayesian structure. Much Bayesian analysis is driven by the search for mathematical beauty, so the prior distribution is often chosen so that the posterior distribution will be elegant, rather than because it is a suitable vehicle to express investor/agent beliefs. Further, elicitation (the act of gleaning information on the prior structure from the relevant agents) is often completely ignored. If Bayesian analysis is to have any added value, elicitation needs to be carried out carefully. I am pleased to say that this paper focuses on both of these aspects of Bayesian analysis.

Our third paper, by Tao Pang, Wei Chen and Le Li, is “On the correlation and parametric approaches to calculation of credit value adjustment”. Credit value adjustment is defined by the authors as “an adjustment added to the fair value of an over-the-counter trade due to the risk of counterparty defaults”. We might think of it as a risk premium. One of the key ideas in this area of research is directional-way risk (DWR). The authors investigate DWR using a parametric approach, where the critical parameter is, in some sense, the correlation between the size of exposure to the counterparty default and the probability of counterparty default.

Nick Georgiopoulos’s “The use of the triangular approximation for some complicated risk measurement calculations” is the fourth and final paper in this issue of The Journal of Risk Model Validation. In it, Georgiopoulos introduces the triangular
approximation to the normal distribution in order to extract closed-form and semi-closed-form solutions that are useful in risk measurement calculations. In particular, he applies the triangular approximation to the normal density to derive closed-form solutions for risk measurement using actuarial models. These include not only insurance risk, such as gamma and Pareto-distributed losses, but also financial risk. In addition, Georgiopoulou approximates the collective risk model under lognormally distributed severities and estimates its value-at-risk. The accuracy of the approximations is checked via Monte Carlo.
Research Paper

Forward ordinal probability models for point-in-time probability of default term structure: methodologies and implementations for IFRS 9 expected credit loss estimation and CCAR stress testing

Bill Huajian Yang

Enterprise Stress Testing, Royal Bank of Canada, 155 Wellington Street West, Toronto M5V 3H6, Canada; email: bill.yang@rbc.com

(Received January 31, 2017; revised June 23, 2017; accepted June 26, 2017)

ABSTRACT

Common ordinal models, including the ordered logit model and the continuation ratio model, are formulated by a common score (ie, a linear combination of given explanatory variables) plus rank-specific intercepts. Sensitivity to the common score is generally not differentiated between rank outcomes. We propose an ordinal model based on forward ordinal probabilities for rank outcomes. In addition to the common score and intercepts, the forward ordinal probabilities are formulated by the rank- and rating-specific sensitivity (for a risk-rated portfolio). This rank-specific sensitivity allows a risk rating to respond to its migrations to default, downgrade, stay and upgrade accordingly. A parameter estimation approach based on maximum likelihood for observing rank-outcome frequencies is proposed. Applications of the proposed model include modeling rating migration probability for point-in-time probability of default term structure for International Financial Reporting Standard 9 expected credit loss estimation and Comprehensive Capital Analysis and Review stress testing. Unlike the rating transition model based on the Merton model, which allows only one sensitivity
parameter for all rank outcomes for a rating and uses only systematic risk drivers, the proposed forward ordinal model allows sensitivity to be differentiated between outcomes, and to include entity-specific risk drivers (e.g., the downgrade history or credit quality changes for an entity in the previous two quarters can be included). No additional estimation of the asset correlation is required. As an example, the proposed model, benchmarked with the rating transition model based on the Merton model, is used to estimate the probability of default term structure for a commercial portfolio, where for each rating the sensitivities are differentiated between migrations to default, downgrade, stay and upgrade. Our results show that the proposed model is more robust.

**Keywords:** ordinal model; forward ordinal probability; common score; rank-specific sensitivity; rating migration probability.

## 1 INTRODUCTION

Let $R$ denote the outcome for a trial with exactly one of the ordinal outcome values $\{1, 2, \ldots, k\}$. The forward (respectively, backward) ordinal probability, for a rank value $i$, is the conditional probability that the outcome value is $i$, given that all outcome ranks are no less (respectively, not larger) than $i$.

Common ordinal models, as reviewed in Section 2, include the ordered logit model (i.e., the proportion-odd model) and the continuation ratio model. For an ordered logit model, the cumulative probabilities for rank outcomes are modeled by a common score (i.e., a linear combination of explanatory variables) together with a rank-specific intercept. For a continuation ratio model, the forward or backward ordinal probabilities for rank outcomes are also modeled by a common score with rank-specific intercept. Sensitivities to the common score are generally not differentiated between rank outcomes.

It is commonly observed that entities with high risk ratings are more sensitive, and vulnerable to adverse shocks, and that entities are more likely to migrate to higher risk grades than to lower risk ratings during downturns. Risk sensitivity is not generally uniform between risk ratings or between outcome ranks.

We propose an ordinal model based on forward ordinal probability (model (3.2) or (3.4); see Section 3). The forward ordinal probabilities are formulated by a common score plus rank-specific sensitivity and intercept. We propose an algorithm for parameter estimation based on maximum likelihood for observing rank-outcome frequencies. The model can be implemented easily by a modeler using, for example, the SAS software procedure PROC NLMIXED (Wolfinger 2008).
Applications of the proposed model include

(a) modeling the rating migration probability for Comprehensive Capital Analysis and Review (CCAR) stress testing (Board of Governors of the Federal Reserve System 2016) and the point-in-time probability of default (PD) term structure for International Financial Reporting Standard 9 (IRFS 9) expected credit loss estimation (Ankarath et al 2010);

(b) estimation of the PD for a low default portfolio, and shadow rating modeling.

The modeling of state transition probabilities dates back to the original CreditMetrics, CreditPortfolioView and CreditRisk+ credit portfolio approaches (Derbali and Hallara 2013; Diaz and Gemmill 2002), and contributions by researchers including Nyström and Skoglund (2006) and Wei (2003). The point-in-time rating transition probability model based on the Merton model (Gordy 2003; Merton 1974; Miu and Ozdemir 2009; Vasicek 2002), which is formulated by a common credit index, was proposed by Miu and Ozdemir (2009), and extended by Yang (2016) to facilitate rating-level sensitivity for CCAR stress testing and IFRS 9 expected credit loss estimation.

Our proposed ordinal model, formulated by a common score plus an outcome rank-specific sensitivity, has several advantages. The outcome rank-specific sensitivity allows a risk rating to respond to its migrations to “default”, “downgrade”, “stay” and “upgrade” accordingly. Monotonicity can be imposed for the sensitivity parameters between initial ratings for each type of migration. Under this model structure, risk for an entity is driven by the common score (as a dynamic) plus the sensitivity in responding to a scenario. Unlike the rating transition model (Yang 2016) based on the Merton model framework, which allows only one sensitivity parameter for all outcomes for a rating and uses only systematic risk drivers, our model can include entity-specific risk drivers and allows for rank-specific sensitivity. No additional estimation for asset correlation is required. Further, the loglikelihood based on the forward PD given by a cumulative distribution function is generally concave, greatly increasing optimization efficiency.

Entity-specific drivers, such as downgrade history or credit quality changes in the last two quarters, can help improve prediction and address the issue of the Markov assumption for most migration models, particularly when the portfolio is small and idiosyncratic risk cannot be diversified away.

The paper is organized as follows. In Section 2, we review two commonly used ordinal regression models: the ordered logit model and the continuation ratio model. In Section 3, we propose the forward ordinal model and show the loglikelihood function and its concavity. A heuristic hard expectation maximization algorithm for parameter estimation is proposed in Section 4. The model is validated and used in Section 5 to
estimate the point-in-time PD structure for a commercial portfolio, where for each rating the sensitivities are differentiated between migrations to default, downgrade, stay and upgrade. The model is benchmarked using a rating transition model based on the Merton model framework. Concluding remarks are given in Section 6.

2 A REVIEW OF ORDINAL REGRESSION MODELS

In this section we review two common ordinal models: ordinal regression and the continuation ratio model.

Let $R$ denote the outcome for a trial with exactly one of the ordinal outcome values $\{1, 2, \ldots, k\}$. Given a scenario consisting of a list of explanatory variables $x_1, x_2, \ldots, x_m$, let $x = (x_1, x_2, \ldots, x_m)$ denote the corresponding vector. Let $F_i(x)$ and $p_i(x)$ denote, respectively, the cumulative and marginal probabilities defined by

$$F_i(x) = P(R \leq i \mid x),$$

$$p_i(x) = P(R = i \mid x).$$

Given $x$ and rank value $i$, the forward ordinal probability $\tilde{p}_i(x)$ and the backward ordinal probability $\tilde{p}_{bi}(x)$ are defined by the following conditional probabilities, respectively:

$$\tilde{p}_i(x) = P(R = i \mid x, R \geq i),$$

$$\tilde{p}_{bi}(x) = P(R = i \mid x, R \leq i).$$

Remark 2.1 We can always model the backward ordinal probability via the forward ordinal probability model: we simply reverse the order of the ordinal outcomes and reindex the resulting forward ordinal probability $\tilde{p}_i(x)$ by replacing $i$ with $(k + 1 - i)$. Therefore, we focus only on the forward ordinal probability model; all discussions for this model apply naturally to the backward ordinal model by an appropriate reversion for the outcome order and the index of the forward probability.

Proposition 2.2 The following equations hold:

$$F_i(x) = p_1(x) + p_2(x) + \cdots + p_i(x), \quad (2.1a)$$

$$\tilde{p}_i(x) = \frac{p_i(x)}{1 - F_{i-1}(x)}, \quad (2.1b)$$

$$p_i(x) = F_i(x) - F_{i-1}(x) = [1 - F_{i-1}(x)]\tilde{p}_i(x), \quad (2.1c)$$

$$[1 - F_i(x)] = [1 - \tilde{p}_1(x)][1 - \tilde{p}_2(x)]\cdots[1 - \tilde{p}_i(x)]. \quad (2.1d)$$

Proof Equation (2.1a) is immediate. Equation (2.1b) follows from the Bayesian theorem, while (2.1c) follows from (2.1a) and (2.1b). By (2.1c), we have

$$1 - F_i(x) = [1 - F_{i-1}(x)] - p_i(x) = [1 - F_{i-1}(x)][1 - \tilde{p}_i(x)].$$

Thus, (2.1d) follows by induction.
For the largest rank outcome \( k \), we have

\[
F_k(x) = 1, \quad \bar{p}_k(x) = 1, \\
p_k(x) = 1 - [p_1(x) + p_2(x) + \cdots + p_{k-1}(x)].
\]

Therefore, by Proposition 2.2, an ordinal model can be chosen to model one of the following components:

(i) the cumulative probabilities \( \{F_i(x) \mid i = 1, 2, \ldots, k-1\} \);
(ii) the marginal probabilities \( \{p_i(x) \mid i = 1, 2, \ldots, k-1\} \);
(iii) the forward ordinal probabilities \( \{\bar{p}_i(x) \mid i = 1, 2, \ldots, k-1\} \).

Marginal probabilities are subject to the following constraints:

\[
p_1(x) + p_2(x) + \cdots + p_i(x) \leq 1, \quad p_1(x) + p_2(x) + \cdots + p_k(x) = 1.
\]

Therefore, modeling marginal probabilities individually introduces additional complexity. In general, we can choose to model either the cumulative probabilities or the forward ordinal probabilities, as reviewed and discussed below.

### 2.1 Ordinal regression models

An ordinal regression model is generally formulated by cumulative probabilities \( \{F_i(x) \mid i = 1, 2, \ldots, k-1\} \) as

\[
F_i(x) = F(b_i + a_1x_1 + a_2x_2 + \cdots + a_mx_m), \quad b_1 \leq b_2 \leq \cdots \leq b_{k-1}, \tag{2.2}
\]

where \( F \) denotes the cumulative distribution for a probability distribution. The coefficients \( a_1, a_2, \ldots, a_m \) in model (2.2) do not depend on index \( i \leq k-1 \).

As they are cumulative probabilities, \( \{F_i(x) \mid i = 1, 2, \ldots, k-1\} \) are required to satisfy the following condition:

\[
F_1(x) \leq F_2(x) \leq \cdots \leq F_{k-1}(x). \tag{2.3}
\]

This is guaranteed for the ordinal regression model by the constraint \( b_1 \leq b_2 \leq \cdots \leq b_{k-1} \) in (2.2). When modeling the cumulative probabilities, condition (2.3) implies the coefficients \( a_1, a_2, \ldots, a_m \) in (2.2) must be the same for all rank outcomes \( \{i = 1, 2, \ldots, k-1\} \), a limitation of choosing to model the cumulative probabilities.

Recall that, given a sample with \( n \) independent trials, where each trial results in exactly one of \( k \) rank outcomes, the probability of observing frequencies \( \{n_i\} \), with frequency \( n_i \) for the \( i \)th outcome, is

\[
\frac{n!}{n_1!n_2!\cdots n_k!} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}, \quad n = n_1 + n_2 + \cdots + n_k.
\]
where \( p_i = p_i(x) \) is the marginal probability for rank outcome \( i \), which can be derived from the cumulative probabilities given in (2.2). Therefore, the parameters for model (2.2) can be estimated by using the maximum likelihood approaches, given a sample for the observed rank-outcome frequencies.

The proportion-odd (or ordered logistic regression) model, a commonly used ordinal model, is given by

\[
\log \left( \frac{P(R \leq i | x)}{P(R > i | x)} \right) = b_i + a_1 x_1 + a_2 x_2 + \cdots + a_m x_m
\]

where \( F_i(x) = P(R \leq i | x) \)

\[
= \frac{1}{1 + \exp(-b_i - a_1 x_1 - a_2 x_2 - \cdots - a_m x_m)}
= F(b_i + a_1 x_1 + a_2 x_2 + \cdots + a_m x_m).
\]

where \( F(x) = 1/(1 + \exp(-x)) \) is the standard logistic cumulative probability distribution. Thus, the proportion-odd model is a special case of the ordinal regression model (2.2), with the link function given by the inverse of the standard logistic cumulative distribution, ie, the logit function.

Ordinal regression models are implemented by SAS, with options for different link functions, including the inverse of standard logistic and the inverse of standard normal cumulative distributions (ie, the logit and probit functions).

### 2.2 Forward/backward continuation ratio model

Recall that the logit function is defined as \( \text{logit}(p) = \log[p/(1 - p)] \) for \( 0 < p < 1 \). Given scenario \( x \) and rank-outcome value \( i \), the forward and backward logistic continuation ratio models, respectively, are formulated as

\[
\text{logit} \left( \frac{P(R = i | x)}{P(R > i | x)} \right) = b_i + a_1 x_1 + a_2 x_2 + \cdots + a_m x_m, \quad (2.4a)
\]

\[
\text{logit} \left( \frac{P(R = i | x)}{P(R \leq i | x)} \right) = b_i + a_1 x_1 + a_2 x_2 + \cdots + a_m x_m. \quad (2.4b)
\]

The coefficients \( a_1, a_2, \ldots, a_m \) do not depend on index \( i \leq k - 1 \). Let \( \tilde{p}_i(x) \) denote the forward ordinal probability \( P(R = i | x, R > i) \) or the backward ordinal probability \( P(R = i | x, R \leq i) \). Then we can reformulate (2.4a) and (2.4b) as

\[
\tilde{p}_i(x) = \frac{1}{1 + \exp(b_i + a_1 x_1 + a_2 x_2 + \cdots + a_m x_m)}
= \Phi(b_i + a_1 x_1 + a_2 x_2 + \cdots + a_m x_m), \quad (2.5)
\]

where \( \Phi \) denotes standard logistic cumulative distribution. This means the logistic forward continuation ratio model is formulated by the forward ordinal probabilities.
forward ordinal probability models for point-in-time probability of default term structure 7

for rank outcomes, with the inverse of the standard logistic cumulative distribution, ie, the logit function, as the link function. The probit continuation ratio model can be formulated similarly using the inverse of the standard normal cumulative distribution (ie, the probit function) as the link function.

3 THE PROPOSED FORWARD ORDINAL MODEL

With ordinal regression model (2.2) and continuation ratio models (2.4a) and (2.4b), the sensitivities for all the rank outcomes are the same, though the intercept can differ between rank outcomes. In this section we propose an ordinal model based on forward ordinal probabilities. This forward ordinal model allows the sensitivities of the rank outcomes to be differentiated.

3.1 The mathematical setup

We assume, given the rank outcome will not be less than \( i \), ie, \( R \geq i \), that there is a latent variable \( y_i \) given by

\[
y_i = -b_i - r_i(a_1x_1 + a_2x_2 + \cdots + a_mx_m) + \varepsilon_i \tag{3.1}
\]

such that \( R > i \) when \( y_i > 0 \) and \( R = i \) if \( y_i \leq 0 \), where \( \varepsilon_i \) is a random variable with zero mean, independent of \( x = (x_1, x_2, \ldots, x_m) \). The coefficients \( \{a_1, a_2, \ldots, a_m\} \) do not depend on index \( i \leq k - 1 \).

By appropriate scaling of both sides of (3.1), we can assume the standard deviation of \( \varepsilon_i \) is 1. We assume that \( \varepsilon_i \) is standard normal. Let \( \Phi \) denote the cumulative distribution for \( \varepsilon_i \). Then, by (3.1), the forward ordinal probability \( \tilde{p}_i(x) \) is

\[
\tilde{p}_i(x) = \Phi(b_i + r_i(a_1x_1 + a_2x_2 + \cdots + a_mx_m)). \tag{3.2}
\]

Let \( c(x) = (a_1x_1 + a_2x_2 + \cdots + a_mx_m) \). We call \( c(x) \) a common score and \( r_i \) the sensitivity for the rank value \( i \leq k - 1 \) with respect to \( c(x) \). For IFRS 9 expected loss estimation and CCAR stress testing, \( c(x) \) can include both systematic and entity-specific risk drivers.

Note that, with model (3.2), an increase (respectively, decrease) in the norm of the parameter vector \( (a_1, a_2, \ldots , a_m) \) during parameter estimation can propagate to the sensitivity parameter vector \( (r_1, r_2, \ldots , r_{k-1}) \) by scaling down (respectively, up). To prevent unnecessary disturbance of parameter estimation and ensure estimation convergence, the following constraints can be imposed:

\[
a_1^2 + a_2^2 + \cdots + a_m^2 = 1. \tag{3.3a}
\]

In practice, the sign of the coefficient \( a_i \) is predetermined. For example, default risk increases as unemployment rate increases. We thus require the unemployment rate
coefficient in the model to be positive. In this case, we can assume that all \( \{a_i\} \) are nonnegative by an appropriate sign scaling to the corresponding variable. Then the following linear constraint can be imposed:

\[
a_1 + a_2 + \cdots + a_m = 1.
\]

(3.3b)

Let \( c(x) = (a_1x_1 + a_2x_2 + \cdots + a_mx_m) \). In the case when the variables \( x_1, x_2, \ldots, x_m \) are common to all entities (e.g., the macroeconomic variables), we obtain the following model, assuming normality for \( c(x) \) with mean \( u \) and standard deviation \( v \):

\[
\tilde{p}_i(x) = \Phi(c_i \sqrt{1 + (r_i v)^2} + r_i (a_1x_1 + a_2x_2 + \cdots + a_mx_m - u)),
\]

(3.4)

where \( c_i \) is the threshold value estimated directly by taking the inverse, \( \Phi^{-1} \), of the long-run average for forward ordinal probability, which can be estimated directly from the sample. Model (3.4) is derived from (3.2) by a well-known lemma (Rosen and Saunders 2009) for the expectation with respect to \( s \):

\[
E_s[\Phi(a + bs)] = \Phi\left(\frac{a}{\sqrt{1 + b^2}}\right), \quad s \sim N(0, 1).
\]

With model (3.4), estimation is required for parameters \( \{a_1, a_2, \ldots, a_m\} \) and \( \{r_i\} \), but not the intercepts \( \{b_i\} \).

### 3.2 The loglikelihood function given the observed rank frequencies

In this section we show the loglikelihood and its concavity for observing rank-outcome frequencies by using the forward ordinal probabilities \( \{\tilde{p}_i(x) \mid i = 1, 2, \ldots, k\} \).

Given a scenario \( x = (x_1, x_2, \ldots, x_m) \), let \( n_i \) denote the corresponding observed frequency for the \( i \)th rank value. Let

\[
n = n_1 + n_2 + \cdots + n_k.
\]

(3.5a)

Define

\[
s_i = n - (n_1 + n_2 + \cdots + n_{i-1}) = n_k + n_{k-1} + \cdots + n_i.
\]

(3.5b)

We focus on the conditional probability space given that the rank value of the outcome \( R \) is not less than \( i \). The loglikelihood of observing frequency \( n_i \) for the \( i \)th rank value and frequency \( s_i - n_i \) for rank values larger than \( i \), given \( x = (x_1, x_2, \ldots, x_m) \), is

\[
L_i(x) = (s_i - n_i) \log[1 - \tilde{p}_i(x)] + n_i \log[\tilde{p}_i(x)]
\]

(3.6)

up to a summand given by the logarithm of a binomial coefficient, which is independent of the model parameters of model (3.2) and (3.4), assuming the occurrence of the \( i \)th rank value is a binary event.
Let $L(x, i, i + h)$ denote the loglikelihood, over this probability space, of observing multiple frequencies $\{n_i, n_{i+1}, \ldots, n_{i+h}\}$, for rank values $\{i, i + 1, \ldots, i + h\}$, and the frequency $s_{i+h+1} = n_k + n_{k-1} + \cdots + n_{i+h+1}$ for rank values larger than $i + h$. We have the following proposition.

**Proposition 3.1** The equations

\[
L(x, i, i + h) = L_i(x) + L_{i+1}(x) + \cdots + L_{i+h}(x), \quad (3.7a)
\]

\[
L(x, 1, k) = L_1(x) + L_2(x) + \cdots + L_k(x) \quad (3.7b)
\]

hold, up to a summand given by the logarithms of some binomial coefficients (independent of the parameters in models (3.2) and (3.4)).

**Proof** We show only (3.7b); the proof for (3.7a) is similar. For simplicity, we write $F_i$ and $\tilde{p}_i$, respectively, for $F_i(x)$ and $\tilde{p}_i(x)$. The marginal probability of the event $\{R = i \mid x\}$ is $(1 - F_{i-1}(x))\tilde{p}_i(x)$. Thus, the probability of observing a frequency $n_i$ for the $i$th rank value is $(1 - F_{i-1})^{n_i} \tilde{p}_i^{n_i}$ up to a multiplicative factor given by the binomial coefficient. Consequently, the probability of observing frequencies $\{n_i\}_{i=1,2,\ldots,k}$ with $n_i$ for the $i$th rank value is

\[
\Delta = \tilde{p}_1^{n_1} \tilde{p}_2^{n_2} \cdots \tilde{p}_k^{n_k} (1 - F_1)^{n_2} (1 - F_2)^{n_3} \cdots (1 - F_{k-1})^{n_k} \quad (3.8)
\]

up to a constant factor given by some binomial coefficients. By (2.1d), we have

\[
(1 - F_1)^{n_2} (1 - F_2)^{n_3} \cdots (1 - F_{k-1})^{n_k} = (1 - \tilde{p}_1)^{n_2} [(1 - \tilde{p}_1) (1 - \tilde{p}_2)]^{n_3} \cdots [(1 - \tilde{p}_1)(1 - \tilde{p}_2) \cdots (1 - \tilde{p}_{k-1})]^{n_k} = (1 - \tilde{p}_1)^{n_2 + n_3 + \cdots + n_k} (1 - \tilde{p}_2)^{n_3 + n_4 + \cdots + n_k} \cdots (1 - \tilde{p}_{k-1})^{n_k} = (1 - \tilde{p}_1)^{s_1-n_1} (1 - \tilde{p}_2)^{s_2-n_2} \cdots (1 - \tilde{p}_{k-1})^{s_{k-1}-n_{k-1}}. \quad (3.9)
\]

This follows from (3.5b). Thus, by (3.8), we have the corresponding loglikelihood:

\[
\log \Delta = [n_1 \log(\tilde{p}_1) + (s_1 - n_1) \log(1 - \tilde{p}_1)] + [n_2 \log(\tilde{p}_2) + (s_2 - n_2) \log(1 - \tilde{p}_2)] + \cdots + [n_k \log(\tilde{p}_k)] = [n_1 \log(\tilde{p}_1) + (s_1 - n_1) \log(1 - \tilde{p}_1)] + [n_2 \log(\tilde{p}_2) + (s_2 - n_2) \log(1 - \tilde{p}_2)] + \cdots + [n_k \log(\tilde{p}_k) + (s_k - n_k) \log(1 - \tilde{p}_k)] = L_1(x) + L_2(x) + \cdots + L_k(x), \quad (3.10)
\]

where the second equality follows from the fact that $(s_k - n_k) = 0$. □

A function is log concave if its logarithm is concave. If a function is concave, a local maximum is a global maximum, and the function is unimodal. This property is important for finding the maximum likelihood estimate.
Proposition 3.2  The loglikelihood function (3.7a) or (3.7b), with \( \Phi \) the standard normal cumulative probability distribution, is concave in the following two cases:

(a) as a function of the \( r \)-parameters \( \{ r_i \} \), or the \( b \)-parameters \( \{ b_j \} \), and the \( a \)-parameters \( \{ a_1, a_2, \ldots, a_m \} \), where \( \tilde{p}_i(x) \) is given by (3.2);

(b) as a function of the \( a \)-parameters \( \{ a_1, a_2, \ldots, a_m \} \), or as a function of the \( r \)-parameters \( \{ r_i \} \), where \( \tilde{p}_i(x) \) is given by (3.4).

Proof It is well known that the standard normal cumulative distribution is log concave. Also, if \( f(x) \) is log concave, then so is \( f(Az + b) \), where \( Az + b : \mathbb{R}^m \to \mathbb{R}^1 \) is any affine transformation from the \( m \)-dimensional Euclidean space to the one-dimensional Euclidean space. Therefore, both the cumulative distributions \( \Phi(x) \) and \( \Phi(-x) \) are log concave. For Proposition 3.2(a), the concavity of (3.7a) and (3.7b) follows from the fact that the sum of concave functions is again concave. For Proposition 3.2(b), the concavity of (3.7a) and (3.7b) as a function of parameters \( \{ a_1, a_2, \ldots, a_m \} \) is also immediate.

For Proposition 3.2(b) and the concavity of (3.7a) and (3.7b) as a function of the \( r \)-parameters \( \{ r_i \} \), recall that \( \tilde{p}_i(x) \) in (3.4) is given by

\[
\tilde{p}_i(x) = \Phi(c_i \sqrt{1 + (r_i v)^2} + r_i (a_1 x_1 + a_2 x_2 + \cdots + a_m x_m - u)).
\]

It suffices to show that the second derivative of the function

\[
L(r) = \log[\Phi(b \sqrt{1 + r^2 + ra})]
\]

is nonpositive for any constants \( a \) and \( b \). This is because either \( \log(\tilde{p}_i(x)) \) or \( \log(1 - \tilde{p}_i(x)) \) will have the form of (3.11) after some appropriate scaling transformations. The second derivative \( d^2[L(r)]/dr^2 \) is given by

\[
\left( \frac{b r}{\sqrt{1 + r^2}} + a \right)^2 - \frac{[\varphi(b \sqrt{1 + r^2 + ra})]^2}{\Phi(b \sqrt{1 + r^2 + ra})^2} + \frac{\varphi'(b \sqrt{1 + r^2 + ra})}{\Phi(b \sqrt{1 + r^2 + ra})} \varphi(b \sqrt{1 + r^2 + ra}) \varphi'(b \sqrt{1 + r^2 + ra}) (1 + r^2)^{-3/2} \right] \Phi(b \sqrt{1 + r^2 + ra}) = I + II, \tag{3.12}
\]

where \( \varphi \) and \( \varphi' \) denote the first and second derivatives of \( \Phi \). Because the factor in the first summand of (3.12),

\[
\left[ \frac{[\varphi(b \sqrt{1 + r^2 + ra})]^2}{\Phi(b \sqrt{1 + r^2 + ra})^2} + \frac{\varphi'(b \sqrt{1 + r^2 + ra})}{\Phi(b \sqrt{1 + r^2 + ra})} \right],
\]

corresponds to the second derivative of \( \log \Phi(z) \) (with respect to \( z = b \sqrt{1 + r^2 + ra} \)), it is nonpositive. Thus, the first summand in (3.12) is nonpositive. The second summand in (3.12) is nonpositive if \( b \leq 0 \). For the \( b > 0 \) case, we can change \( b \) back to the negative case using the function \( F(x) = \Phi(-x) \) and repeat the same discussion to obtain nonpositivity of the second derivative of (3.11). \( \square \)
4 PARAMETER ESTIMATION BY MAXIMUM LIKELIHOOD APPROACHES

In this section we propose an algorithm for parameter estimation for models (3.2) and (3.4) by maximizing the loglikelihood for observing rank-outcome frequencies. This generic algorithm works for one forward ordinal model. For modeling rating migration for a risk-rated portfolio, multiple forward ordinal models are required: one for each of the nondefault initial risk ratings (see Section 5 for the model formulation and the adapted algorithm for parameter fitting).

4.1 Estimation of parameters for model (3.2)

The algorithm proposed is essentially a heuristic hard expectation maximization algorithm.

Parameter initialization: initially, set \(r_1, r_2, \ldots, r_{k-1}\) to 1. Estimate the parameters \(\{a_1, a_2, \ldots, a_m\}\) and \(\{b_1, b_2, \ldots, b_{k-1}\}\) without constraints (3.3a) and (3.3b), by maximizing the loglikelihood of (3.7b). Recall that (3.7b) is concave by Proposition 3.2(a). Therefore, global maximum estimates are guaranteed. Rescale the \(a\)-parameter estimates by a scalar \(\rho > 0\) to make \((a_1, a_2, \ldots, a_m)\) a unit vector, and then set each of \(r_1, r_2, \ldots, r_{k-1}\) to \(1/\rho\). This completes the initialization for all parameters.

Step 1: assume that the parameters \(\{a_1, a_2, \ldots, a_m\}\) and \(\{b_1, b_2, \ldots, b_{k-1}\}\) are given. Estimate the sensitivities \(r_1, r_2, \ldots, r_{k-1}\) by maximizing the loglikelihood of (3.7b).

Recall that, by Proposition 3.2(a), global maximum estimates are guaranteed.

Step 2: assume that the sensitivities \(r_1, r_2, \ldots, r_{k-1}\) are given. Estimate the parameters \(\{a_1, a_2, \ldots, a_m\}\) and \(\{b_1, b_2, \ldots, b_{k-1}\}\) by maximizing the loglikelihood of (3.7b).

Global maximum estimates are granted by Proposition 3.2(a). Rescale the \(a\)-parameter estimates by a scalar \(\rho > 0\) to make \((a_1, a_2, \ldots, a_m)\) a unit vector, and then scale the vector \((r_1, r_2, \ldots, r_{k-1})\) by \(1/\rho\).

Step 3: iterate the above three steps until a convergence is reached.

Steps 1 and 2 are repeated until convergence is reached, ie, the maximum deviation for all parameter estimates for \(\{b_1, b_2, \ldots, b_{k-1}\}\), \(\{a_1, a_2, \cdots, a_m\}\) and \(\{r_1, r_2, \ldots, r_{k-1}\}\), in two consecutive iterations, is less than \(10^{-4}\).

We implement the above three-step optimization process by using the SAS procedure PROC NLMIXED.
4.2 Estimation of parameters for model (3.4)

For model (3.4), we follow the steps above to fit for the coefficients \( \{a_1, a_2, \ldots, a_m\} \) for common score \( c(x) = (a_1 x_1 + a_2 x_2 + \cdots + a_m x_m) \). When this common score is known, we estimate \( \{r_1, r_2, \ldots, r_{k-1}\} \) by maximizing (3.7b), with \( \tilde{p}_i(x) \) given by (3.4). Global maximum estimates are guaranteed by Proposition 3.2(b).

5 AN EMPIRICAL EXAMPLE: RATING MIGRATION PROBABILITY AND PROBABILITY OF DEFAULT TERM STRUCTURE FOR A COMMERCIAL PORTFOLIO

In this section we apply the proposed ordinal model to estimate the rating transition probability for a risk-rated commercial portfolio, where a point-in-time PD term structure for IFRS 9 expected credit loss estimation and CCAR stress testing is derived.

The sample contains quarterly rating migration frequencies between 2006 Q3 and 2016 Q4 for a commercial portfolio, created synthetically by scrambling the default rate using an appropriate scaling. There are twenty-one risk ratings, with \( R_{21} \) the default rating and \( R_1 \) the best quality rating.

Because we are more concerned with the default outcome and default risk, we model rating migration probability with the backward ordinal model, starting with a rating-level default risk. As noted in Section 2, a backward ordinal model can be viewed as a forward ordinal model after an appropriate reversion of the outcome order and of the index of the resulting forward ordinal probability.

The backward ordinal model is benchmarked using the rating transition model based on the Merton model framework proposed by Yang (2016). Additional benchmark comments for SAS ordinal regression using SAS PROC LOGISTIC are given at the end of this section.

5.1 The backward ordinal and benchmark models for IFRS 9 expected credit loss estimation and CCAR stress testing

5.1.1 Formulation of the models

Backward ordinal model for rating migration probability.

Given a nondefault initial risk rating \( R_i \) at the beginning of the quarter, there are twenty-one possible ordinal outcomes at the end of the quarter: an entity can migrate to a default rating or any of the other twenty ratings. Given a scenario \( x = (x_1, x_2, \ldots, x_m) \), let \( \tilde{p}_{ij}(x) \) denote the backward ordinal probability that the rating \( R_i \) migrates to rating \( R_j \) given that it will migrate only to a rating with rank no higher than \( j \). Bearing in mind that a backward ordinal model can be viewed as a forward ordinal model by an outcome order and probability index reversion, we can
model $\tilde{p}_{ij}(x)$ using models (3.2) and (3.4), respectively, as follows:

$$\tilde{p}_{ij}(x) = \Phi(b_{ij} + r_{ij}(a_1x_1 + a_2x_2 + \cdots + a_mx_m)), \quad (5.1a)$$

$$\tilde{p}_{ij}(x) = \Phi(c_{ij}\sqrt{1 + (r_{ij}v)^2} + r_{ij}(a_1x_1 + a_2x_2 + \cdots + a_mx_m - u)), \quad (5.1b)$$

We assume that, for each initial rating $R_i$, the sensitivity parameters $r_{ij}$ are the same for rank-outcome values $j$ when $i < j < 21$ (downgrade) or $1 < j < i$ (upgrade). Denote the downgrade sensitivity by $r_{id}$ and the upgrade sensitivity by $r_{iu}$. Let $r_{id}$ and $r_{is}$ denote the sensitivities for outcome cases $j = 21$ (default) and $j = i$ (stay), respectively. Then (5.1a) and (5.1b) reduce to

$$\tilde{p}_{ij}(x) = \Phi(b_{ij} + r_i(a_1x_1 + a_2x_2 + \cdots + a_mx_m)), \quad (5.2a)$$

$$\tilde{p}_{ij}(x) = \Phi(c_{ij}\sqrt{1 + (r_{i}v)^2} + r_i(a_1x_1 + a_2x_2 + \cdots + a_mx_m - u)), \quad (5.2b)$$

where $r_i = r_{id}, r_{id}, r_{is}, r_{iu}$, respectively, for default, downgrade, stay and upgrade. The marginal probability is given by

$$p_{ij}(x) = (1 - F_{ij}(x))\tilde{p}_{ij}(x),$$

where $F_{ij}(x) = p_{i21}(x) + p_{i20}(x) + \cdots + p_{i22-j}(x)$ is the cumulative probability. The constraint (3.3a) (respectively, (3.3b)) is imposed for the proposed backward ordinal model (5.2a) (respectively, (5.2b)).

### Rating transition model under the Merton model framework

The point-in-time rating transition probability model, based on the Merton framework, was proposed by Miu and Ozdemir (2009), and extended by Yang (2016) to facilitate rating-level sensitivity for CCAR stress testing and IFRS 9 expected credit loss estimation.

Let $t_{ij}(x)$ denote the transition probability from an initial rating $R_i$ at the beginning of the quarter to rating $R_j$ at the end of the quarter, given a macroeconomic scenario $x = (x_1, x_2, \ldots, x_m)$. Let $\Phi$ denote the standard normal cumulative distribution. Under the Merton model framework (Gordy 2003; Merton 1974; Miu and Ozdemir 2009; Vasicek 2002), it can be shown (Yang 2016) that

$$t_{ij}(x) = \Phi(\tilde{q}_{i(k-j+1)} + \tilde{r}_i\tilde{c}_i(x)) - \Phi(\tilde{q}_{i(k-j)} + \tilde{r}_i\tilde{c}_i(x))$$

$$= \Phi[\tilde{q}_{i(k-j+1)} + \tilde{r}_i(\tilde{a}_1\tilde{x}_1 + \tilde{a}_2\tilde{x}_2 + \cdots + \tilde{a}_m\tilde{x}_m)]$$

$$- \Phi[\tilde{q}_{i(k-j)} + \tilde{r}_i(\tilde{a}_1\tilde{x}_1 + \tilde{a}_2\tilde{x}_2 + \cdots + \tilde{a}_m\tilde{x}_m)], \quad (5.3)$$

where $\tilde{q}_{ih} = q_{ih}\sqrt{1 + r_i^2}$; the quantities $\{q_{ij}\}$ are the threshold values given by $q_{ij} = \Phi^{-1}(\tilde{p}_{ij})$, where $\tilde{p}_{ij}$ is the through-the-cycle transition probability from $R_i$.

www.risk.net/journals   Journal of Risk Model Validation

To subscribe to a Risk Journal visit subscriptions.risk.net/journals or email info@risk.net
to $R_j$, which can be estimated directly from the historical sample. The sensitivity parameter $\tilde{r}_1$ is the same for all rank outcomes for a given rating $R_j$.

The credit index $c_i(x) = a_1 \tilde{x}_1 + a_2 \tilde{x}_2 + \cdots + a_m \tilde{x}_m$ is derived by a normalization from a linear combination $a_1 \tilde{x}_1 + a_2 \tilde{x}_2 + \cdots + a_m \tilde{x}_m$, with which the model $\{p_i(x)\}$ best predicts the portfolio default risk, in the sense of maximum likelihood, for observing default frequencies, where

$$p_i(x) = \Phi[c_i + \tilde{r}_1(a_1 \tilde{x}_1 + a_2 \tilde{x}_2 + \cdots + a_m \tilde{x}_m)]$$

(5.4)

is a model predicting the PD for rating $R_i$ and no constraint is imposed for intercept $c_i$. The quantity $\tilde{r}_1$ is driven by

$$\tilde{r}_1 = \frac{r_i \lambda}{\sqrt{1 + r_i^2(1 - \lambda^2)}}, \quad 0 \leq \lambda \leq 1,$$

(5.5)

where $r_i = \sqrt{p_i} / \sqrt{1 - \rho_i}$, and $\rho_i$ is the asset correlation in the Merton model for rating $R_i$ (Yang 2016).

**Remark 5.1** We can choose to fit for $\{a_1, a_2, \ldots, a_m\}$ without constraint (5.5). The unconstrained result is always better than the constrained one in the sense of a higher likelihood value.

### 5.1.2 Fitting for parameters

We focus on macroeconomic scenarios and consider parameter fitting only for models (5.2b) and (5.3).

For models (5.2b) and (5.3), parameter fitting follows the two steps below.

1. Fit for the macroeconomic variable coefficients $\{a_1, a_2, \ldots, a_m\}$ by using maximum likelihood to observe rating-level default frequencies, with default probability $p_i(x)$ for rating $R_i$ given by (5.4) without constraint (5.5). This can be done via Steps 1–3 in Section 4.

2. When credit index $c_i(x) = a_1 \tilde{x}_1 + a_2 \tilde{x}_2 + \cdots + a_m \tilde{x}_m$ has been determined, we need to fit only for the risk sensitivity parameters $\{r_i\}$ for model (5.3), and $\{r_{id}, r_{id}, r_{is}, r_{iu}\}$ for model (5.2b), for all risk ratings. For model (5.3), we can choose to fit for $\{r_i\}$ either separately for each rating $R_i$, or as a combination of all ratings, using the appropriate likelihood function (3.7b) for all rating migration frequencies, or (3.7a) for downgrade or default frequencies only. The corresponding loglikelihood function is concave by Proposition 3.2(b). For model (5.2b), we fit for each of the four groups $\{r_{id}, r_{id}, r_{is}, r_{iu}\}$ separately, using the appropriate likelihood function (3.7a) for the corresponding migration frequency.
In general, monotonicity is imposed for sensitivity between ratings; specifically, we require that \( \{r_i\} \), \( \{r_{id}\} \) and \( \{r_{id}\} \) are nondecreasing and that \( \{r_{is}\} \) and \( \{r_{iu}\} \) are nonincreasing for a higher risk rating.

### 5.2 Validation results

We use the following labels for the backward ordinal and the benchmark models.

- BORD: the backward ordinal model (5.2b).
- RTGM: the rating migration model based on the Merton model framework (5.3).

Both models use the same variables, provided by the US Federal Reserve:

- three-month treasury bill interest rate \( (v_1) \);
- unemployment rate \( (v_2) \).

The macro coefficients for credit index \( ci(x) = a_1 x_1 + a_2 x_2 + \cdots + a_m x_m \) are fitted as described in Section 5.1 in the same way for both the BORD and RTGM, so both models have the same macro coefficient estimates. Table 1 records the estimates for these two variable coefficients, with the variable \( p \)-values \( p_1 \) and \( p_2 \).

For the BORD the sensitivity parameter estimates are reported as in Table 2 for twenty nondefault ratings for default, downgrade and stay, with monotonicity constraint being imposed. The sensitivity estimates for upgrade are all close to zero (reflecting the fact that the upgrade probability is slim in the stress period), and are not shown in the table. The RTGM estimates the sensitivity parameters by maximum likelihood for observing only the default frequency. Thus, for default it has the same sensitivity estimates as the BORD.

Table 3 shows the backtest performance for two \( R \)-squared-based models for predicting portfolio cumulative default rates for one, four, six, twelve and sixteen quarters.

The results show performance improves for the BORD when the sensitivities are differentiated between migrations to default, downgrade, stay and upgrade. This improvement is a trade-off with the addition of more sensitivity parameters.
TABLE 2  Sensitivity parameter estimates.

<table>
<thead>
<tr>
<th>Migration</th>
<th>Default</th>
<th>Downgrade</th>
<th>Stay</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.001</td>
<td>0.128</td>
<td>1.992</td>
</tr>
<tr>
<td>2</td>
<td>0.001</td>
<td>0.129</td>
<td>1.992</td>
</tr>
<tr>
<td>3</td>
<td>0.002</td>
<td>0.130</td>
<td>0.258</td>
</tr>
<tr>
<td>4</td>
<td>0.003</td>
<td>0.131</td>
<td>0.258</td>
</tr>
<tr>
<td>5</td>
<td>0.004</td>
<td>0.132</td>
<td>0.258</td>
</tr>
<tr>
<td>6</td>
<td>0.005</td>
<td>0.133</td>
<td>0.144</td>
</tr>
<tr>
<td>7</td>
<td>0.017</td>
<td>0.134</td>
<td>0.133</td>
</tr>
<tr>
<td>8</td>
<td>0.059</td>
<td>0.135</td>
<td>0.051</td>
</tr>
<tr>
<td>9</td>
<td>0.059</td>
<td>0.136</td>
<td>0.051</td>
</tr>
<tr>
<td>10</td>
<td>0.059</td>
<td>0.151</td>
<td>0.051</td>
</tr>
<tr>
<td>11</td>
<td>0.059</td>
<td>0.228</td>
<td>0.051</td>
</tr>
<tr>
<td>12</td>
<td>0.059</td>
<td>0.229</td>
<td>0.051</td>
</tr>
<tr>
<td>13</td>
<td>0.059</td>
<td>0.230</td>
<td>0.051</td>
</tr>
<tr>
<td>14</td>
<td>0.059</td>
<td>0.231</td>
<td>0.051</td>
</tr>
<tr>
<td>15</td>
<td>0.059</td>
<td>0.248</td>
<td>0.051</td>
</tr>
<tr>
<td>16</td>
<td>0.059</td>
<td>0.249</td>
<td>0.050</td>
</tr>
<tr>
<td>17</td>
<td>0.060</td>
<td>0.362</td>
<td>0.050</td>
</tr>
<tr>
<td>18</td>
<td>0.060</td>
<td>0.504</td>
<td>0.050</td>
</tr>
<tr>
<td>19</td>
<td>0.769</td>
<td>1.139</td>
<td>0.050</td>
</tr>
<tr>
<td>20</td>
<td>0.769</td>
<td>—</td>
<td>0.050</td>
</tr>
</tbody>
</table>

TABLE 3  $R$-squared values for predicting portfolio cumulative default rate.

<table>
<thead>
<tr>
<th>Quarters</th>
<th>1</th>
<th>4</th>
<th>8</th>
<th>12</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BORD</td>
<td>0.420</td>
<td>0.575</td>
<td>0.570</td>
<td>0.792</td>
<td>0.777</td>
</tr>
<tr>
<td>RTGM</td>
<td>0.420</td>
<td>0.558</td>
<td>0.518</td>
<td>0.726</td>
<td>0.660</td>
</tr>
</tbody>
</table>

We end this section by commenting on an additional benchmark based on SAS ordinal regression using SAS PROC LOGISTIC, with both logit and probit as the link functions, via the “class” and “by” options.

When the “by” statement is used for initial ratings, for each initial rating $R_i$, SAS fits an ordinal regression model of the form

$$F_{ij}(x) = \Phi(b_{ij} + a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{im}x_m)$$

for the cumulative probability for rank outcome $j < 21$. This model has redundant coefficients (depending on the rating index $i$) for such a short time series sample,
causing an overfitting issue. More importantly, it is not formulated using a common score or sensitivity. We do not recommend this model.

When the “class” statement is used, the initial risk rating is treated as a class variable in the model, and for each initial risk rating $R_i$, SAS fits an ordinal model of the form

$$F_{ij}(x) = \Phi(b_{ij} + a_1x_1 + a_2x_2 + \cdots + a_mx_m)$$

for the cumulative probability for the rank outcome $j < 21$. The intercept vectors for initial risk rating $R_i$ and $R_1$ satisfy the following equation:

$$(b_{i1}, b_{i2}, \ldots, b_{i20}) = (d_i + b_{11}, d_i + b_{12}, \ldots, d_i + b_{120})$$  \hspace{1cm} (5.6)

with constant $d_i$ corresponding to the $i$th level of the class variable. That is, the intercept vector for $R_i$ is a translation of the intercept vector for $R_1$. As expected, this model fails to predict the default risk and other migration risk. It overestimates PD for the high risk ratings $R_{19}$ and $R_{20}$, and significantly underestimates the PD for other ratings. We do not recommend this model.

6 CONCLUSIONS

Ordinal regression models are widely used for modeling rating migration. Results are generally not very optimistic, partly due to the lack of flexibility with respect to the sensitivity (between rank outcomes and between risk ratings). In this paper, we proposed an ordinal model based on forward ordinal probabilities. Under this model, forward ordinal probabilities are formulated by a common score plus rank- and rating-specific sensitivity. The rank-specific sensitivity allows a risk rating to respond to its own migration patterns to default, downgrade, stay and upgrade accordingly. Empirical results show our model is more robust than the rating transition model based on the Merton model framework. Unlike the latter model, which allows only one sensitivity parameter for all rank outcomes for a rating, and uses only systematic risk drivers, our proposed ordinal model differentiates sensitivities between outcomes and includes entity-specific risk drivers. No estimation for asset correlation is required. The model can be implemented by using, for example, the SAS PROC NLMIXED procedure. This forward ordinal model will provide a useful tool for practitioners to estimate the point-in-time PD term structure for IFRS 9 expected credit loss estimation as well as multiperiod scenario loss projection for CCAR stress testing.

DECLARATION OF INTEREST

The author reports no conflicts of interest. The author alone was responsible for the content and writing of the paper. The views expressed in this paper are not necessarily those of Royal Bank of Canada or any of its affiliates.
ACKNOWLEDGEMENTS

The author thanks both referees for many insights and valuable suggestions. Special thanks go to Carlos Lopez for his consistent support for this research, and Clovis Sukam for his critical reading of the manuscript. The author also thanks Glenn Fei, Biao Wu, Zunwei Du, Wallace Law and Kaijie Cui for many valuable discussions.

REFERENCES


Research Paper

Bayesian analysis in an aggregate loss model: validation of the structure functions

A. Hernández-Bastida, J. M. Pérez-Sánchez and M. P. Fernández-Sánchez

Department of Quantitative Methods, University of Granada, Campus Cartuja s/n 18011, Granada, Spain; emails: bastida@ugr.es, pilarfs@ugr.es, josemag@ugr.es

(Received November 4, 2016; revised February 21, 2017; accepted February 21, 2017)

ABSTRACT

The aggregate loss, $S$, is the sum of the individual claim sizes, ie, $S = \sum_{i=1}^{K} X_i$ for $K > 0$, and $S = 0$ for $K = 0$. The distributions of $K$ and $X_i$ are termed the primary distribution and the secondary distribution, respectively. This paper considers the empirical evaluation of a collective risk model with the geometric as the primary distribution and the exponential as the secondary distribution. We develop a Bayesian analysis for three risk variables (number of claims, severity and aggregate loss), and apply it to sixteen real portfolios of policyholders, to illustrate the validity of the model. The net premium is obtained for each variable and the results are compared with those derived from a frequentist approach given by maximum likelihood estimation. From a Bayesian standpoint, the interest of this study lies in its specification of the prior distributions (structure functions) and in how the hyperparameters are elicited. In order to validate the specification of the prior information, we consider two different distributions for structure functions (the beta and uniform distributions) and, in each case, study the determination of the hyperparameters. We analyze the moment and maximum likelihood methods and propose a new possibility based on prior information specified by an expert in terms of quantiles, suggesting three different scenarios.
These comparisons show that in all cases the Bayes premium is practically equal to the net premium with the parameter estimated by maximum likelihood. This finding indicates that in nonlife insurance problems the importance of the data, modeled by the likelihood, outweighs the prior information.

**Keywords:** model selection; risk premiums; validation of prior information; aggregate loss models; uncertainty modeling.

## 1 INTRODUCTION

The coverage of a risk by an insurance company is established with the guarantee of a contract (the policy), and requires the insured to pay a price (the premium). For the insurance company, this premium must incorporate the cost of the claims made plus a profit margin. The first of these costs is stochastic and is discussed in this paper. It is important for an actuary to correctly apply the principles of premium calculation. Basic texts in this respect include Gerber (1979), Heilmann (1989), Hürlimann (1994) and Sarabia et al (2007). For an actuary, a risk is a random variable and a premium calculation principle is a function that assigns a real number (the premium) to the risk. A wide variety of premium calculation principles have been developed in actuarial research. One method for developing these principles (see, for example, Heilmann 1989) is based on the use of loss functions and on minimizing expected losses. The net premium principle is obtained from the quadratic loss function. More precisely, in most nonlife insurance, the risk variable, $\xi$, is an observable random variable (discrete or continuous) with a likelihood density given by $g(y \mid \theta)$. The risk parameter $\theta$ has a probability density function, $u(\theta)$. It is assumed that the marginal distribution,

$$m(y \mid u) = \int g(y \mid \theta)u(\theta)\,d\theta,$$

the posterior distribution,

$$u(\theta \mid data) = \frac{g(y \mid \theta)u(\theta)}{m(y \mid u)},$$

and the predictive distribution,

$$p(z \mid data) = \int g(z \mid \theta)u(\theta \mid data)\,d\theta,$$

can be determined. Under the net premium principle, the true risk premium is defined as

$$NP(\theta) = E[\xi] = \int yg(y \mid \theta)\,dy.$$
Several alternatives are possible for the unknown parameter $\theta$. One possibility is to obtain the point estimation of $\theta$ with a frequentist alternative, thus obtaining a point maximum likelihood estimator of $\theta$ from the available data. Otherwise, the posterior mode can be used, which is a Bayesian approach to the problem. Another possibility consists in eliminating parameter averaging with respect to a probability distribution: via the posterior distribution if data is available, or the prior distribution otherwise. The following definitions are well known.

- The collective premium (CP) is the expected value of $NP(\theta)$ with respect to the prior distribution, which is equal to the expected value of the marginal distribution, ie,
  \[ CP = \int NP(\theta)u(\theta) \, d\theta = \int y m(y \mid u) \, dy. \]

- The Bayes premium (BP) is the expected value of $NP(\theta)$ with respect to the posterior distribution, which is equal to the expected value of the predictive distribution, ie,
  \[ BP = \int NP(\theta)u(\theta \mid \text{data}) \, d\theta = \int z p(z \mid \text{data}) \, dz. \]

The collective risk model (CRM) is described by a frequency distribution for the number of claims $K$ and a sequence of independent and identically distributed nonnegative random variables representing the size of the single claim, $X_i$.

We assume that $K$ and $X_i$ are conditionally independent, given the parameters. This hypothesis is a convenient way to obtain the density of the aggregate loss, the $k$-fold convolution of $f(x)$, in a mathematically manageable way. Without this hypothesis, it would be necessary to know the bi-dimensional distribution of $K$ and $X_i$ in order to determine the convolution. The aggregate loss, $S$, is the sum of the individual claim sizes, ie, $S = \sum_{i=1}^{K} X_i$ for $K > 0$, and $S = 0$ for $K = 0$. The distributions of $K$ and $X_i$ are termed the primary and secondary distributions, respectively. Traditionally, the Poisson has been taken as the primary distribution (termed the compound Poisson model). However, it is well known that the presence of a large volume of data presents a problem of overdispersion, ie, the variance is larger than the mean, and the Poisson distribution does not properly reflect this phenomenon. For this reason, as Nikoloulopoulos and Karlis (2008) pointed out, overdispersed models (in contrast to the simple Poisson model) provide an alternative means of describing these situations. In particular, mixed Poisson distributions, which are obtained by mixing the Poisson with other distributions, have been widely employed for the random variable $K$. For a comprehensive review of this topic, see Klugman (1992) and Nadarajah and Kotz (2006a,b).
We assume that the sample information for a portfolio with $n$ policies has the following elements.

- $k_i$ is the number of claims of the policy $i$ for $i = 1, \ldots, n$. A large number of policies have made no claim.
- $x_j^{(i)}$, $j = 1, \ldots, k_i$, is the size of the single claim in the policy $i$ for $i = 1, \ldots, n$.

From this information, we have the following.

- $s_i = \sum_{j=1}^{k_i} x_j^{(i)}$, $i = 1, \ldots, n$, which is the aggregate loss of the policy $i$. A large number of these aggregate losses will be 0, and $n_0$ is the number of zeros.

In this context, we may be interested in the “number of declared claims” risk variable, which is the most usual case, or in the “severity of each declared claim” variable, or even in the aggregate loss variable for each policy.

This paper focuses on an aggregation of the two first cases, ie, through a convolution of these two variables, so that all three can be studied. We consider a CRM with the geometric distribution (Geo) as the primary distribution and an exponential one (Exp) as the secondary distribution. This is a well-known model, and its analytic development can be consulted in various papers. Hu and Lin (2001) analyzed the geometric compounding model and obtained an interesting upper bound for the tail probability. Willmot (2002) and Willmot et al (2005) used this representation to derive the reliability properties of compound distributions, together with an explicit analytic representation of the premiums of interest in connection with insurance claim modeling.

In this paper, we propose a validation process for several specifications of the prior distributions in order to compare the results with those for a frequentist distribution. In this respect, Jacobs et al (2015) presented a Bayesian-based credit risk stress-testing methodology, concluding that the Bayesian model outperformed the frequentist one.

As the geometric distribution can be obtained from a Poisson mixture, it is overdispersed, and the probability of observing a zero value is always higher than under a simple Poisson distribution with the same mean. This is thus a reasonable option for the purposes of the present study. Further, for situations including common real data cases, the geometric distribution assigns a lower probability to the value 1 (for values of the mode approximately less than 2.4) and a higher one to the values $\{2, 3, \ldots \}$ (for values of the mode approximately less than 0.82). The negative binomial distribution would probably be a better option, but we rejected this possibility in order to limit the number of parameters involved.

The aim of this paper is to analyze the specification and validation of the structure function, taking into account the premium calculation (ie, the net premium), paying special attention to the difficulties encountered in eliciting the hyperparameters. We
analyze the three risk variables (number of claims, severity and the aggregate loss) and apply these to real data sets. From a nonlife insurance industry standpoint, the most relevant cases are those based on the number of claims and aggregate loss, but we also analyze the case of claims severity for completeness. For all cases, the problem of eliciting prior information is carefully addressed by taking a very realistic approach. The values of the premiums from these specifications are systematically compared with those obtained by maximum likelihood. Our (in our opinion very significant) conclusions are that extremely similar values are obtained for all the premiums. In short, data sets in actuarial statistics are usually so large that the sample information tends to dominate the reasonable prior information available for the calculation of premiums.

The rest of this paper is structured as follows. In Sections 2 and 3, we briefly develop the data and the model, respectively. Sections 4–6 consider the problem of the elicitation and validation of the prior information for each of the risk variables. Finally, the main conclusions are presented in Section 7.

2 THE DATA

To apply the study method to real data, we consider sixteen portfolios: those labeled C1–C15 contain data only for the number of claims, while C16 contains information on both the number of claims and the severity (size) of the claims in each policy in the portfolio. The portfolios were obtained from the following sources: C1, Thyrion (1960); C2, Tröblinger (1961); C3, Bühlmann (1970); C4, Johnson and Hey (1971); C5, Lemaire (1985); C6, Besson and Partrat (1992); C7, C8, Denuit (1997); C9, Hossack et al (1999); C10, Morillo and Bermúdez (2003); C11, Vilar Zanón et al (2004); C12, Boucher et al (2007); and C13–C15, Alvarez-Jareño (2009). As a whole, these portfolios were considered by Alvarez-Jareño and Muñíz-Rodríguez (2010). The last portfolio in the set, C16, was obtained from De Jong and Heller (2008). Some descriptive statistics of the severity of C16 are the following: mean $\bar{1}37.27$, standard deviation $\bar{1}056.30$, min $\bar{1}0$ and max $\bar{1}55922.10$. Table 1 summarizes the characteristics of these portfolios.

3 THE MODEL

The number of claims, $K$, follows a geometric distribution with parameter $\theta_1$, $\text{Geo}(\theta_1)$, with probability mass function given by

$$P_{\text{Geo}}[K = k | \theta_1] = \theta_1 (1 - \theta_1)^k$$

where $\theta_1$ is the probability at $k = 0$. The values of the mean and the variance are $E_{\text{Geo}}[K] = (1 - \theta_1)/\theta_1$ and $V_{\text{Geo}}[K] = (1 - \theta_1)/\theta_1^2$, respectively. The moment-generating function is given by $M_1(t, \theta_1) = \theta_1/[1 - (1 - \theta_1)e^t]$. 
### TABLE 1
Data used in the numerical analysis of the sixteen nonlife insurance portfolios analyzed.

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>C8</th>
<th>C9</th>
<th>C10</th>
<th>C11</th>
<th>C12</th>
<th>C13</th>
<th>C14</th>
<th>C15</th>
<th>C16</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7 840</td>
<td>20 592</td>
<td>103 704</td>
<td>370 412</td>
<td>96 978</td>
<td>881 705</td>
<td>57 178</td>
<td>118 700</td>
<td>565 664</td>
<td>122 618</td>
<td>2 196 808</td>
<td>513 814</td>
<td>434 698</td>
<td>378 289</td>
<td>371 481</td>
<td>63 232</td>
</tr>
<tr>
<td>1</td>
<td>1 317</td>
<td>2 651</td>
<td>14 075</td>
<td>46 545</td>
<td>9 240</td>
<td>142 217</td>
<td>5 617</td>
<td>11 468</td>
<td>68 714</td>
<td>21 686</td>
<td>161 913</td>
<td>32 296</td>
<td>39 914</td>
<td>30 518</td>
<td>26 784</td>
<td>4 333</td>
</tr>
<tr>
<td>2</td>
<td>239</td>
<td>297</td>
<td>1 766</td>
<td>3 935</td>
<td>704</td>
<td>18 088</td>
<td>446</td>
<td>930</td>
<td>5 177</td>
<td>4 014</td>
<td>10 976</td>
<td>2 493</td>
<td>3 970</td>
<td>2 629</td>
<td>2 118</td>
<td>271</td>
</tr>
<tr>
<td>3</td>
<td>42</td>
<td>41</td>
<td>255</td>
<td>317</td>
<td>704</td>
<td>2 118</td>
<td>50</td>
<td>90</td>
<td>365</td>
<td>832</td>
<td>10 976</td>
<td>243</td>
<td>435</td>
<td>240</td>
<td>174</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>41</td>
<td>45</td>
<td>28</td>
<td>432</td>
<td>203</td>
<td>74</td>
<td>90</td>
<td>24</td>
<td>224</td>
<td>882</td>
<td>203</td>
<td>74</td>
<td>27</td>
<td>18</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>7</td>
<td>6</td>
<td>3</td>
<td>224</td>
<td>203</td>
<td>8</td>
<td>11</td>
<td>24</td>
<td>68</td>
<td>90</td>
<td>203</td>
<td>8</td>
<td>5</td>
<td>18</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>17</td>
<td>11</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>17</td>
<td>11</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9 461</td>
<td>23 589</td>
<td>119 853</td>
<td>421 240</td>
<td>106 974</td>
<td>1 044 454</td>
<td>63 299</td>
<td>131 182</td>
<td>639 950</td>
<td>149 473</td>
<td>2 370 683</td>
<td>548 830</td>
<td>479 107</td>
<td>411 708</td>
<td>400 579</td>
<td>67 856</td>
</tr>
<tr>
<td></td>
<td>0.82867</td>
<td>0.87295</td>
<td>0.86526</td>
<td>0.87934</td>
<td>0.90656</td>
<td>0.84418</td>
<td>0.90330</td>
<td>0.90485</td>
<td>0.88392</td>
<td>0.82034</td>
<td>0.92666</td>
<td>0.93620</td>
<td>0.90731</td>
<td>0.91883</td>
<td>0.92369</td>
<td>0.93186</td>
</tr>
<tr>
<td></td>
<td>0.2144</td>
<td>0.1442</td>
<td>0.1551</td>
<td>0.1317</td>
<td>0.1011</td>
<td>0.1782</td>
<td>0.1057</td>
<td>0.1036</td>
<td>0.1255</td>
<td>0.2251</td>
<td>0.0788</td>
<td>0.0692</td>
<td>0.1034</td>
<td>0.0889</td>
<td>0.0789</td>
<td>0.0728</td>
</tr>
<tr>
<td></td>
<td>0.2889</td>
<td>0.1639</td>
<td>0.1793</td>
<td>0.1385</td>
<td>0.1074</td>
<td>0.1974</td>
<td>0.1149</td>
<td>0.1115</td>
<td>0.1300</td>
<td>0.2966</td>
<td>0.0847</td>
<td>0.0763</td>
<td>0.1175</td>
<td>0.0984</td>
<td>0.0867</td>
<td>0.2782</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of claims</th>
<th>Number of policies</th>
<th>Frequency at 0</th>
<th>( \bar{k} )</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Note: The table includes the number of claims and frequency for each portfolio, along with the number of policies, frequency at 0, and variance.
If \( k = (k_1, \ldots, k_n) \) is a simple random sample of size \( n \) from a Geo(\( \theta_1 \)) distribution, the likelihood function is \( L_{\text{Geo}}(k \mid \theta_1) = \theta_1^n (1 - \theta_1)^{nk} \), where \( \tilde{k} \) is the sample mean. The maximum likelihood estimator is obtained by direct calculation and is expressed as \( \hat{\theta}_{1}^{\text{ML}} = 1/(1 + \tilde{k}) \). For real data sets, in practice, the likelihood function \( L_{\text{Geo}}(k \mid \theta_1), \theta_1 \in [0, 1] \), is a function that takes extremely small values. As a function of \( \theta_1 \) the maximum is reached at \( \hat{\theta}_1^{\text{ML}} \) and the value of this maximum is \( (\tilde{k})^{nk}/[(1 + \tilde{k})^n+n\tilde{k}] \), which is extremely small. Accordingly, any calculation involving direct algebra with the likelihood function will be very difficult.

As a secondary distribution, we consider the exponential distribution. Let \( X \) be the random variable claim size assumed to follow an exponential distribution with parameter \( \theta_2 \), \( \text{Exp}(\theta_2) \), with a density function given by
\[
f(x \mid \theta_2) = \theta_2 e^{-\theta_2 x} \quad \text{with } x > 0.\]
The expected value and the variance are \( E_{\text{Exp}}[X] = 1/\theta_2 \) and \( V_{\text{Exp}}[X] = 1/\theta_2^2 \), respectively. The moment-generating function is \( M_2(t, \theta_2) = \theta_2/(\theta_2 - t) \).

If \( x = (x_1, \ldots, x_n) \) is a simple random sample of size \( n \) from an \( \text{Exp}(\theta_2) \) distribution, the likelihood function is \( L_{\text{Exp}}(x \mid \theta_2) = \theta_2^n e^{-n\bar{x}\theta_2} \), where \( \bar{x} \) is the sample mean. The likelihood estimator is \( \hat{\theta}_2^{\text{ML}} = 1/\bar{x} \).

For the CRM, the density function of the random variable \( S \) is
\[
f_{\text{CRM}}(s \mid \theta_1, \theta_2) = \begin{cases} \theta_1(1 - \theta_1)\theta_2 e^{-\theta_1 \theta_2 s} & \text{if } s > 0, \\ \theta_1 & \text{if } s = 0. \end{cases}\]
The cumulative distribution function is given by
\[F_{\text{CRM}}(s \mid \theta_1, \theta_2) = 1 - (1 - \theta_1) e^{-\theta_1 \theta_2 s}\text{.}\]
The moment-generating function is obtained from \( M_1(\log M_2(t, \theta_2), \theta_1) \) and is given by \( M_3(t, \theta_1, \theta_2) = \theta_1(\theta_2 - t)/(\theta_1 \theta_2 - t) \). The \( k \)-th order moment is given by \( E_{\text{CRM}}[S^k] = k!/(1 - \theta_1)/(\theta_1 \theta_2)^k \). From this, the expected value and the variance can be expressed as
\[E_{\text{CRM}}[S] = \frac{1 - \theta_1}{\theta_1 \theta_2} = E_{\text{Geo}}[K]E_{\text{Exp}}[X] \quad \text{and} \quad V_{\text{CRM}}[S] = \frac{1 - \theta_2^2}{\theta_1^2 \theta_2^2},\]
respectively.

Let \( s = (s_1, \ldots, s_n) \) be a simple random sample of size \( n \). The likelihood function is given by \( L_{\text{CRM}}(s \mid \theta_1, \theta_2) = \theta_2^n (1 - \theta_1)^{n-n_0} \theta_2^{n-n_0} \exp(-n\bar{s}\theta_1 \theta_2) \), where \( n_0 \) is the number of zeros in the sample and \( \bar{s} \) is the sample mean. The maximum likelihood estimator, by direct calculation, is equal to
\[\left(\hat{\theta}_1, \hat{\theta}_2\right)^{\text{ML}} = \left(\frac{n_0}{n}, \frac{n-n_0}{n_0 \bar{s}}\right)\text{.}\]
4 NUMBER OF CLAIMS

Under the net premium principle in the geometric model, the true risk premium is given by

$$\text{NP}_{\text{Geo}}(\theta_1) = \frac{1 - \theta_1}{\theta_1}.$$  

In a frequentist approach, the parameter is estimated by maximum likelihood, $\hat{\theta}_1^{\text{ML}} = 1/(1 + \tilde{k})$, and an estimation of the true risk premium is given by

$$\tilde{\text{NP}}^{[1]}_{\text{Geo}} = \text{NP}_{\text{Geo}}(\hat{\theta}_1^{\text{ML}}) = \tilde{k}.$$  

The penultimate column in Table 1 shows the values of $\tilde{\text{NP}}^{[1]}_{\text{Geo}}$ for the sixteen portfolios.

We now consider the specification and the validation of the prior distribution with which a Bayesian analysis will be performed. First, we use the standard beta distribution as a structure function, containing the noninformative uniform distribution in the interval $[0, 1]$, and then a uniform distribution in an interval $[\alpha, \beta]$ that is much smaller than $[0, 1]$.

4.1 Structure function Beta($a, b$) in the interval $[0, 1]$

In this case, the density function is given by

$$\text{Be}(\theta_1) = \frac{\theta_1^{a-1}(1 - \theta_1)^{b-1}}{B(a, b)},$$

where $B(a, b) = \Gamma(a) \Gamma(b)/\Gamma(a + b)$ and $\Gamma(\cdot)$ is the gamma function. The values of the mean and the variance are given by

$$E_{\text{Be}}[\theta_1] = \frac{a}{a + b} \quad \text{and} \quad \text{Var}_{\text{Be}}[\theta_1] = \frac{ab}{(a + b)^2(a + b + 1)}.$$  

respectively. This function is unimodal, with $a > 1$ and $b > 1$ and $\text{Mode}_{\text{Be}}[\theta_1] = (a - 1)/(a + b - 2)$.

The marginal distribution is given by

$$m_{\text{Geo}}(k \mid \text{Be}) = \frac{B(a + n; b + nk)}{B(a, b)}.$$  

The mean and the variance of this marginal distribution are given by

$$E_{m_{\text{Geo}}}[K] = \frac{b}{a - 1} \quad \text{and} \quad \text{Var}_{m_{\text{Geo}}}[K] = \frac{ab(a + b - 1)}{(a - 1)^2(a - 2)}.$$  

To subscribe to a Risk Journal visit subscriptions.risk.net/journals or email info@risk.net
The posterior distribution is \( \text{Beta}(a + n, b + n\bar{k}) \). Hence, the beta distribution is a conjugate distribution for the geometric distribution. The mode of the posterior distribution is

\[
\hat{\theta}_1^{\text{Mode}} = \frac{a + n - 1}{a + b + n + n\bar{k} - 2},
\]

which is a very convenient point estimator for the parameter. A second point estimator can be obtained for the true risk premium, given by

\[
\hat{\theta}_1^{\text{Mode}}(\hat{\theta}_1^{\text{Mode}}) = \frac{n\bar{k} + b - 1}{n + a - 1}.
\]

In this case, the predictive function is given by

\[
p_{\text{Geo}}(k_{n+1} \mid k) = \frac{B(a + n + 1; b + n\bar{k} + k_{n+1})}{B(a + n; b + n\bar{k})}.
\]

The collective premium is \( \text{CP}[\text{Be}(a, b)] = b / (a - 1) \) and the Bayes premium is

\[
\text{BP}[\text{Be}(a, b)] = \frac{b + n\bar{k}}{a + n - 1} = \frac{a - 1}{a + n - 1} \frac{b}{a - 1} + \left(1 - \frac{a - 1}{a + n - 1}\right)\bar{k},
\]

which is given by a credibility formula following Jewell (1974). For the uniform distribution \( \text{Unif}[0, 1] = \text{Beta}(1, 1) \), it follows that

\[
\text{BP}[\text{Be}(1, 1)] = \bar{k} + \frac{1}{n} = \text{NP}_{\text{Geo}}^{[1]} + \frac{1}{n}.
\]

Clearly,

\[
\lim_{n \to \infty} \{\text{BP}[\text{Be}(a, b)]\} = \bar{k} = \text{NP}_{\text{Geo}}^{[1]}.
\]

In every case, \( \text{BP}[\text{Be}(1, 1)] > \text{NP}_{\text{Geo}}^{[1]} \). Further, the difference is \( 1/n \), which is very small, in practice, for nonlife insurance portfolios. However, if we compare it with the \( \text{BP}[\text{Be}(a, b)] \) premium, with nonuniform \( \text{Be}(a, b) \), any of the inequalities can occur. It is straightforward to obtain that

\[
\text{BP}[\text{Be}(a, b)] - \text{NP}_{\text{Geo}}^{[1]} = \frac{b - (a - 1)\bar{k}}{a + n - 1},
\]

and this difference is positive when \( \bar{k} < b / (a - 1) \). Otherwise,

\[
\text{BP}[\text{Be}(1, 1)] - \text{BP}[\text{Be}(a, b)] = \frac{(n\bar{k} + 1)(a - 1) - n(b - 1)}{n(n + a - 1)}.
\]

and this difference is positive when

\[
\bar{k} > \frac{b - 1}{a - 1} - \frac{1}{n}.
\]
Consequently, we have three cases, according to the order of the premiums indicated.

1. For the portfolios and elicitations of the hyperparameters verifying

\[
\bar{k} < \frac{b - 1}{a - 1} - \frac{1}{n},
\]

that is,

\[
a < \frac{(b - 1)n}{nk + 1} + 1,
\]

it follows that

\[
\widehat{NP}_{\text{Geo}}^{[1]} < \text{BP}[\text{Be}(1, 1)] < \text{BP}[\text{Be}(a, b)].
\]

2. For the portfolios and elicitations of the hyperparameters verifying

\[
\frac{b - 1}{a - 1} - \frac{1}{n} < \bar{k} < \frac{b}{a - 1},
\]

that is,

\[
\frac{(b - 1)n}{nk + 1} + 1 < a < \frac{b}{k} + 1,
\]

it follows that

\[
\widehat{NP}_{\text{Geo}}^{[1]} < \text{BP}[\text{Be}(a, b)] < \text{BP}[\text{Be}(1, 1)].
\]

3. Finally, for the portfolios and elicitations of the hyperparameters verifying

\[
\bar{k} > \frac{b}{a - 1},
\]

that is, \(a > b/\bar{k} + 1\), it follows that

\[
\text{BP}[\text{Be}(a, b)] < \widehat{NP}_{\text{Geo}}^{[1]} < \text{BP}[\text{Be}(1, 1)].
\]

Let us now consider the elicitation of the hyperparameters. In previous studies, this was often performed by means of the marginal distribution, in seeking to achieve greater objectivity; hence, a moment or maximum likelihood method was used to determine values for the hyperparameters. We analyze both possibilities.

In the moment method, the expected value of the marginal distribution is set equal to the simple mean of the data. The collective premium then coincides with the sample mean. Further, in cases where the Bayes premium is a convex combination of the collective premium and the sample mean, it follows that the Bayes premium is equal to the sample mean. This is the case, for example, with the geometric–exponential or exponential–gamma pairs. For this reason, we must discard this method of hyperparameter elicitation.
Let us again consider the marginal distribution, this time estimating the hyperparameters by means of the maximum likelihood method. In a given sample, the likelihood function is

$$L(a, b) \equiv L(k_0, k_1, \cdots | a, b)$$

$$= \prod_{i \geq 0} m(k_i | a, b)$$

$$= [m(0 | a, b)]^{n_0}[m(1 | a, b)]^{n_1}[m(2 | a, b)]^{n_2}[m(3 | a, b)]^{n_3} \cdots$$

$$= \left[ \frac{a}{a+b} \right]^{n_0} \left[ \frac{ab}{(a+b)(a+b+1)} \right]^{n_1} \times \left[ \frac{ab(b+1)}{(a+b)(a+b+1)(a+b+2)} \right]^{n_2}$$

$$\times \left[ \frac{ab(b+1)(b+2)}{(a+b)(a+b+1)(a+b+2)(a+b+3)} \right]^{n_3} \cdots .$$

This approach obtains expressions given by the quotients of high-order polynomials, and calls for a complicated maximization of these functions or of their logarithms. Table 2 shows the results obtained for the sixteen portfolios considered. For each case, we show the values of the mean, mode and variance of the corresponding prior distribution and the value of the collective premium.

With the exception of the C1 portfolio, the hyperparameters shown in the table present extremely large values, and $a$ is always much larger than $b$. On analyzing the characteristics of these prior distributions, we conclude that all the beta distributions obtained are unimodal. The values for the mode and the mean are very similar, with differences of less than $10^{-3}$, except for the C1 portfolio. Finally, in many portfolios, the values of the variances are so small that they have practically degenerated to one-point distributions. This is perhaps to be expected, because the elicitation performed is of the ML-II type (see Berger 1985) and it is well known that for the family of all distributions the solution is given by the degenerate distribution at the maximum likelihood. In our opinion, this type of elicitation should be excluded because the prior distribution obtained does not model a reasonable grade of prior information, but reiterates (almost pathologically) the observed data.

In our opinion, both the elicitation methods presented above have the drawback of abandoning the prior information and adopting a solution of compromise, close to the data. In order to validate the above results, we now consider more reasonable specifications of the available prior information. Thus, we continue to assume a prior beta distribution, and for the specification of the hyperparameters we assume that an expert has specified the prior information in terms of quantiles.
### TABLE 2  Prior beta distributions for the sixteen portfolios when the hyperparameters are elicited by maximum likelihood in the marginal distribution.

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>(a)</th>
<th>(b)</th>
<th>(E_{Be[\theta_1]})</th>
<th>(\text{Mode}_{Be[\theta_1]})</th>
<th>(\text{Var}_{Be[\theta_1]})</th>
<th>(CP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>22.794410</td>
<td>4.670693</td>
<td>0.82994</td>
<td>0.85585</td>
<td>0.004958</td>
<td>0.214307</td>
</tr>
<tr>
<td>C2</td>
<td>11736.620</td>
<td>1692.604</td>
<td>0.87396</td>
<td>0.87402</td>
<td>0.000008</td>
<td>0.144228</td>
</tr>
<tr>
<td>C3</td>
<td>3591.8823</td>
<td>557.0896</td>
<td>0.86573</td>
<td>0.86590</td>
<td>0.000028</td>
<td>0.155140</td>
</tr>
<tr>
<td>C4</td>
<td>989461.75</td>
<td>130348.9</td>
<td>0.88359</td>
<td>0.88359</td>
<td>0.0000009</td>
<td>0.131737</td>
</tr>
<tr>
<td>C5</td>
<td>653876.19</td>
<td>66094.12</td>
<td>0.90819</td>
<td>0.90819</td>
<td>0.000001</td>
<td>0.101081</td>
</tr>
<tr>
<td>C6</td>
<td>31998.715</td>
<td>5702.673</td>
<td>0.84874</td>
<td>0.84876</td>
<td>0.000003</td>
<td>0.178221</td>
</tr>
<tr>
<td>C7</td>
<td>277658.77</td>
<td>29349.74</td>
<td>0.90440</td>
<td>0.90440</td>
<td>0.000003</td>
<td>0.105705</td>
</tr>
<tr>
<td>C8</td>
<td>778912.25</td>
<td>80716.26</td>
<td>0.90610</td>
<td>0.90610</td>
<td>0.0000009</td>
<td>0.103627</td>
</tr>
<tr>
<td>C9</td>
<td>973487.53</td>
<td>122134.9</td>
<td>0.88852</td>
<td>0.88852</td>
<td>0.0000009</td>
<td>0.125461</td>
</tr>
<tr>
<td>C10</td>
<td>30.590489</td>
<td>6.661685</td>
<td>0.82117</td>
<td>0.83939</td>
<td>0.003839</td>
<td>0.225129</td>
</tr>
<tr>
<td>C11</td>
<td>47497.071</td>
<td>3745.341</td>
<td>0.92691</td>
<td>0.92693</td>
<td>0.000001</td>
<td>0.078856</td>
</tr>
<tr>
<td>C12</td>
<td>66.826617</td>
<td>4.556172</td>
<td>0.93617</td>
<td>0.94874</td>
<td>0.000826</td>
<td>0.069215</td>
</tr>
<tr>
<td>C13</td>
<td>73.134502</td>
<td>7.459448</td>
<td>0.90744</td>
<td>0.91781</td>
<td>0.001029</td>
<td>0.103410</td>
</tr>
<tr>
<td>C14</td>
<td>135.44930</td>
<td>11.96173</td>
<td>0.91885</td>
<td>0.92461</td>
<td>0.000502</td>
<td>0.089968</td>
</tr>
<tr>
<td>C15</td>
<td>117.94143</td>
<td>9.235519</td>
<td>0.92738</td>
<td>0.93421</td>
<td>0.000525</td>
<td>0.078975</td>
</tr>
<tr>
<td>C16</td>
<td>25448.335</td>
<td>1851.377</td>
<td>0.93218</td>
<td>0.93221</td>
<td>0.000002</td>
<td>0.072753</td>
</tr>
</tbody>
</table>

The value for the collective premium, \(CP(\text{Be}(a, b))\) is also shown.
Bayesian analysis in an aggregate loss model

TABLE 3  Prior beta distributions for the sixteen portfolios when the hyperparameters are elicited with the quantiles of order 0.1% and 95% in the columns labeled $P_{0.001}$ and $P_{0.95}$.

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>$P_{0.001}$</th>
<th>$P_{0.95}$</th>
<th>$a$</th>
<th>$b$</th>
<th>$E_{Be}[	heta_1]$</th>
<th>$\text{Var}_{Be}[	heta_1]$</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1, C10</td>
<td>0.7787</td>
<td>0.95</td>
<td>76.685</td>
<td>8.190</td>
<td>0.90350</td>
<td>0.00101</td>
<td>0.10821</td>
</tr>
<tr>
<td>C2, C4</td>
<td>0.8229</td>
<td>0.95</td>
<td>116.813</td>
<td>11.104</td>
<td>0.91319</td>
<td>0.00061</td>
<td>0.09588</td>
</tr>
<tr>
<td>C3</td>
<td>0.8152</td>
<td>0.95</td>
<td>104.757</td>
<td>10.246</td>
<td>0.91090</td>
<td>0.00069</td>
<td>0.09875</td>
</tr>
<tr>
<td>C5, C7, C8, C13</td>
<td>0.8565</td>
<td>0.949</td>
<td>147.717</td>
<td>13.249</td>
<td>0.91768</td>
<td>0.00046</td>
<td>0.09031</td>
</tr>
<tr>
<td>C6</td>
<td>0.7941</td>
<td>0.95</td>
<td>84.515</td>
<td>8.773</td>
<td>0.90595</td>
<td>0.00090</td>
<td>0.10505</td>
</tr>
<tr>
<td>C9</td>
<td>0.8339</td>
<td>0.949</td>
<td>128.553</td>
<td>11.929</td>
<td>0.91508</td>
<td>0.00055</td>
<td>0.09352</td>
</tr>
<tr>
<td>C11, C15</td>
<td>0.8766</td>
<td>0.95</td>
<td>146.198</td>
<td>8.828</td>
<td>0.94305</td>
<td>0.00034</td>
<td>0.06080</td>
</tr>
<tr>
<td>C12, C16</td>
<td>0.8862</td>
<td>0.95</td>
<td>125.888</td>
<td>5.969</td>
<td>0.95472</td>
<td>0.00032</td>
<td>0.04780</td>
</tr>
<tr>
<td>C14</td>
<td>0.8688</td>
<td>0.945</td>
<td>150.075</td>
<td>11.202</td>
<td>0.93054</td>
<td>0.00039</td>
<td>0.07514</td>
</tr>
</tbody>
</table>

The value for the collective premium, $CP(\text{Be}(a, b))$, is also shown.

This approach to the problem has a strong intuitive content and leads to affirmations such as the following.

(i) The $\theta_1$ parameter, which is the probability at zero, never takes values below a certain level specified by the expert, denoted by $P_1$. This can be expressed by indicating that $P_1$ is a quantile of order, say, $\varepsilon_1$, which is very small, e.g., $\varepsilon_1 = 0.001$. This is denoted as $P_{\varepsilon_1}$.

(ii) The $\theta_1$ parameter is very unlikely to take any value higher than $P_2$. This can be expressed by indicating that $P_2$ is a quantile of order, say, $\varepsilon_2$, which is very high, e.g., $\varepsilon_2 = 0.95$. This is denoted as $P_{\varepsilon_2}$.

A beta distribution that satisfies these conditions can be determined by successive approximations.

We first elicit the hyperparameters by assuming specifications (i) and (ii). Again, our analysis refers to the sixteen portfolios described in Table 1. For (i), we take as $P_{\varepsilon_1}$ the sample frequency of approximately zero observed claims in all the portfolios decreased by 5%, and as $P_{\varepsilon_2}$ the sample frequency increased by 5%. Accordingly, the portfolios are divided into groups with similar specifications, and hence the specified beta distribution is the same. By means of straightforward, albeit tedious, calculations we obtain the results in Table 3.

Let us now consider a context of less precise prior information – a situation that is easier for the expert – in which we take just one specification as type (i). To do so, the portfolios are split into three groups, with the same beta distribution in each one. The order of the quantile is 0.1% for all three groups. Portfolios C1, C6 and C10 belong to the first group, and its quantile is $P_{0.001} = 0.8$. Portfolios C2–C5, C7–C9 and C13
TABLE 4  Prior beta distributions for the sixteen portfolios when the hyperparameters are elicited with the quantile of order 0.1% in the column labeled $P_{0.001}$.

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>$P_{0.001}$</th>
<th>$a$</th>
<th>$b$</th>
<th>$E_{Be}[\theta_1]$</th>
<th>$\text{Var}_{Be}[\theta_1]$</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1, C6, C10</td>
<td>0.80</td>
<td>57</td>
<td>4</td>
<td>0.93442</td>
<td>0.00098</td>
<td>0.0714285</td>
</tr>
<tr>
<td>C2, C3, C4, C5, C7, C8, C9, C13</td>
<td>0.85</td>
<td>79</td>
<td>4</td>
<td>0.95181</td>
<td>0.00054</td>
<td>0.0512820</td>
</tr>
<tr>
<td>C11, C12, C14, C15, C16</td>
<td>0.90</td>
<td>122</td>
<td>4</td>
<td>0.96825</td>
<td>0.00024</td>
<td>0.0330578</td>
</tr>
</tbody>
</table>

The value for the collective premium, $CP(\theta(a), b)$, is also shown.

TABLE 5  The values for the Bayes premium, $BP(\theta(a), b)$, in the cases indicated for the sixteen portfolios.

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>BP</th>
<th>$BP[\theta(a, b)]^{[1]}_{\text{Geo}}$</th>
<th>$BP[\theta(1, 1)]$</th>
<th>$BP[\theta(a, b)]^{[2]}_{\text{Geo}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>0.214353</td>
<td>0.999999</td>
<td>0.999506</td>
<td>1.004922</td>
</tr>
<tr>
<td>C2</td>
<td>0.144222</td>
<td>1.000018</td>
<td>0.999725</td>
<td>1.001963</td>
</tr>
<tr>
<td>C3</td>
<td>0.155140</td>
<td>1</td>
<td>0.999946</td>
<td>1.000052</td>
</tr>
<tr>
<td>C4</td>
<td>0.131737</td>
<td>1.000002</td>
<td>0.999982</td>
<td>1.000005</td>
</tr>
<tr>
<td>C5</td>
<td>0.101081</td>
<td>0.999999</td>
<td>0.999907</td>
<td>1.000013</td>
</tr>
<tr>
<td>C6</td>
<td>0.178184</td>
<td>1.000006</td>
<td>1.000001</td>
<td>1.000005</td>
</tr>
<tr>
<td>C7</td>
<td>0.105704</td>
<td>1.000004</td>
<td>0.999851</td>
<td>1.000027</td>
</tr>
<tr>
<td>C8</td>
<td>0.103627</td>
<td>1.000000</td>
<td>0.999926</td>
<td>1.000011</td>
</tr>
<tr>
<td>C9</td>
<td>0.125461</td>
<td>1.000002</td>
<td>0.999987</td>
<td>1.000005</td>
</tr>
<tr>
<td>C10</td>
<td>0.225144</td>
<td>0.999999</td>
<td>0.999970</td>
<td>1.000029</td>
</tr>
<tr>
<td>C11</td>
<td>0.078857</td>
<td>0.999999</td>
<td>0.999994</td>
<td>1.000005</td>
</tr>
<tr>
<td>C12</td>
<td>0.069214</td>
<td>1</td>
<td>0.999973</td>
<td>1.000026</td>
</tr>
<tr>
<td>C13</td>
<td>0.103411</td>
<td>1</td>
<td>0.999979</td>
<td>1.000020</td>
</tr>
<tr>
<td>C14</td>
<td>0.088968</td>
<td>1</td>
<td>0.999972</td>
<td>1.000027</td>
</tr>
<tr>
<td>C15</td>
<td>0.078975</td>
<td>1</td>
<td>0.999968</td>
<td>1.000031</td>
</tr>
<tr>
<td>C16</td>
<td>0.072756</td>
<td>0.999987</td>
<td>0.999784</td>
<td>1.000147</td>
</tr>
</tbody>
</table>

belong to the second group, and its quantile is $P_{0.001} = 0.85$. The third group, with quantile $P_{0.001} = 0.9$ contains portfolios C11, C12 and C14–C16. Table 4 shows a possible (and easier) elicitation obtained when the second hyperparameter is $b = 4$. This small-valued integer was chosen only for simplicity. An integer value is obtained for the first hyperparameter by successive approximations.

Tables 2–4 show three possible elicitations of hyperparameters. The most suitable would seem to be those in Tables 3 and 4, using prior information that would be considered essential for the expert.

Table 5 shows the Bayes premium values for each of the three elicitations of hyperparameter contexts. For validation purposes, the table also presents a comparison
**TABLE 5** Continued.

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>BP</th>
<th>$BP / NP^{[1]}_{Geo}$</th>
<th>$BP / BP[Be(1, 1)]$</th>
<th>$BP / NP^{[2]}_{Geo}$</th>
<th>BP</th>
<th>$BP / NP^{[1]}_{Geo}$</th>
<th>$BP / BP[Be(1, 1)]$</th>
<th>$BP / NP^{[2]}_{Geo}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>0.213511</td>
<td>0.996070</td>
<td>0.995579</td>
<td>1.000491</td>
<td>0.213512</td>
<td>0.996076</td>
<td>0.995585</td>
<td>1.000492</td>
</tr>
<tr>
<td>C2</td>
<td>0.143983</td>
<td>0.998362</td>
<td>0.998069</td>
<td>1.000293</td>
<td>0.143913</td>
<td>0.997876</td>
<td>0.997583</td>
<td>1.000294</td>
</tr>
<tr>
<td>C3</td>
<td>0.155091</td>
<td>0.999685</td>
<td>0.999632</td>
<td>1.000054</td>
<td>0.155072</td>
<td>0.999564</td>
<td>0.999511</td>
<td>1.000054</td>
</tr>
<tr>
<td>C4</td>
<td>0.131727</td>
<td>0.999925</td>
<td>0.999907</td>
<td>1.000018</td>
<td>0.131722</td>
<td>0.999887</td>
<td>0.999869</td>
<td>1.000018</td>
</tr>
<tr>
<td>C5</td>
<td>0.101065</td>
<td>0.999854</td>
<td>0.999761</td>
<td>1.000092</td>
<td>0.101044</td>
<td>0.999641</td>
<td>0.999548</td>
<td>1.000092</td>
</tr>
<tr>
<td>C6</td>
<td>0.178177</td>
<td>0.999967</td>
<td>0.999962</td>
<td>1.000005</td>
<td>0.178177</td>
<td>0.999946</td>
<td>0.999962</td>
<td>1.000005</td>
</tr>
<tr>
<td>C7</td>
<td>0.105669</td>
<td>0.999663</td>
<td>0.999514</td>
<td>1.000149</td>
<td>0.105637</td>
<td>0.999366</td>
<td>0.999217</td>
<td>1.000149</td>
</tr>
<tr>
<td>C8</td>
<td>0.103612</td>
<td>0.999856</td>
<td>0.999783</td>
<td>1.000073</td>
<td>0.103595</td>
<td>0.999699</td>
<td>0.999626</td>
<td>1.000073</td>
</tr>
<tr>
<td>C9</td>
<td>0.125455</td>
<td>0.999949</td>
<td>0.999937</td>
<td>1.000012</td>
<td>0.125452</td>
<td>0.999928</td>
<td>0.999915</td>
<td>1.000012</td>
</tr>
<tr>
<td>C10</td>
<td>0.220851</td>
<td>0.999737</td>
<td>0.999707</td>
<td>1.000029</td>
<td>0.220586</td>
<td>0.999744</td>
<td>0.999714</td>
<td>1.000029</td>
</tr>
<tr>
<td>C11</td>
<td>0.078855</td>
<td>0.999985</td>
<td>0.999981</td>
<td>1.000005</td>
<td>0.078855</td>
<td>0.999970</td>
<td>0.999965</td>
<td>1.000005</td>
</tr>
<tr>
<td>C12</td>
<td>0.069209</td>
<td>0.999929</td>
<td>0.999903</td>
<td>1.000026</td>
<td>0.069206</td>
<td>0.999988</td>
<td>0.999858</td>
<td>1.000026</td>
</tr>
<tr>
<td>C13</td>
<td>0.103407</td>
<td>0.999961</td>
<td>0.999941</td>
<td>1.000020</td>
<td>0.103402</td>
<td>0.999918</td>
<td>0.999897</td>
<td>1.000020</td>
</tr>
<tr>
<td>C14</td>
<td>0.088963</td>
<td>0.999943</td>
<td>0.999916</td>
<td>1.000027</td>
<td>0.088952</td>
<td>0.999815</td>
<td>0.999788</td>
<td>1.000027</td>
</tr>
<tr>
<td>C15</td>
<td>0.078969</td>
<td>0.999916</td>
<td>0.999885</td>
<td>1.000031</td>
<td>0.078962</td>
<td>0.999824</td>
<td>0.999793</td>
<td>1.000031</td>
</tr>
<tr>
<td>C16</td>
<td>0.072711</td>
<td>0.999369</td>
<td>0.999167</td>
<td>1.000202</td>
<td>0.072686</td>
<td>0.999029</td>
<td>0.998826</td>
<td>1.000202</td>
</tr>
</tbody>
</table>
by quotient between the corresponding Bayes premium, $\text{BP}[\text{Be}(a, b)]$, and the estimations $\hat{N}_\text{Geo}^{[1]} = \bar{k}$ and $\hat{N}_\text{Geo}^{[2]} = (n\bar{k} + b - 1)/(n + a - 1)$, respectively, and between the Bayes premium and the noninformative Bayes premium, $\text{BP}[\text{Be}(1, 1)] = \bar{k} + 1/n$.

Table 5 shows the premium obtained by maximum likelihood, the premium obtained with the posterior distribution mode and the Bayes premium obtained with the beta prior distribution, together with the hyperparameters obtained for each of these three approaches. The conclusions are clear and incontrovertible: in all cases the premiums obtained have practically the same value.

### 4.2 Uniform structure function in an interval $[\alpha, \beta] \subseteq [0, 1]$, where the hyperparameters are specified by the expert

We now consider the case in which there is less precise prior information. Specifically, the prior distribution is the same case as in (i) and (ii), or just (i) without the assumption of a beta distribution. The difference is not very significant, in view of the flexibility of the beta distribution family, but nevertheless it is worth studying. Specifically, the prior distribution is

$$\text{Unif}(\theta_1) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } \theta_1 \in [\alpha, \beta], \\ 0 & \text{otherwise.} \end{cases}$$

We assume that the hyperparameters $\alpha$ and $\beta$ are specified by the expert. In practice, these are two extreme-order quantiles, as in the previous section. For the numerical analysis, $(\alpha, \beta)$ is taken as $(P_{0.001}, P_{0.95})$, shown in the second and third columns of Table 3. We then consider the pair $(\alpha, \beta)$ to be $(P_{0.001}, 1)$, where $P_{0.001}$ is as in the second column of Table 4.

The marginal distribution is given by

$$m(\bar{k} \mid \text{Unif}[\alpha, \beta]) = \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} \theta_1^n (1 - \theta_1)^{n\bar{k}} d\theta_1,$$

where $n$ is the total number of policies in the portfolio and $n\bar{k}$ is the total amount of the claims made. As mentioned above, the values for $n$ and $n\bar{k}$ in all the portfolios have an order of magnitude such that, in practice, the integral $\int_{\alpha}^{\beta} \theta_1^n (1 - \theta_1)^{n\bar{k}} d\theta_1$ will have extremely small values. Expressions involving this integer are very difficult to manipulate algebraically, a difficulty that extends to the calculation of the posterior distribution:

$$\pi(\theta_1 \mid k_0, k_1, \ldots) = \frac{\theta_1^n (1 - \theta_1)^{n\bar{k}}}{\int_{\alpha}^{\beta} \theta_1^n (1 - \theta_1)^{n\bar{k}} d\theta_1} I_{[\alpha, \beta]}(\theta_1).$$
Under the net premium principle, the collective premium is obtained directly, by

$\text{CP}(\text{Unif}[\alpha, \beta]) = \int_{\alpha}^{\beta} \frac{1 - \theta_1}{\beta - \alpha} \frac{1}{\theta_1} \, d\theta_1 = \frac{1}{\beta - \alpha} \ln \left( \frac{\beta}{\alpha} \right) - 1.$

Again, calculation of the Bayes premium presents some difficulty:

$\text{BP}(\text{Unif}[\alpha, \beta]) = \frac{\int_{\alpha}^{\beta} \theta_1^n (1 - \theta_1)^{n\theta_1 + 1} \, d\theta_1}{\int_{\alpha}^{\beta} \theta_1^n (1 - \theta_1)^{n\theta_1} \, d\theta_1}.$

For this expression, an approximate calculation of the numerator and denominator, with, for example, Mathematica, for all portfolios, produces a zero value. Accordingly, we obtain an approximate expression for the fraction, using integration by parts in the numerator of the above expression. This produces

$\text{BP}(\text{Unif}[\alpha, \beta]) = \tilde{k} + \frac{1}{n} + \frac{(1 - \beta)g(\beta) - (1 - \alpha)g(\alpha)}{n \int_{\alpha}^{\beta} \theta_1^n (1 - \theta_1)^{n\theta_1} \, d\theta_1},$

where $g(x) = x^n (1 - x)^{n\tilde{k}}$.

The fraction in the latter expression can be approximated as follows: let $g(x) = x^n (1 - x)^{n\tilde{k}}$, $x \in [0, 1]$, be the function that verifies $g(0) = g(1) = 0$. Hence, it has a unique absolute maximum from which it is easy to obtain that $\hat{x} = 1/(1 + \tilde{k})$.

With the interval $[\alpha, \beta] \not\subset [0, 1]$ we can study three cases related to its position with respect to the abscissa of the maximum.

**Case 1** ($\hat{x} = 1/(1 + \tilde{k}) < \alpha$) $g(x)$ decreases in $[\alpha, \beta]$, and has a maximum at $x = \alpha$ and a minimum at $x = \beta$.

**Case 2** ($\hat{x} = 1/(1 + \tilde{k}) > \beta$) $g(x)$ increases in $[\alpha, \beta]$, and has a maximum at $x = \beta$ and a minimum at $x = \alpha$.

The third case is the most reasonable for specific data:

**Case 3** ($\alpha < \hat{x} = 1/(1 + \tilde{k}) < \beta$) $g(x)$ increases from $x = \alpha$ to $\hat{x}$. It has an absolute maximum and then decreases to $x = \beta$. Comparison of $g(\alpha)$ and $g(\beta)$ shows that

$g(\alpha) > g(\beta) \iff \tilde{k} > \frac{\ln(\beta/\alpha)}{\ln((1 - \alpha)/(1 - \beta))}$.

For an interval $[\alpha, \beta] \not\subset [0.5, 1]$, it is always true that

$\frac{\beta}{\alpha} < \frac{1 - \alpha}{1 - \beta} \iff \frac{\ln(\beta/\alpha)}{\ln((1 - \alpha)/(1 - \beta))} < 1.$
and therefore, in this third case

- if
  \[ \tilde{k} > \frac{\ln(\beta/\alpha)}{\ln((1 - \alpha)/(1 - \beta))}, \]
  then \( g(\beta) < g(\alpha) < g(\hat{x}) \) and \( (1 - \beta)g(\beta) - (1 - \alpha)g(\alpha) < 0 \), and

- if
  \[ \tilde{k} < \frac{\ln(\beta/\alpha)}{\ln((1 - \alpha)/(1 - \beta))}, \]
  then \( g(\alpha) < g(\beta) < g(\hat{x}) \) and \( (1 - \beta)g(\beta) - (1 - \alpha)g(\alpha) > 0 \).

Now we approximate the integral in the marginal distribution:

\[
\int_{\alpha}^{\beta} \theta_1^n (1 - \theta_1)^{n\tilde{k}} \, d\theta_1 \approx \frac{1}{2} (\beta - \alpha) [\inf\{g(x) : x \in [\alpha, \beta]\} + \sup\{g(x) : x \in [\alpha, \beta]\}].
\]

Then, considering the third case, it follows that

\[
\tilde{k} > \frac{\ln(\beta/\alpha)}{\ln((1 - \alpha)/(1 - \beta))} \quad \Rightarrow \quad \int_{\alpha}^{\beta} \theta_1^n (1 - \theta_1)^{n\tilde{k}} \, d\theta_1 \approx \frac{1}{2} (\beta - \alpha) [g(\beta) + g(\hat{x})],
\]

\[
\tilde{k} < \frac{\ln(\beta/\alpha)}{\ln((1 - \alpha)/(1 - \beta))} \quad \Rightarrow \quad \int_{\alpha}^{\beta} \theta_1^n (1 - \theta_1)^{n\tilde{k}} \, d\theta_1 \approx \frac{1}{2} (\beta - \alpha) [g(\alpha) + g(\hat{x})].
\]

In the special cases of intervals such as \([\alpha, 1] \subset [0.5, 1]\), the above expressions can be simplified to obtain

\[
\int_{\alpha}^{\beta} \theta_1^n (1 - \theta_1)^{n\tilde{k}} \, d\theta_1 \approx \frac{1}{2} (1 - \alpha) g(\hat{x}).
\]

Finally, the Bayes premium is given by the following: if

\[
\tilde{k} > \frac{\ln(\beta/\alpha)}{\ln((1 - \alpha)/(1 - \beta))},
\]

then

\[
\text{BP(Unif}[\alpha, \beta]) \approx \text{BP(Unif}[0, 1]) + \frac{2}{(\beta - \alpha)n} \frac{(1 - \beta)g(\beta) - (1 - \alpha)g(\alpha)}{[g(\beta) + g(\hat{x})]}.
\]

The final term is negative and so the Bayes premium decreases with respect to the BP(Unif[0, 1]).

Moreover, if

\[
\tilde{k} < \frac{\ln(\beta/\alpha)}{\ln((1 - \alpha)/(1 - \beta))},
\]
then

\[
BP(\text{Unif}[\alpha, \beta]) \approx BP(\text{Unif}[0, 1]) + \frac{2}{(\beta - \alpha)n} \frac{(1 - \beta)g(\beta) - (1 - \alpha)g(\alpha)}{[g(\alpha) + g(\hat{x})]}.
\]

Here, the final term is positive, and so the Bayes premium increases with respect to \(BP(\text{Unif}[0, 1])\).

In the special case when \(\beta = 1\), it follows that

\[
BP(\text{Unif}[\alpha, 1]) \approx BP(\text{Unif}[0, 1]) - \frac{2g(\alpha)}{g(\hat{x})}.
\]

Table 6 shows the values for the collective and Bayes premiums in each of the elicitation contexts considered. Again, in order to validate the results, we also perform a comparison by quotient between the corresponding Bayes premium, \(BP(\text{Unif}[\alpha, \beta])\), and the estimation \(\hat{\lambda}_\text{Exp} = \hat{k}\), and between the Bayes premium and the noninformative prior Bayes premium, \(BP(\text{Unif}[0, 1]) = \hat{k} + 1/n\).

The conclusions drawn are similar to those for Table 5. In all the cases considered, the Bayes premium is practically identical to the one calculated by maximum likelihood.

5 SIZE OF SINGLE CLAIMS

Under the net premium principle, the true risk premium in the exponential model is

\[
\text{NP}_{\text{Exp}}(\theta_2) = \frac{1}{\theta_2}.
\]

In a frequentist approach, the parameter is estimated by maximum likelihood, \(\hat{\theta}_2^{\text{ML}} = 1/\bar{x}\), and an estimator of the true risk premium is

\[
\hat{\lambda}_\text{Exp} = \text{NP}_{\text{Exp}}(\hat{\theta}_2^{\text{ML}}) = \bar{x}.
\]

In this case, the parametric space is not bounded and so is not as direct as a proper noninformative prior distribution. However, in an improper prior distribution given by \(\pi_2(\theta_2) = 1\), the marginal distribution is

\[
m(x_1, \ldots, x_n \mid \pi_2) = \frac{n!}{(n\bar{x})^{n+1}},
\]

where \(\bar{x}\) is the sample mean, and the posterior distribution, which is a proper distribution, is given by

\[
\pi_2(\theta_2 \mid \text{data}) = \frac{(n\bar{x})^{n+1}}{n!} \theta_2^n e^{-n\bar{x}\theta_2}.
\]
TABLE 6  The values of the collective and Bayes premiums in the cases indicated.

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>(α, β) = (P_{0.001}, P_{0.95}) as in Table 3</th>
<th>(α, 1) = (P_{0.001}, 1) as in Table 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CP</td>
<td>BP</td>
</tr>
<tr>
<td>C1</td>
<td>0.160748</td>
<td>0.214459</td>
</tr>
<tr>
<td>C2</td>
<td>0.130034</td>
<td>0.144262</td>
</tr>
<tr>
<td>C3</td>
<td>0.135226</td>
<td>0.155148</td>
</tr>
<tr>
<td>C4</td>
<td>0.130034</td>
<td>0.131739</td>
</tr>
<tr>
<td>C5</td>
<td>0.108697</td>
<td>0.101089</td>
</tr>
<tr>
<td>C6</td>
<td>0.149792</td>
<td>0.178184</td>
</tr>
<tr>
<td>C7</td>
<td>0.108697</td>
<td>0.105720</td>
</tr>
<tr>
<td>C8</td>
<td>0.108697</td>
<td>0.103634</td>
</tr>
<tr>
<td>C9</td>
<td>0.12333</td>
<td>0.125463</td>
</tr>
<tr>
<td>C10</td>
<td>0.160748</td>
<td>0.225151</td>
</tr>
<tr>
<td>C11</td>
<td>0.095520</td>
<td>0.078857</td>
</tr>
<tr>
<td>C12</td>
<td>0.089644</td>
<td>0.069216</td>
</tr>
<tr>
<td>C13</td>
<td>0.108697</td>
<td>0.103413</td>
</tr>
<tr>
<td>C14</td>
<td>0.103307</td>
<td>0.088971</td>
</tr>
<tr>
<td>C15</td>
<td>0.095520</td>
<td>0.078978</td>
</tr>
<tr>
<td>C16</td>
<td>0.089644</td>
<td>0.072772</td>
</tr>
</tbody>
</table>
The Bayes premium, with an appropriate noninformative distribution, is given directly by
\[
BP[\pi_2] = \frac{(n \bar{x})^{n+1}}{n!} \int_0^\infty \theta_2^{n-1} e^{-n \bar{x} \theta_2} = \frac{(n \bar{x})^{n+1}}{n!} \frac{(n - 1)!}{(n \bar{x})^n} = \bar{x}.
\]

Accordingly, the Bayes premium for a noninformative prior distribution coincides with the maximum likelihood estimation for the true risk premium.

Let us now consider a common modeling approach for the structure function, given by the gamma distribution. As in the previous section, we focus on eliciting the hyperparameters, by considering a Gamma\((c, d)\) distribution for the parameter of the secondary distribution. Its density function is given by
\[
\text{Ga}(\theta_2) = \frac{d^c}{\Gamma(c)} \theta_2^{c-1} e^{-d \theta_2} \text{ for } \theta_2 > 0, c, d > 0.
\]
The mean and variance are given by \(E_{\text{Ga}}[\theta_2] = c/d\) and \(V_{\text{Ga}}[\theta_2] = c/d^2\), respectively. When \(c > 1\), the distribution is unimodal and the corresponding prior mode for the parameter \(\theta_2\) is \(\text{Mode}_{\text{Ga}}[\theta_2] = (c - 1)/d\).

The marginal distribution is a Pareto distribution given by
\[
m_{\text{Exp}}(x \mid \text{Ga}) = \frac{cd^c}{(d + x)^{c+1}}.
\]
The marginal distribution function is given by
\[
M_{\text{Exp}}(x \mid \text{Ga}) = 1 - \frac{d^c}{(d + x)^c}.
\]
The first- and second-order moments and the variance are given by
\[
E_{m_{\text{Exp}}}[X] = \frac{d}{c - 1}, \quad E_{m_{\text{Exp}}}[X^2] = \frac{2d^2}{(c - 1)(c - 2)}, \quad \text{Var}_{m_{\text{Exp}}}[X] = \frac{cd^2}{(c - 1)^2(c - 2)},
\]
respectively.

The posterior distribution is \(\text{Ga}(c + n, d + n \bar{x})\). Hence, the gamma distribution is a conjugate distribution for the exponential distribution. The posterior mode is
\[
\hat{\theta}_2^{\text{Mode}} = \frac{c + n - 1}{d + n \bar{x}},
\]
which is a very convenient point estimator for the parameter. Thus, we obtain a second point estimator of the true risk premium:
\[
\hat{\text{NP}}^{[2]}_{\text{Exp}} = \text{NP}_{\text{Exp}}(\hat{\theta}_2^{\text{Mode}}) = \frac{n \bar{x} + d}{n + c - 1}.
\]
The predictive function is given by

\[
p_{\text{Exp}}(x_{n+1} \mid x) = \frac{(c + n)(d + n\bar{x})^{c+n}}{(d + n\bar{x} + x_{n+1})^{c+n}}.
\]

The collective premium is \(\text{CP}[\text{Ga}] = d/(c-1)\) when \(c > 1\). Otherwise, the integral does not converge and, consequently, does not exist.

The Bayes premium is also given by a credibility formula

\[
\text{BP}[\text{Ga}] = \frac{d + n\bar{x}}{c + n - 1} = \frac{c - 1}{c + n - 1} d + (1 - \frac{c - 1}{c + n - 1})\bar{x}.
\]

Unless the \(c\) hyperparameter takes an extremely high value, in practice the term \((c - 1)/(c + n - 1)\) will be very small, and therefore the Bayes premium will be practically equal to the sample mean. The following numerical analysis corroborates this affirmation.

We verify that \(\lim_{n \to \infty} \{\text{BP}[\text{Ga}(c, d)]\} = \bar{x} = \text{NP}_{\text{Exp}}^{[1]}\).

We now consider the elicitation of the hyperparameters. As in the previous section, we reject the use of the marginal distribution with a moment method, and instead use the marginal distribution with a maximum likelihood estimator for the hyperparameters.

From the likelihood

\[
L(c, d) = \prod_{i \geq 1}^n m(x_i \mid c, d) = \prod_{i \geq 1} \frac{cd^c}{(d + x_i)^{c+1}} = \frac{c^n d^{nc}}{\prod_{i \geq 1} (d + x_i)^{c+1}},
\]

we derive the following system:

\[
\frac{n}{c} + n \ln(d) - \sum_{i \geq 1} \ln(d + x_i) = 0,
\]

\[
\frac{nc}{d} - (c + 1) \sum_{i \geq 1} \frac{1}{d + x_i} = 0.
\]

Evidently, the application of this system would be very tedious and so this alternative is rejected. In our opinion, it is also clear that we must take into account the data, which depends on, among other aspects, the unit of measurement adopted. For this reason, we continue to use the marginal distribution but make a more intuitive assumption about the quantiles. A reasonable approximation is to consider that the expert is able to determine the values of the quantiles of orders \(\varepsilon_1\) and \(\varepsilon_2\), denoted by \(P_{\varepsilon_1}\) and \(P_{\varepsilon_2}\), respectively, for the data, i.e., for the marginal distribution. This leads to a system of equations such as

\[
\frac{d^c}{(d + P_{\varepsilon_1})^c} = 1 - \varepsilon_1,
\]

\[
\frac{d^c}{(d + P_{\varepsilon_2})^c} = 1 - \varepsilon_2.
\]
TABLE 7  The value of the Bayes premium in portfolio C16, with $\bar{x} = 137.2701$ and $n = 67856$.

<table>
<thead>
<tr>
<th>$P_{0.92}$</th>
<th>$P_{0.99}$</th>
<th>$c$</th>
<th>$d$</th>
<th>BP</th>
<th>$\text{BP/}Np^{[2]}_{\text{Exp}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3000</td>
<td>0.259725</td>
<td>0.000059</td>
<td>137.2716</td>
<td>1.000000018</td>
</tr>
<tr>
<td>1</td>
<td>3500</td>
<td>0.254818</td>
<td>0.000049</td>
<td>137.2716</td>
<td>0.999999945</td>
</tr>
<tr>
<td>1</td>
<td>4000</td>
<td>0.250716</td>
<td>0.000042</td>
<td>137.2717</td>
<td>1.000000613</td>
</tr>
<tr>
<td>1</td>
<td>4500</td>
<td>0.247205</td>
<td>0.000036</td>
<td>137.2717</td>
<td>1.000000562</td>
</tr>
<tr>
<td>5</td>
<td>3000</td>
<td>0.32509</td>
<td>0.002113</td>
<td>137.2717</td>
<td>1.000000252</td>
</tr>
<tr>
<td>5</td>
<td>3500</td>
<td>0.317436</td>
<td>0.001752</td>
<td>137.2715</td>
<td>1.00000014</td>
</tr>
<tr>
<td>5</td>
<td>4000</td>
<td>0.311092</td>
<td>0.001489</td>
<td>137.2715</td>
<td>1.000000046</td>
</tr>
<tr>
<td>5</td>
<td>4500</td>
<td>0.305704</td>
<td>0.001291</td>
<td>137.2716</td>
<td>1.000000695</td>
</tr>
<tr>
<td>10</td>
<td>3000</td>
<td>0.364635</td>
<td>0.009821</td>
<td>137.2715</td>
<td>1.000000106</td>
</tr>
<tr>
<td>10</td>
<td>3500</td>
<td>0.355027</td>
<td>0.008141</td>
<td>137.2715</td>
<td>1.000000693</td>
</tr>
<tr>
<td>10</td>
<td>4000</td>
<td>0.347107</td>
<td>0.006921</td>
<td>137.2715</td>
<td>1.000000576</td>
</tr>
<tr>
<td>10</td>
<td>4500</td>
<td>0.340409</td>
<td>0.005997</td>
<td>137.2715</td>
<td>1.000000478</td>
</tr>
<tr>
<td>15</td>
<td>3000</td>
<td>0.392591</td>
<td>0.024141</td>
<td>137.2714</td>
<td>1.000000516</td>
</tr>
<tr>
<td>15</td>
<td>3500</td>
<td>0.381469</td>
<td>0.020006</td>
<td>137.2714</td>
<td>1.000000353</td>
</tr>
<tr>
<td>15</td>
<td>4000</td>
<td>0.372334</td>
<td>0.017004</td>
<td>137.2714</td>
<td>1.000000218</td>
</tr>
<tr>
<td>15</td>
<td>4500</td>
<td>0.364634</td>
<td>0.014732</td>
<td>137.2714</td>
<td>1.000000105</td>
</tr>
</tbody>
</table>

Several cases of prior information are shown in the first two columns.

This system can be solved straightforwardly by taking

$$c = \frac{\log(1 - \varepsilon_1)}{\log(d/(d + P_{\varepsilon_1}))},$$

and obtaining $d$ by successive approximations in the equation

$$\log \left( \frac{1 - \varepsilon_2}{1 - \varepsilon_1} \right) \log(d) + \log(1 - \varepsilon_1) \log(d + P_{\varepsilon_2}) - \log(1 - \varepsilon_2) \log(d + P_{\varepsilon_1}) = 0.$$

For real data sets, it is common to find that a large number of claims are zero, and therefore the first quantile is $P_{\varepsilon_1} = 0$. In this case, the first equation of the system has no solution. Accordingly, an approximation must be performed, by taking a small value, insignificant from a practical point of view, for $P_{\varepsilon_1}$. This was the approach taken to obtain the first column in Table 7. For the second column, which contains the quantile $P_{\varepsilon_2}$, we considered a small range of values taken from data in portfolio C16.

The specified prior information is intuitive and simple. It is easy to show that the solutions to the proposed system of equations are obtained with $c < 1$, and hence the collective premium will not exist. If we wish the system to have a solution with $c > 1$, we need

$$d > \max \left\{ \frac{1 - \varepsilon_1}{\varepsilon_1} P_{\varepsilon_1}, \frac{1 - \varepsilon_2}{\varepsilon_2} P_{\varepsilon_2} \right\}.$$
For usual quantile values, this condition is, for example, \( d > 30 \), and with these values there is no solution to the system. Accordingly, the solutions \((c, d)\) obtained for the system and shown in Table 7 are those in which the collective premiums do not exist. Finally, the table also contrasts the Bayes premium \(BP_{\text{GeExp}}\) with the estimation \(\hat{\text{NP}}_{\text{Exp}[1]} = \bar{x}\). The comparison with the estimation \(\hat{\text{NP}}_{\text{Exp}[2]}\) is direct.

On comparing the last two columns in Table 7 with the mean \(\bar{x} = 137.2701\), we reach the same conclusion as in the previous section, namely that the Bayes premium is virtually the same whether it is estimated by maximum likelihood, with the posterior mode, with improper noninformative prior distributions or with hyperparameters elicited by considering quantiles of the marginal distribution.

### 6 AGGREGATE LOSS

In the CRM, under the net premium principle, the true risk premium is

\[
\text{NP}_{\text{GeExp}}(\theta_1, \theta_2) = \frac{1 - \theta_1}{\theta_1 \theta_2}.
\]

Here it is reasonable to adopt a frequentist approach, similar to that used in the previous sections. Thus, we estimate the parameters \((\theta_1, \theta_2)\) by maximum likelihood:

\[
(\hat{\theta}_1, \hat{\theta}_2)_{\text{ML}} = \left( \frac{n_0}{n}, \frac{n - n_0}{n0\bar{s}} \right).
\]

Then it follows that

\[
\hat{\text{NP}}_{\text{GeExp}[1]} = \text{NP}_{\text{GeExp}}((\hat{\theta}_1, \hat{\theta}_2)_{\text{ML}}) = \bar{s}.
\]

When the risk profiles, \(\theta_1\) and \(\theta_2\), are independent, the bivariate prior distribution \(\text{BeGa}(\theta_1, \theta_2)\) is the product of the marginal distributions. The following results are then very useful. Let \(\varepsilon_1, \varepsilon_2\) and \(\varepsilon_3\) be real numbers, and let \(d\) and \(s\) be real and positive numbers. Then we obtain the following equality:

\[
\text{Int} \equiv \text{Int}(\varepsilon_1, \varepsilon_2, \varepsilon_3, s) = \frac{1}{\theta_1 = 0, \theta_2 = 0} \int_{\theta_1 = 0}^{1} \theta_1^{\varepsilon_1} (1 - \theta_1)^{\varepsilon_2} \theta_2^{\varepsilon_3} e^{-d\theta_2} e^{-\theta_1 \theta_2 s} \ d\theta_1 \ d\theta_2
\]

\[
= \frac{B(\varepsilon_1 + 1, \varepsilon_2 + 1) \Gamma(\varepsilon_3 + 1)}{d^{\varepsilon_3 + 1}} 2F_1 \left( \varepsilon_1 + 1, \varepsilon_3 + 1; \varepsilon_1 + \varepsilon_2 + 2; -\frac{s}{d} \right).
\]

Indeed,

\[
\text{Int} = \int_{\theta_2 = 0}^{\infty} \theta_2^{\varepsilon_3} e^{-d\theta_2} d\theta_2 \int_{\theta_1 = 0}^{1} \theta_1^{\varepsilon_1} (1 - \theta_1)^{\varepsilon_2} e^{-\theta_1 \theta_2 s} d\theta_1.
\]
Using an integral representation of the function $1_{F_1}$, it follows that

$$\text{Int} = B(\varepsilon_1 + 1, \varepsilon_2 + 1) \int_{\theta_2=0}^{\infty} \theta_2^\varepsilon_3 e^{-\theta_2^t} 1_{F_1}(\varepsilon_1 + 1, \varepsilon_2 + 2; -s \theta_2) \, d\theta_2,$$

and by setting $\theta_2 d = t$ we get

$$\text{Int} = B(\varepsilon_1 + 1, \varepsilon_2 + 1) \frac{1}{d^{\varepsilon_3+1}} \int_0^{\infty} t^{\varepsilon_3} e^{-t} 1_{F_1}(\varepsilon_1 + 1, \varepsilon_2 + 2; -\frac{st}{d}) \, dt.$$

By using an integral representation of the function $2_{F_1}$, it follows that

$$\text{Int} = \frac{B(\varepsilon_1 + 1, \varepsilon_2 + 1) \Gamma(\varepsilon_3 + 1)}{d^{\varepsilon_3+1}} 2_{F_1}(\varepsilon_1 + 1, \varepsilon_3 + 1; \varepsilon_1 + \varepsilon_2 + 2; -\frac{s}{d}).$$

The marginal distribution is given by

$$m_{\text{CRM}}(s \mid \text{BeGa}) = C_1 F_{2,1}(a + n, c + n - n_0; a + b + 2n - n_0; -\frac{n \tilde{s}}{d}),$$

where

$$C_1 = \frac{B(a + n, b + n - n_0) \Gamma(c + n - n_0)}{d^{n-n_0} B(a, b) \Gamma(c)}.$$

This is obtained straightforwardly by observing that the marginal adopts the form

$$\frac{d^c}{B(a, b) \Gamma(c)} \text{Int}(a + n - 1, b + n - n_0 - 1, c + n - n_0 - 1, n \tilde{s})$$

and using the previous expression.

We define the posterior distribution as follows:

$$\text{BeGa}(\theta_1, \theta_2 | s) = \frac{L_{\text{CRM}}(s | \theta_1, \theta_2) \text{BeGa}(\theta_1, \theta_2)}{m_{\text{CRM}}(s | \text{BeGa})} = \frac{d^{c}}{C_1 B(a, b) \Gamma(c)} \frac{\theta_1^{n+1-1}(1 - \theta_1)^{n-n_0-b-1} \theta_2^{n-n_0+c-1} e^{-n \tilde{s} \theta_1 \theta_2 - d \theta_2}}{F_{2,1}(a + n, c + n - n_0; a + b + 2n - n_0; -n \tilde{s}/d)}.$$

It is then easy to obtain that the posterior means are given by

$$E_{\text{BeGa}(\theta_1, \theta_2 | s)}[\theta_1] = C_2 \frac{F_{2,1}(a + n + 1, c + n - n_0; a + b + 1 + 2n - n_0; -n \tilde{s}/d)}{F_{2,1}(a + n, c + n - n_0; a + b + 2n - n_0; -n \tilde{s}/d)}.$$

---

1 See http://functions.wolfram.com/07.20.07.0001.01.
2 See http://functions.wolfram.com/07.23.07.0003.01.
and

\[ E_{\text{BeGa}}(\theta_1, \theta_2 | s)[\theta_2] = C_3 \frac{F_{2,1}(a + n, c + n - n_0 + 1; a + b + 2n - n_0: -n \tilde{s}/d)}{F_{2,1}(a + n, c + n - n_0; a + b + 2n - n_0: -n \tilde{s}/d)}, \]

where

\[ C_2 = \frac{a + n}{a + b + 2n - n_0} \quad \text{and} \quad C_3 = \frac{c + n - n_0}{d}. \]

The mode of the posterior distribution, which is an adequate point estimator, is obtained by direct calculation, and given by \((\hat{\theta}_1, \hat{\theta}_2)^\text{Mode} = (\hat{\theta}_1, \hat{\theta}_2)\), where

\[ \hat{\theta}_2 = \frac{c + n - n_0 - 1}{d + n \tilde{s} \hat{\theta}_1} \quad \text{and} \quad \hat{\theta}_1 = \frac{-B + \sqrt{B^2 - 4AC}}{2A}, \]

with \(A = (a + b - c - 1 + n)n \tilde{s}\), \(B = -(a - c + n_0)n \tilde{s} + d(a + b + 2n - n_0 - 2)\) and \(C = -d(a + n - 1)\).

Using \((\hat{\theta}_1, \hat{\theta}_2)^\text{Mode}\) as a point estimator of the parameters \((\theta_1, \theta_2)\), we derive a second point estimator for the true risk premium:

\[ \hat{\text{NP}}_{\text{GeExp}}^{[2]} = \text{NP}_{\text{GeExp}}((\hat{\theta}_1, \hat{\theta}_2)^\text{Mode}). \]

The predictive density function is

\[
p_{\text{CRM}}(s_{n+1} | s) = C_4 \frac{F_{2,1}(a + n + 1, c + n - n_0 + 1; a + b + 2n - n_0 + 2: -(n \tilde{s} + s_{n+1})/d)}{F_{2,1}(a + n, c + n - n_0; a + b + 2n - n_0: -n \tilde{s}/d)},
\]

where

\[ C_4 = \frac{(a + n)(b + n - n_0)(c + n - n_0)}{d(a + b + 2n - n_0)(a + b + 2n - n_0 + 1)}. \]

This is obtained directly by observing that it includes the function \(\text{Int}(a + n, b + n - n_0, c + n - n_0, n \tilde{s} + s_{n+1})\).

From the independence assumption between the risk profiles, it follows that the collective premium is \(\text{CP[BeGa]} = \text{CP[Be]} \cdot \text{CP[Ga]} = bd/((a - 1)(c - 1))\), which does not exist for values of \(c < 1\).

The Bayes premium is given by

\[ \text{BP[BeGa]} = C_5 \frac{F_{2,1}(a + n - 1, c + n - n_0 - 1; a + b + 2n - n_0: -n \tilde{s}/d)}{F_{2,1}(a + n, c + n - n_0; a + b + 2n - n_0: -n \tilde{s}/d)}, \]

where

\[ C_5 = \frac{(b + n - n_0)d}{(a + n - 1)(c + n - n_0 - 1)}. \]
TABLE 8 The value of the Bayes premium in the collective risk model, in portfolio C16, with \( \tilde{s} = 137.2701, n = 67856 \) and \( n_0 = 63232 \).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th>( \hat{NP}_{GeExp}^{[0]} )</th>
<th>( \hat{BP}_{BeGa} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>125.888</td>
<td>5.969</td>
<td>0.259725</td>
<td>0.000042</td>
<td>137.17567</td>
<td>137.20386</td>
</tr>
<tr>
<td>125.888</td>
<td>5.969</td>
<td>0.355027</td>
<td>0.008141</td>
<td>137.17303</td>
<td>137.20065</td>
</tr>
<tr>
<td>122</td>
<td>4</td>
<td>0.259725</td>
<td>0.000042</td>
<td>137.12913</td>
<td>137.15675</td>
</tr>
<tr>
<td>122</td>
<td>4</td>
<td>0.355027</td>
<td>0.008141</td>
<td>137.12650</td>
<td>137.15412</td>
</tr>
</tbody>
</table>

The first two columns show the results of the hyperparameters.

We then take into account that

\[
E_{BeGa} \left[ \frac{(1 - \theta_1) L_{CRM}(s | \theta_1, \theta_2)}{\theta_1 \theta_2} \right] = \frac{d^c}{B(a, b) \Gamma(c)} \text{Int}(a + n - 2, b + n - n_0, c + n - n_0 - 2, n\tilde{s}).
\]

Table 8 shows the results obtained for portfolio C16. We consider the values obtained in Sections 5 and 6 as hyperparameters. Table 8 also shows the values of the \( \hat{NP}_{GeExp}^{[2]} \) estimation and the Bayes premium \( \hat{BP}_{BeGa} \). As we have already established,

\[
\hat{NP}_{GeExp}^{[1]} = \tilde{s}.
\]

We draw similar conclusions to those from Tables 5 to 7.

7 CONCLUSIONS

In this study, for a typical nonlife insurance model, we developed a collective risk model with primary distribution given by the geometric distribution and with an exponential distribution as the secondary distribution. We considered three risk variables (number of claims, size of claims and aggregate loss) and obtained the net premiums, using the classical approach and from the Bayesian standpoint. In all cases, we obtained a frequentist estimation of the premium, an estimation using the posterior mode and the determination of the Bayes premium. These results were then compared in order to validate the models. We studied several cases of prior information applied to a real data set of sixteen portfolios. The comparison between frequentist and Bayesian models showed that in all cases the Bayes premium is practically equal to the net premium when the parameter is estimated by maximum likelihood. Accordingly, we consider that in nonlife insurance problems, with large sample sizes, the data, modeled by the likelihood, is more important than the prior information.
DECLARATION OF INTEREST

This work was supported by Ministerio de Economía y Competitividad under grant ECO2013-47092. The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

ACKNOWLEDGEMENTS

The authors thank the anonymous referee for their detailed and helpful comments.

REFERENCES


Research Paper

On the correlation and parametric approaches to calculation of credit value adjustment

Tao Pang,¹ Wei Chen² and Le Li²

¹Department of Mathematics, Box 8205, North Carolina State University, Raleigh, NC 27695-8205, USA; email: tpang@ncsu.edu
²SAS Institute Inc., 100 SAS Campus Drive, Cary, NC 27513, USA; emails: wei.chen@sas.com, le.li@sas.com

(Received October 31, 2016; revised February 22, 2017; accepted March 2, 2017)

ABSTRACT

Credit value adjustment (CVA) is an adjustment added to the fair value of an over-the-counter trade due to the risk of counterparty defaults. When the exposure to the counterparty and the counterparty default risk tend to change in the same direction, so-called wrong-way risk (WWR) must be taken into account. Right-way risk takes place when the two factors move in opposite directions. These two comovement effects are also called directional-way risk (DWR). Many efforts have been made to reduce the computational burden of calculating CVA with DWR. The two most popular approaches are the parametric approach and the correlation approach. In this paper, we develop a connection between these two approaches. In particular, by decomposing the DWR into a robust correlation coefficient and a profile multiplier, we bring the parametric approach into the correlation approach framework. This allows us to explain the parameters in the parametric approach. Our results suggest that the parametric approach can become sensitive when calculating the WWR in certain scenarios. For risk model governance and validation purposes, caution should be exercised when using the parametric approach for CVA calculation.

Keywords: credit value adjustment (CVA); wrong-way risk (WWR); right-way risk; correlation approach; parametric approach.
1 INTRODUCTION

Over-the-counter derivative dealers face the risk of defaults by their counterparties. When a dealer has positive exposure to the counterparty, a counterparty default can induce a significant loss. Credit value adjustment (CVA) is thus introduced to capture the actual fair value to the dealer. It is a value added to the market value of the derivative itself. That means that the fair value of an over-the-counter derivative should depend on the credit quality of the counterparty as well as the value of the derivative. After the financial crisis in 2008, use of CVA gained increasing attention for both fair valuation and capital requirement purposes. CVA is usually quoted as a part of over-the-counter derivatives. Basel III (Basel Committee on Banking Supervision 2011) has also enhanced its requirement on the CVA capital charge.

There are two main risk sources in CVA. One is the collateralized and netted exposure at default to the counterparty. The other is the probability of default (PD) of the counterparty. For simplicity, in the past, over-the-counter derivatives used to be priced with CVA by assuming the exposure and the counterparty credit risk were independent. However, in reality, these two risk sources often move at the same time. A wrong-way risk (WWR) intensifies counterparty risk when the dealer’s exposure to the counterparty tends to increase, while the counterparty credit quality deteriorates. A famous example of WWR is credit default swap (CDS) contracts during the recent financial crisis, where the credit protection buyers faced a significant credit risk from major CDS sellers such as AIG, while the underlying protection got more valuable. On the other hand, right-way risk (RWR), when the two factors move in opposite directions (ie, when the counterparty credit risk intensifies, the exposure to the counterparty goes down or vice versa), is also plausible. RWR essentially builds in a hedge between the underlying derivative risk and the counterparty risk. Because of this hedging effect, RWR receives less attention from the regulators. For example, Basel III does not offer any incentive to reward RWR as much as it punishes WWR. The two comovement effects WWR and RWR are often called directional-way risk (DWR).

According to Gregory (2012, p. 380), there are two main approaches to modeling CVA with DWR: the correlation approach and the parametric approach. The former attempts to capture the underlying risk through the correlation between the exposure and the counterparty credit risk. In most cases, the solution to this approach requires simulations. Pykhtin and Rosen (2010) consider the correlation structure of portfolio exposures and the idiosyncratic default risk driver and derive analytical results based on the joint normal distribution. Rosen and Saunders (2012) use similar assumptions about portfolio exposures and the idiosyncratic default risk driver and propose an algorithm to incorporate different default scenarios with pre-generated exposures to model WWR. Brigo and Alfonsi (2005) and Brigo and Pallavicini (2008) examine the
relationship between counterparty default risk and market-observed CDSs and then price counterparty risk based on the stochastic process of the CDS. Skoglund et al (2013) and Pang et al (2015) use joint stochastic processes to model the risk drivers of the underlying and the CDS. Pang et al propose a robust correlation method to model the correlation of the underlying processes for exposures and PDs.

The parametric approach, on the other hand, can be viewed as a reduced-form approach, where the model does not necessarily explain the underlying correlation but only aims to capture the comovement itself. Examples of this approach include Hull and White (2012), Ghamami and Goldberg (2014) and Ruiz et al (2015). Hull and White assume the hazard rate of the default follows an exponential function of the exposure, and the WWR is modeled by a single parameter. Ghamami and Goldberg assume the exposure is a function of a uniform random variable. Using the Hull–White model, they find that independent CVA can exceed the wrong-way CVA under some circumstances. Ruiz et al carry out calibrations for some parametric models with different forms of function.

In this paper, we first establish a connection between the parametric and correlation approaches. This allows us to discover economic interpretations of the parameters in the latter approach. We apply the DWR decomposition approach (referred to here as the Pang–Chen–Li (PCL) approach) introduced by Pang et al (2015) to the parametric approach proposed in Hull and White (2012). As a result, we derive a decomposition and an analytical and economic anatomy of the parameter $b$ (see (3.1)) in the Hull–White approach. In addition, our results provide an explanation to the finding of the effect of the underlying price volatility on the CVA value observed by Ruiz et al (2015). The parameter $b$ in the Hull–White approach blends both the robust correlation and the profile multiplier in the PCL approach. The CVA value can be quite sensitive when $b$ is small. This is a very important consideration for model risk in real-world implementations.

The paper is organized as follows. In Section 2, we briefly review the CVA and the DWR. In Section 3, we introduce the parametric approach proposed by Hull and White (2012). In Section 4, we briefly review the PCL approach and DWR decomposition proposed by Pang et al (2015). The key results of the analysis are stated and discussed in Section 5. In Section 6, the major findings are discussed. A numerical study based on interest rate swaps is provided in Section 7, with further discussions. We conclude in Section 8.

### 2 THE CREDIT VALUE ADJUSTMENT CALCULATION

CVA is an adjustment added to the fair value of an over-the-counter trade to take into account counterparty credit risk. For simplicity, throughout the paper we consider
only unilateral CVA. A general expression for CVA is

\[ \text{CVA} = (1 - R)\mathbb{E}[\mathbf{1}_{\{\tau \leq T\}} V^+(\tau)], \]

where \( R \) is the constant recovery rate, \( \tau \) is the time to default, \( T \) is the time to expiration, \( \mathbf{1}_{\{\tau \leq T\}} \) is the default indicator function, which is 1 if the default time \( \tau \leq T \) and 0 otherwise, and \( V^+(t) \) is the risk-neutral discounted exposure at time \( t \) subject to netting and collateral.

Equation (2.1) can also be expressed as

\[ \text{CVA} = (1 - R)\left[ \int_0^T \mathbb{E}[V^+(t)] dF(t) \right], \]

where \( F(t) \) is the cumulative distribution function (CDF) of the counterparty’s time to default \( \tau \).

If the PD \( F(t) = \Pr(\tau \leq t) \) is independent of the credit exposure \( V(t) \), the CVA calculation is relatively easy. However, in reality, \( V(t) \) and the counterparty PD often depend on each other. When the credit exposure and the PD tend to move in the same direction, the loss due to the counterparty default will be amplified. In this case, a WWR is presented. On the other hand, if the credit exposure and the PD tend to move in opposite directions, the loss due to the counterparty default is mitigated. In this case, a RWR is presented.

The presence of the DWR makes the CVA calculation more challenging, as the dependence structure between the credit exposure and the counterparty default risk is not easy to characterize quantitatively. Next, we shall introduce our two approaches: the Hull–White parametric approach and the PCL correlation approach.

3 THE HULL–WHITE PARAMETRIC APPROACH

Hull and White (2012) proposed a parametric approach to model the dependence of the PD on the credit exposure. This approach permits the evolution of the conditional PD to depend explicitly on the counterparty credit exposure process \( V(t) \) at time \( t \). In particular, they assume that the hazard rate \( h(t) \) at the counterparty default time is an exponential function of the credit exposure \( V(t) \):

\[ h(t) = \exp(a(t) + bV(t)), \]

where \( a(t) \) is a function of time \( t \) and \( b \) is a constant that describes the dependence of the PD on the credit exposure. Here we outline the approach. More details can be found in Hull and White (2012).

Conditional on there being no earlier default by time \( t \), the probability of default in any small time increment \( \Delta t \) is thus \( h(t)\Delta t \). At time 0, the probability of default
by time $t > 0$ is thus

$$F(t) = 1 - \exp\left(-\int_0^t h(u) \, du\right).$$  \hfill (3.2)

The corresponding discrete form is

$$F(t) = 1 - \exp\left(- \sum_{t_j \leq t} h(t_j) \Delta t\right).$$  \hfill (3.3)

Hence, the PD in terms of hazard rates is

$$q(t_i) = \exp\left(- \sum_{k=1}^{i-1} h(t_k) \Delta t\right) - \exp\left(- \sum_{k=1}^i h(t_k) \Delta t\right).$$  \hfill (3.4)

Hazard rates are not directly observable from the market. But the counterparty credit spread $s(t)$ is observable. The connection between the hazard rate function $h(t)$ and the credit spread $s(t)$ is given by

$$\exp\left(-\int_0^t h(u) \, du\right) = \exp\left(-s(t) t \frac{1}{1 - R}\right),$$  \hfill (3.5)

where $s(t)$ is the counterparty credit spread with maturity $t$, assuming constant recovery rate $R$. For a stochastic hazard rate, the following relationship must be satisfied:

$$\mathbb{E}\left[ \exp\left(-\int_0^t h(u) \, du\right) \right] = \exp\left(-s(t) t \frac{1}{1 - R}\right).$$  \hfill (3.6)

The corresponding discrete version is

$$\mathbb{E}\left[ \exp\left(- \sum_{i=1}^j h_i \Delta t\right) \right] = \exp\left(-s_j t_j \frac{1}{1 - R}\right),$$  \hfill (3.7)

where $h_i$ and $s_j$ denote $h(t_i)$ and $s(t_j)$, respectively, and $t_i, i = 0, 1, 2, \ldots$, are the discrete times corresponding to the different maturities of the counterparty credit spread.

Based on (3.1), the discrete version of the Hull–White model is

$$h_i = \exp(a(t_i) + bV(t_i)),$$  \hfill (3.8)

where $b$ is a constant that measures the amount of DWR and $a(t_i)$ is a function of $t_i$. The values of $a(t_i)$ and $b$ can be calibrated with market data using (3.7) and (3.8).

Intuitively, the larger the value of $b$, the greater the dependency between the portfolio value and the counterparty PD. However, whether the dependence comes from the correlation, the exposure profile or both is not clear. This is the problem we try to solve in this paper.
4 THE PANG–CHEN–LI CORRELATION APPROACH

Pang et al. (2015) proposed a correlation approach. They used a semi-parametric method to decompose the CVA dependence into a robust correlation coefficient and a profile multiplier. It turns out that the multiplier only depends on the first two moments of the credit exposures and the PDs, so the dependence is reflected in the robust correlation coefficient. Next we will briefly introduce the approach.

By discretization of (2.2) into \( t_0 \leq t_1 \leq \ldots \leq t_K = T \), and defining

\[ q(t_i) = F(t_i) - F(t_{i-1}) \]  

as the probability of default at \( (t_{i-1}, t_i] \), we can rewrite (2.2) as

\[ \text{CVA} = (1 - R) \sum_{i=1}^{K} \mathbb{E}[V^+(t_i)q(t_i)]. \]  

When the underlying exposure and counterparty credit risk are independent, (4.2) becomes

\[ \text{CVA}_{\text{ind}} = (1 - R) \sum_{i=1}^{K} \mathbb{E}[V^+(t_i)]\mathbb{E}[q(t_i)]. \]  

Pang et al. (2015) proposed a CVA DWR multiplier decomposition using a simple correlation between the exposure and the counterparty risk. Moreover, they showed that the DWR CVA can be expressed as a factor adjustment to the CVA under the following independence assumption:

\[ \text{CVA}_{\text{DWR}} = (1 + \tilde{\rho}C_p)\text{CVA}_{\text{ind}}, \]  

where

\[ \tilde{\rho} = \frac{\sum_{j=1}^{K} \rho(t_j) \sigma_V(t_j) \sigma_{PD}(t_j)}{\sum_{j=1}^{K} \sigma_V(t_j) \sigma_{PD}(t_j)} \]  

is called the robust correlation coefficient and

\[ C_p = \frac{\sum_{j=1}^{K} \sigma_V(t_j) \sigma_{PD}(t_j)}{\sum_{j=1}^{K} \mu_V(t_j) \mu_{PD}(t_j)} \]  

is called the profile multiplier. In the two equations above, \( \mu_V(t_j), \mu_{PD}(t_j), \sigma_V(t_j) \) and \( \sigma_{PD}(t_j) \) are the means and standard deviations of the underlying exposure and the counterparty’s default probability at time \( t_j \), respectively; \( \rho(t_j) \) is the correlation between the exposure and the counterparty’s PD at time \( t_j \). It is easy to see that \(-1 \leq \tilde{\rho} \leq 1\). \( C_p \) describes a quasi-coefficient of variation of the defaultable value, that is, the product of default-free value and PD.
Based on (4.4), we can define the CVA ratio as

\[
\text{CVA}_{\text{ratio}} = \frac{\text{CVA}_{\text{DWR}}}{\text{CVA}_{\text{ind}}} = 1 + \hat{\rho}C_p. \tag{4.7}
\]

If there is no DWR, the ratio is 1. For WWR the ratio will be larger than 1, and for RWR the ratio is less than 1. In Basel III (Basel Committee on Banking Supervision 2011), the suggested value for this ratio is 1.4 for WWR.

Under the PCL approach framework, the profile multiplier \(C_p\) does not change with the correlation level and can be derived from a case of independent exposure and default. The robust correlation coefficient \(\hat{\rho}\) can be calibrated as a function of the underlying correlation that correlates the stochastic processes driving exposure and default. Pang \textit{et al} (2015) also discussed the stability and reusability of the two measures \(\hat{\rho}\) and \(C_p\) at certain confidence levels, giving a series of numerical examples. They proposed an efficient algorithm to compute CVA with DWR based on these two measures. A major advantage of the PCL approach is that a risk manager can infer the confidence interval of CVA price estimate based on the confidence interval of the underlying correlation without adding significant computational burden.

5 ANALYSIS

We first derive some analytic results. The analysis in this section is based on the following assumptions and notation:

(1) the market is observable daily and there are \(K\) time periods;

(2) \(\Delta t = \frac{1}{252}\) and \(t_j = j\Delta t\) for \(j = 1, 2, \ldots, K\);

(3) the profit and loss at \(t_j\) is denoted by \(X_j\);

(4) \(X_j\) follows normal distributions with mean \(\mu_j\) and variance \(\sigma_j^2\), and \(X_i\) and \(X_j\) are independent for \(i \neq j\);

(5) the initial value of the portfolio is a positive constant, \(V_0\);

(6) the observed credit spread with maturity \(t_j\) is \(s_j\);

(7) the recovery rate and the discount rate are zero;

(8) the exposure at time \(t_j\) is \(V_j\), and \(V_j = V_0 + \sum_{i=1}^{j} X_i\).

We denote the hazard rate at time \(t_j\) by \(h_j\). Following Hull and White (2012), \(h_j\) is a function of \(V_j\), and

\[
h_j = g(V_j) = \exp(a_j + bV_j). \tag{5.1}
\]
where $b$ is a predetermined constant and $a_i$ is time dependent and should be calibrated with the market data. The market-implied probability of default between time $t_{j-1}$ and $t_j$, denoted by $C_j$, is

$$C_j = \exp(-s_{j-1}t_{j-1}) - \exp(-s_j t_j).$$

(5.2)

We define the means and variances of the exposures and PDs as follows:

$$\mathbb{E}[V_j] = \mu_V(t_j), \quad \text{var}(V_j) = \sigma^2_V(t_j),$$

$$\mathbb{E}[h_j \Delta t] = \mu_{PD}(t_j), \quad \text{var}(h_j \Delta t) = \sigma^2_{PD}(t_j).$$

Therefore, we get

$$\mu_V(t_j) = V_0 + \sum_{i=1}^j \mu_i, \quad \sigma^2_V(t_j) = \sum_{i=1}^j \sigma^2_i,$$

(5.3)

$$\mu_{PD}(t_j) = C_j, \quad \sigma^2_{PD}(t_j) = C_j^2 \left[ \exp \left( b^2 \sum_{i=1}^j \sigma^2_i \right) - 1 \right].$$

(5.4)

where the derivations of $\mu_{PD}(t_j)$ and $\sigma^2_{PD}(t_j)$ can be found in Appendix A online. For each time node, if we consider the correlation coefficient of $V_j$ and the probability of default at $t_j$ directly, namely $\rho(t_j)$ in the presence of parameter $b$, we have

$$\rho(t_j) = \frac{b \sqrt{\sum_{i=1}^j \sigma^2_i}}{\sqrt{\exp \left( b^2 \sum_{i=1}^j \sigma^2_i \right) - 1}}.$$

(5.5)

The derivation of $\rho(t_j)$ is also given in Appendix A online.

By virtue of (5.3) and (5.4), we obtain the robust correlation coefficient $\tilde{\rho}$ given by (4.5) and the profile multiplier $C_p$ given by (4.6) as follows:

$$\tilde{\rho} = \frac{b \sum^K_{j=1} (C_j \sum_{i=1}^j \sigma^2_i)}{\sum^K_{j=1} \left[ C_j \sqrt{\sum_{i=1}^j \sigma^2_i} \sqrt{\exp \left( b^2 \sum_{i=1}^j \sigma^2_i \right) - 1} \right]},$$

(5.6)

$$C_p = \frac{\sum^K_{j=1} \left[ C_j \sqrt{\sum_{i=1}^j \sigma^2_i} \sqrt{\exp \left( b^2 \sum_{i=1}^j \sigma^2_i \right) - 1} \right]}{\sum^K_{j=1} [C_j (\sum_{i=1}^j \mu_i + V_0)]}.$$

(5.7)

Then, from (4.7), we get

$$\text{CVA}_{\text{ratio}} = 1 + \tilde{\rho} C_p = 1 + \frac{b \sum^K_{j=1} (C_j \sum_{i=1}^j \sigma^2_i)}{\sum^K_{j=1} [C_j (\sum_{i=1}^j \mu_i + V_0)]}.$$

(5.8)
From (5.7), it is easy to see that the profile multiplier $C_p$ increases as $b$ increases in magnitude. Next, we consider the sensitivity of the robust correlation coefficient $\tilde{\rho}$ to the parameter $b$.

First, we claim the magnitude of $\rho(t_j)$ decreases as $b$ increases in magnitude. It is sufficient to show this when $b > 0$ because $\rho(t_j)$ is symmetric with respect to the origin. Let

$$x = b \sqrt{\sum_{j=1}^{K} \sigma_j^2}. \quad (5.9)$$

The square of (5.5) becomes

$$\rho^2(t_j) = \frac{x^2}{\exp(x^2) - 1}. \quad (5.10)$$

Taking the first-order derivative of the above equation with respect to $x$, we get

$$\frac{d(\rho^2(t_j))}{dx} = -\frac{2x[(x^2 - 1) \exp(x^2) + 1]}{[\exp(x^2) - 1]^2}. \quad \text{(5.11)}$$

It is easy to see that the numerator, $2x[(x^2 - 1) \exp(x^2) + 1]$, is an increasing function, and its value is 0 at $x = 0$. Therefore, the first-order derivative of $\rho^2(t_j)$ with respect to $x$ is less than 0 when $x > 0$. Clearly, $\rho^2(t_j)$ is a decreasing function in $x$. Since $x$ is an increasing function in $b$, $\rho^2(t_j)$ is a decreasing function in $b$. Moreover, when $x$ is positive, $\rho(t_j)$ is positive and has the same monotonicity as $\rho^2(t_j)$. Thus, $\rho(t_j)$ decreases in magnitude as $b$ increases in magnitude.

Because $\tilde{\rho}$ is a weighted average of $\rho(t_j)$ with all positive weights from (4.5), we can further claim that, for given $\sigma_Y(t_j)$ and $\sigma_{PD}(t_j)$, the magnitude of the robust correlation coefficient $\tilde{\rho}$ decreases as $b$ increases in magnitude.

Note that, based on (5.9) and (5.10), we obtain

$$\lim_{b \to 0^+} \rho(t_i) = 1, \quad \lim_{b \to 0^-} \rho(t_i) = -1.$$

Therefore,

$$\lim_{b \to 0^+} \tilde{\rho} = 1, \quad \lim_{b \to 0^-} \tilde{\rho} = -1.$$

Thus, the limit does not exist as $b$ approaches 0, and the Hull–White parametric approach could be very sensitive as $b$ approaches 0. We will discuss this further in the next section.

Further, we want to investigate the sensitivity of CVA to the underlying credit exposure volatility, ie, CVA Vega. The CVA Vega is the partial derivative of the CVA price with respect to the underlying volatility. We denote the underlying volatility by
\( \sigma_u \). The volatility of the credit exposure at time \( t_i \), \( \sigma_i \), is a function of \( \sigma_u \), ie. \( \sigma_i = f_i(\sigma_u) \). Obviously, \( f_i(\sigma_u) \) is always nonnegative and nondecreasing for all \( t_i \leq T \). By virtue of (5.8),

\[
\text{CVA}_{\text{ratio}} = 1 + \frac{b \sum_{j=1}^{K} (C_j \sum_{i=1}^{j} \sigma_i^2)}{\sum_{j=1}^{K} [C_j (\sum_{i=1}^{j} \mu_i + V_0)]} = 1 + \frac{b \sum_{j=1}^{K} (C_j \sum_{i=1}^{j} f_i(\sigma_u)^2)}{\sum_{j=1}^{K} [C_j (\sum_{i=1}^{j} \mu_i + V_0)].}
\]

(5.11)

Taking the derivative of (5.11) with respect to \( \sigma_u \) yields

\[
\frac{d\text{CVA}_{\text{ratio}}}{d\sigma_u} = \frac{b \sum_{j=1}^{K} (C_j \sum_{i=1}^{j} 2 f_i(\sigma_u) f_i'(\sigma_u))}{\sum_{j=1}^{K} [C_j (\sum_{i=1}^{j} \mu_i + V_0)]}.
\]

(5.12)

so the first-order derivative or CVA Vega is positive when \( b > 0 \) and negative when \( b < 0 \). In other words, \( \text{CVA}_{\text{ratio}} \) is increasing with respect to the underlying volatility when there is WWR, and decreasing when there is RWR.

6 IMPLICATIONS FOR CREDIT VALUE ADJUSTMENT ESTIMATION

From the analysis in the last section, we can see that the single parameter \( b \) in the Hull–White approach actually plays two roles: it models the same directional effect as the profile multiplier \( C_p \), and it has a negative impact on the robust correlation coefficient \( \tilde{\rho} \). In fact, from (5.7) and (5.6), we can see that the profile multiplier \( C_p \) increases with a rate

\[
\sqrt{\exp \left( b^2 \sum_{i=1}^{j} \sigma_i^2 \right) - 1},
\]

while the robust correlation coefficient \( \tilde{\rho} \) decreases with a slower rate of

\[
b \sqrt{\exp \left( b^2 \sum_{i=1}^{j} \sigma_i^2 \right) - 1}.
\]

Therefore, overall the increase in the profile multiplier \( C_p \) dominates the decrease in the robust correlation coefficient \( \tilde{\rho} \), and the combined effect is positive with respect to \( b \) (see (5.8)). In other words, the parameter \( b \) has an overall positive impact on the DWR adjustment on the CVA. The numerical results, presented in the next section (with the recovery rate \( R = 0.4 \)), demonstrate the above theoretical result (see Figures 2(b) and 4(b)).

In the Hull–White approach, the credit exposure \( V(t) \) needs to be estimated first, and then the PD hazard rate \( h(t) \) will be estimated by calibration with the market
data. If $V(t)$ can be simulated or estimated easily, then the Hull–White approach will be relatively easy to implement. In Hull and White (2012), the hazard rate is given in a very specific exponential form (see (3.1)) to describe the dependence of the counterparty default on the credit exposure. Other forms of function describing the dependence can also be considered. For example, Ruiz et al. (2015) consider the probability of default as a function of some market factor and examine functional forms such as power, exponential, logarithm and linear. In the Hull–White approach, the dependence is modeled by a single parameter, $b$, but the sensitivity with respect to $b$ does not seem very clear. In addition, once the dependence structure changes, the calibration needs to be performed again before reestimating $b$. Therefore, the Hull–White approach should be used with caution.

Unlike the Hull–White approach, the PCL approach does not depend on any specific form of the default hazard function. Moreover, in the latter approach, the profile multiplier $C_p$ and the robust correlation coefficient $\rho$ are considered separately, so that a more granular insight into the DWR can be obtained in the CVA calculation. The profile multiplier $C_p$ can capture well the CVA effect caused by the volatilities of the exposure and the counterparty credit quality. The DWR due to the correlation between the credit exposure $V(t)$ and the PD $q(t)$ is reflected in the robust correlation coefficient $\rho$. To implement the approach, we need to generate or simulate both the credit exposure and the PD simultaneously. Based on such data, we can obtain $C_p$, which does not change with the dependency between the credit exposure and the PD. In other words, $C_p$ stays the same across all levels of underlying correlations. The dependence will be characterized only by the robust correlation coefficient $\rho$, which can be estimated with the efficient curve-fitting algorithm given in Pang et al. (2015). When the risk manager does not have an accurate estimate of the correlation and only a confidence interval is given on the underlying or the robust correlation coefficient, a corresponding confidence interval for CVA can be derived. Moreover, if the CVA value has to be reestimated over time, then this approach can be quite efficient. In particular, when the dependence structure changes due to changes in market conditions, we simply change the value of $\rho$ and do not need to resimulate the credit exposure or the PDs.

The CVA DWR is only modeled by the single parameter $b$ in the Hull–White approach. However, even when $b$ is fixed, the CVA price can decrease as the underlying volatility increases in the presence of RWR, and it can increase as the underlying volatility increases in the presence of WWR. These phenomena are observed, for example, in the empirical results presented in Ruiz et al. (2015, Figure 9). On the other hand, the PCL approach with CVA DWR decomposition offers valuable details and provides an intuitive explanation for these phenomena. From (5.12), we can see that, for CVA Vega, the change in the underlying volatility has a negative effect on the CVA price if there is an RWR ($b < 0$) and a positive effect on the CVA price if there
is a WWR \( (b > 0) \). In other words, it is clear from (5.12) that the CVA price depends not only on the value of \( b \) but also on the portfolio profile (such as the underlying volatility). Therefore, the same value of \( b \) in the Hull–White approach does not always imply the same level of DWR, and the portfolio profile matters. Therefore, both the robust correlation and the profile multiplier can play key roles in DWR modeling. Considering only one of them may not be adequate.

In the Hull–White approach, the dependency between the default and the exposure is modeled with functions of certain forms, and there is no assumption made on the distributions of the credit exposure or the PD, so it is robust for all distributions. However, in the PCL approach, the dependence is described by a robust correlation coefficient, and it does not require any particular distribution assumption either. The approach works as long as the first and second moments exist and the linear correlation is adequate to describe the underlying dependency. In fact, most elliptic distributions can meet these assumptions. Although the analysis in Section 5 is based on the assumptions of normal distributions and the recovery rate \( R = 0 \), numerical results with \( R \neq 0 \) and no distribution assumption also indicate the same behavior. The numerical results and further discussion are presented in the next section.

7 NUMERICAL EXAMPLES

Typically, risk-neutral parameter calibration is carried out based on market prices. It is a straightforward exercise to calibrate the means and standard deviations using the market price of the underlying exposure and credit spread of the counterparty as long as they are available. However, calibration of the robust correlation coefficient \( \hat{\rho} \) could be difficult due to the lack of market price of DWR CVA. One approach is to use the observed historical correlation \( \rho \). In Hull and White (2012), \( a(t) \) is calibrated with the market-observed credit spread of the counterparty. Hull and White also present two approaches to estimate the value of \( b \): one uses historical data for credit exposure and the credit spread, and the other involves subjective judgment of the DWR. Ruiz et al (2015) presented some calibration results for models similar to the Hull–White approach. Aiming to investigate the dependence of the default density on market factors, they proposed several functional forms including linear, exponential and logarithmic. The models were calibrated using market-observed credit spread and different market factors, such as equity price, foreign exchange rates and commodity prices, but not calibrated using CVA prices.

In this paper, we focus on the analytical connection between the correlation and parametric approaches. Model calibration is left to a separate study. Instead, we illustrate our findings via a series of numerical examples for vanilla interest rate swaps. The numerical examples presented here are not based on assumptions of normal
distributions, but we do observe phenomenons that are consistent with the analytical results in Section 5.

7.1 Simulation models

For illustration, we consider two very common risk-free interest rate models. In particular, we use the short-rate Cox–Ingersoll–Ross (CIR) model (Cox et al 1985),

\[
dr_t = \kappa_r (\theta_r - r_t) \, dt + \sigma_r \sqrt{r_t} \, dW_t, \tag{7.1}
\]

and the model proposed by Vasicek (1977),

\[
dr_t = \kappa_r (\theta_r - r_t) \, dt + \sigma_r \, dW_t, \tag{7.2}
\]

where \( W_t \) is a standard Brownian motion, and \( \kappa_r, \theta_r \) and \( \sigma_r \) are the corresponding mean-reverting rate, mean-reverting level and volatility, respectively.

Assume the trader has exposures to the pay leg of a three-year fixed interest rate swap with quarterly payments. The parameters are chosen to be \( \kappa_r = 0.1, \theta_r = 0.05 \) and \( \sigma_r = 0.06 \), and we assume that at \( t_0 \) the interest rate term structure is flat and equal to \( \theta_r \). The fixed rate is also set to be \( \theta_r \). The simulated future exposures have time-dependent mean \( \mu_V(t_i) \) and volatility \( \sigma_V(t_i) \). This set of parameters is also used in Skoglund et al (2013).

DWR dependency is modeled with the Hull–White approach. Parameter \( b \) ranges from \(-0.4\) to \(0.4\). A recovery rate of \( R = 0.4 \) is used. Based on the simulated data, for each given \( b \) value, we calculate the profile multiplier \( C_p \), the robust correlation coefficient \( \tilde{\rho} \) and the combined CVA ratio \( (1 + \tilde{\rho} C_p) \). We investigate their sensitivities to the parameter \( b \) for both the CIR and Vasicek models. The numerical results are shown in Figures 1–4.

7.2 Numerical results and further discussion

From Figures 1 and 3, we can see that the profile multiplier \( C_p \) increases as the magnitude of \( b \) increases. This is true for both RWR and WWR. This result also holds for interest rates given by either the CIR model or the Vasicek model. In the presence of RWR, the magnitude of the profile multiplier ranges from 0 to slightly over 1. However, if WWR is presented, this number can reach almost 50 for the CIR model and 100 for the Vasicek model (see Figures 1 and 3). In other words, it seems like the profile multiplier \( C_p \) is much more sensitive to the parameter \( b \) with WWR (\( b > 0 \)) than with RWR (\( b < 0 \)).

The significant difference between RWR and WWR on the impact of the profile multiplier for the Hull–White approach is very interesting. It means that the Hull–White approach can catch the WWR well with a single variable, \( b \). Since the RWR is not of great concern to the dealer or the regulators, the significant lower sensitivity of
the profile multiplier to the value of $b$ is acceptable. This behavior is mainly due to the fact that the credit exposure truncation ($V^+(t) = \max\{V(t), 0\}$) causes a volatility change in the PDs. The PD volatility $\sigma_{PD}$ when WWR is presented ($b > 0$) tends to be larger than that when RWR is presented ($b < 0$). Now, from the profile multiplier
FIGURE 2 (a) Robust correlation coefficient $\tilde{\rho}$ versus $b$ and (b) CVA ratio versus $b$ for the CIR model.

definition (see (4.6)), it is easy to see that $C_P$ tends to be bigger when $b > 0$ and smaller when $b < 0$.

From Figures 2(a) and 4(a), we can see that the absolute value of the robust correlation coefficient decreases as $b$ increases in magnitude. This is consistent with the analysis in Section 5. Even though the exposures of vanilla swaps may not strictly
follow our assumptions (4), (8) and (9) in Section 5, we observe the same monotonicity in this numerical analysis.

Further, from Figures 2(b) and 4(b) we can see that the CVA ratio is increasing in $b$. Obviously a WWR trade faces a larger potential credit loss if the counterparty
default is more uncertain. Although Basel III does not award RWR, risk managers can still benefit from such a risk.

Another important finding is that, for the Hull–White approach, when $b$ is positive but very close to 0, the CVA ratio tends to be more sensitive to $b$ (see Figures 2(b)
From Figure 2(b), we can see that for the CIR interest rate model the CVA ratio changes from 1 to 5 when $b$ changes from 0 to 0.1. From Figure 4(b), we can see that for the Vasicek interest rate model the CVA ratio changes from 1 to 6.5 when $b$ changes from 0 to 0.1. As a reference, in Basel III (Basel Committee on Banking Supervision 2011) the suggested CVA ratio is only 1.4 for the WWR. Therefore, when the Hull–White approach is used, the CVA with WWR is very sensitive to $b$ when $b$ is small. Thus, more caution is needed for small values of $b$. A bad estimate of $b$ around 0 can be misleading.

8 CONCLUSION

In this paper, we developed a connection between the Hull–White parametric approach and the PCL correlation approach for CVA calculation. Hull and White (2012) provided a concise approach to capture DWR through a single parameter $b$. We provide a theoretical analysis of the dependence parameter $b$ in the Hull–White approach via the effect of the robust correlation and the profile multiplier introduced by Pang et al (2015), which sheds more light on the parameter in terms of economic interpretation and effect from the dependency and underlying risks. We find that $b$ in the Hull–White approach is affected not only by the robust correlation (an indicator of the underlying dependency) but also by the profile multiplier (severity of the underlying risk). The Hull–White approach also shows that these two underlying effects partly offset each other when they are both reflected through the single parameter $b$. Whether this offset is desirable in model validation or not requires further attention. In addition, the Hull–White model could be very sensitive to $b$ when it is positive and very close to 0. Our results help us to better understand the Hull–White approach. In addition, in most cases the PCL correlation approach by Pang et al (2015) is easier to understand while also offering straightforward implementation.

DECLARATION OF INTEREST

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

REFERENCES


Research Paper

The use of the triangular approximation for some complicated risk measurement calculations

Nick Georgiopoulos

Financial Stability Department, Bermuda Monetary Authority, 43 Victoria Street, Hamilton, Bermuda; email: ngeorgiopoulos@bma.bm

(Received August 19, 2016; revised December 16, 2016; accepted February 21, 2017)

ABSTRACT

We introduce the triangular approximation to the normal distribution in order to extract closed- and semi-closed-form solutions that are useful in risk measurement calculations. In risk measurement models there is usually a normal distribution together with some other distributions in a portfolio of risks. Exceedance probability or value-at-risk (VaR) calculations for these portfolios require simulations. However, with the use of the triangular approximation to the normal density we can have closed-form solutions for risk measurements using actuarial models that include not only insurance risk, such as gamma- and Pareto-distributed losses, but also financial risk. We also approximate the collective risk model under lognormally distributed severities and estimate its VaR. We evaluate the accuracy of the analytic solutions versus the Monte Carlo estimates and, according to our results, the analytic solutions requiring much less computational time are quite good.

Keywords: normal distribution; triangular approximation; collective risk model; model validation; value-at-risk (VaR); insurance.
1 INTRODUCTION

We will show how the triangular distribution that can approximate the normal distribution can be used to derive interesting results in risk management. The triangular distribution receives little attention in the literature, primarily because it lacks the theoretical foundations of other distributions. Yet its simplicity, as we shall show, enables us to approximate risk measurements with simple analytic functions, which would otherwise require numerical approximation or simulation.

This not only allows much faster calculation but also offers true insight into the key risk drivers, since sensitivities can easily be calculated. In particular, we will calculate probabilities that insurance losses will exceed financial losses in the cases of gamma- and Pareto-distributed insurance losses and normally distributed financial losses. In addition, we will study the collective risk insurance model with lognormal severities, and derive closed-form solutions for its probability distribution and value-at-risk (VaR).

This paper differentiates itself from typical approaches to risk measurement in the literature, such as those of Jamshidian and Zhu (1997), who use simulations for risk measurement of VaR, or Panjer (1981), who approximates the distribution function of the collective risk model by taking advantage of the properties of the frequency distribution. Other risk measurement techniques include the scenario-based optimization approach by Rockafellar and Uryasev (2000), who use discrete scenarios to jointly estimate VaR and conditional VaR by taking advantage of the properties of the conditional VaR using convex analysis. Our technique establishes closed-form solutions for cases when normal and lognormal variables need to be either convoluted (collective risk model) or used in multidimensional integrals with other distributions (default probabilities). The scope of this paper is further enhanced by the prevalence of the normal and lognormal distributions in insurance and financial risk management.

Nowadays computing power is cheap; therefore, the use of simulation is ubiquitous in the finance industry. Quantities such as default probabilities and upper percentiles are used for the calculation of capital requirements and solvency capital requirements as well as for economic capital budgeting. Despite the low cost of computing power, computing resources require time and programming effort. However, closed-form solutions for relevant risk quantities are equally effective, and they have almost zero computing requirements. Finally, simulation approaches make it more difficult to explicitly understand the risk drivers and perform sensitivity analysis, even though there are techniques to make simulation more transparent.

The key result of our method is a rather obscure technique from the engineering world that approximates the normal distribution with a triangular distribution. This approximation replaces the complicated normal density function with a simple polynomial function of a first-degree polynomial. Thus, integrals of normal densities are
easily replaced by integrals of polynomial functions. This approximation was presented by Scherer et al (2003), who showed by matching the mean and the standard deviation that a triangular density is a good approximation to the normal density. They perform extensive Monte Carlo simulations to validate their result and conclude that the triangular approximation works well for approximations within 2.4 standard deviations above the mean. In this paper we will extend the result of Scherer et al so as to capture the upper percentiles of a normal distribution in order to estimate risk measures such as VaR.

This approximation technique has been used in operational research, management and engineering. Samorani and Ganguly (2016) use this approximation to solve an optimal sequencing problem of nonpunctual patients in a clinic. Schmitt et al (2009) use it to solve problems of inventory management with stochastic supply and demand. There are many papers in the field of operations research employing this technique, but, to the best of the author’s knowledge, the actuarial mathematics, finance and risk management literature contains no paper using the triangular approximation.

Therefore, this paper serves as an introduction of this useful technique to real-life applications in actuarial mathematics, and risk management in particular. Risk managers and actuaries will be able to replace black box simulation with models that can be implemented very easily in a simple spreadsheet or as routines in a larger enterprise risk management model. In this paper we provide both the formulas and validation results. In particular, the closed-form solution for the probability distribution of the collective risk model will warrant multiple validation checks, since the result hinges on extreme value theory, which facilitates the algebra of the calculation.

The structure of this paper is as follows. First, we briefly describe the triangular approximation to the normal distribution. Second, we show a series of results relevant for financial and actuarial calculations, and we extract closed-form solutions for probabilities when there are gamma- and Pareto-distributed insurance losses and a normally distributed financial risk variable. Finally, we show a closed-form solution for the distribution of the collective risk model with lognormal severities. Throughout the paper we validate our results with the use of Monte Carlo experiments. The final section states our conclusions.

2 THE TRIANGULAR APPROXIMATION

If we observe the normal density, it looks like a nice, symmetric bell. At the same time, if we draw an isosceles triangle around the bell, we observe that the triangle can contain the normal density without leaving too much density outside the triangle (Figure 1).
Although the fitting is not perfect, the nature of the normal distribution, which has rather thin tails, allows a good approximation of the normal density with the triangular density. Scherer et al (2003) showed using Monte Carlo experiments that this approximation works very well by fitting a triangular distribution with first and second moment matching. Without delving into the details of the derivation of the result, we show that the approximation of a triangular density to the normal density with parameters $\mu$ and $\sigma$ is expressed mathematically as

$$f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & a \leq x \leq c, \\ \frac{2(b-x)}{(b-a)(b-c)} & c < x \leq b, \\ 0 & \text{otherwise}, \end{cases}$$

(2.1)

with

$$c = \mu,$$  
(2.2)

$$a = c - \sigma \sqrt{6},$$  
(2.3)

$$b = c + \sigma \sqrt{6}.$$  
(2.4)

Substituting (2.2)–(2.4) into (2.1), we have the following triangular density function that approximates the normal density function:

$$f(x) = \begin{cases} \frac{x-c + \sigma \sqrt{6}}{6\sigma^2} & a \leq x \leq c, \\ \frac{c + \sigma \sqrt{6} - x}{6\sigma^2} & c < x \leq b, \\ 0 & \text{otherwise}. \end{cases}$$

(2.5)
With this result we will calculate risk-related quantities when there is either both financial and insurance risk or some complicated form of insurance risk. In this paper, financial risk will be described by a normal distribution; for insurance risk, known distributions such as the gamma, the Pareto and the compound Poisson will be used.

3 GAMMA CLAIMS AND NORMAL ASSET RETURNS

Gamma claims arise naturally in actuarial risk theory. Compound Poisson models of a portfolio of losses are usually approximated by the so-called translated gamma distribution. For details of this approximation, the reader is referred to Dickson (2010, p. 86). A useful risk measure (not to be confused with other risk measures such as VaR) is the probability that the insurance losses of an insurance company, which collects premiums for a portfolio of risks and then invests the proceeds in financial markets, will surpass the financial losses of the investment portfolio.

Assume that the random variable of insurance losses is $Y_t$, with density function $g(y)$, while the value of the investment portfolio is the random variable $X_t$, with density function $f(x)$. Asset values are independent of insurance losses, which makes sense since we would not expect events such as automobile crashes or medical liability to have an affect on financial markets. The insurance company receives a premium $C = E[Y_t]$, assuming that the premium is calculated according to the expectation method. We can also include premium loadings for expenses and a capital buffer, but the mathematics will not change. The assets grow according to a geometric Brownian motion, ie,

$$X_{t+dt} = X_t \exp((\mu - \frac{1}{2}\sigma^2)dt + \sigma(W_{t+dt} - W_t)).$$  \hspace{1cm} (3.1)

In (3.1) the term $W_{t+dt} - W_t$ is the increment of a Wiener process, while the parameters $\mu$ and $\sigma$ denote the drift and the volatility of the geometric Brownian motion, respectively. If we assume that $dt$ is small, we can log linearize (3.1) using a first-order Taylor expansion of the exponential function such that the change in asset value is

$$X_{t+dt} - X_t \approx X_t((\mu - \frac{1}{2}\sigma^2)dt + \sigma(W_{t+dt} - W_t)).$$  \hspace{1cm} (3.2)

\[1\] From now on $dt = 1$, meaning that the time step is “one period” to simplify the mathematics. The unit value can be interpreted as one day; therefore, the parameters can be assumed to have been estimated from daily data. Nevertheless, in practice the risk manager will scale the time to match the sample frequency.
In fact, (3.2) says that the returns are transformed into log returns, since for a small \( dt \) the approximation \( \ln x \approx x - 1 \) holds. For the model in (3.2) we have

\[
\begin{align*}
\bar{\mu} &= E[\bar{X}_{t+dt}] = (\mu - \frac{1}{2}\sigma^2)X_t \, dt, \\
\bar{\sigma}^2 &= \text{Var}[\bar{X}_{t+dt}] = \sigma^2 X_t^2 \, dt.
\end{align*}
\]

(3.3) (3.4)

Before we proceed, it should be kept in mind that nonlinear financial assets such as derivatives cannot be covered by the model described above. However, in practice, insurers mostly invest in bonds and plain equity; therefore, we would expect that most insurers’ asset portfolios can be described within this setting. Since all the premium is invested in financial assets, we have that \( C = E[Y_t] = X_t \). The probability of default is defined as

\[
\Pr(Y_{t+dt} \geq \bar{X}_{t+dt}).
\]

(3.5)

Now, this probability does not imply default in the pure legal sense. It is rather a convenient name to allow us to understand the mathematical setting. “Default” in this paper is the excess of losses compared with losses of assets that are funded by insurance liabilities. To solve for this probability we have a two-dimensional domain of integration, shown in Figure 2. The domain of integration is the region above the red line. With the domain of integration in place, we now have the following:

\[
\begin{align*}
\Pr(Y_{t+dt} \geq \bar{X}_{t+dt}) &= \int_{-\infty}^{0} \int_{0}^{\infty} f(x)g(y) \, dy \, dx + \int_{0}^{\infty} \int_{x}^{\infty} f(x)g(y) \, dy \, dx \\
&= \Pr(\bar{X}_{t+dt} \leq 0) + \int_{0}^{\infty} \int_{x}^{\infty} f(x)g(y) \, dy \, dx.
\end{align*}
\]

(3.6)
If we assume gamma-distributed losses \( Y_t \) with parameters \( \alpha \) and \( \beta \), we can write (3.6) as

\[
\Pr(\bar{X}_{t+dt} \leq 0) + \int_0^\infty \int_x^\infty f(x)g(y) \, dy \, dx \\
= \Pr(\bar{X}_{t+dt} \leq 0) + \int_0^\infty \left( 1 - \frac{\gamma(\alpha, x/\beta)}{\Gamma(\alpha)} \right) f(x) \, dx. \tag{3.7}
\]

The integral \( \int_x^\infty g(y) \, dy \) is in fact the probability \( \Pr(Y_{t+dt} > x) \), and it is replaced by the cumulative gamma distribution function. The function \( \gamma(\alpha, x/\beta) \) is the lower incomplete gamma function. Moreover, using the triangular approximation, (3.7) is written as

\[
\Pr(Y_{t+dt} \geq \bar{X}_{t+dt}) \\
= 1 - \frac{1}{6\sigma^2 \Gamma(\alpha)} \left( \int_0^{\bar{\mu}} \gamma(\alpha, \frac{x}{\beta})(x - \bar{\mu} + \bar{\sigma} \sqrt{6}) \, dx \\
+ \int_{\bar{\mu}}^b \gamma(\alpha, \frac{x}{\beta})(\bar{\mu} + \bar{\sigma} \sqrt{6} - x) \, dx \right), \quad b = \bar{\mu} + \bar{\sigma} \sqrt{6}. \tag{3.8}
\]

The lower incomplete gamma function can be evaluated using an infinite series representation of the form

\[
\gamma(\alpha, \frac{x}{\beta}) = \sum_{k=0}^{\infty} \frac{(-1)^k (x/\beta)^{\alpha + k}}{k! \alpha + k}. \tag{3.9}
\]

Observing (3.8) and (3.9), we see that we have integrals of polynomials that can be solved analytically. We now define the following integrals:

\[
I_1 = \int_0^{\bar{\mu}} x \gamma(\alpha, \frac{x}{\beta}) \, dx, \tag{3.10}
\]

\[
I_2 = \int_0^{\bar{\mu}} \gamma(\alpha, \frac{x}{\beta}) \, dx, \tag{3.11}
\]

\[
I_3 = \int_{\bar{\mu}}^b \gamma(\alpha, \frac{x}{\beta}) \, dx, \tag{3.12}
\]

\[
I_4 = \int_{\bar{\mu}}^b x \gamma(\alpha, \frac{x}{\beta}) \, dx. \tag{3.13}
\]
We calculate each integral separately:

\[ I_1 = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_{0}^{\tilde{\mu}} x \frac{(x/\beta)^{\alpha+k}}{\alpha+k} \, dx \]

\[ = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{\beta^{-(\alpha+k)}}{(\alpha+k)(\alpha+k+2)} \tilde{\mu}^{\alpha+k+2}, \quad (3.14) \]

\[ I_2 = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_{0}^{\tilde{\mu}} x \frac{(x/\beta)^{\alpha+k}}{\alpha+k} \, dx \]

\[ = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{\beta^{-(\alpha+k)}}{(\alpha+k)(\alpha+k+1)} \tilde{\mu}^{\alpha+k+1}, \quad (3.15) \]

\[ I_3 = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_{b}^{\mu} x \frac{(x/\beta)^{\alpha+k}}{\alpha+k} \, dx \]

\[ = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{\beta^{-(\alpha+k)}}{(\alpha+k)(\alpha+k+1)} (b^{\alpha+k+1} - \tilde{\mu}^{\alpha+k+1}), \quad (3.16) \]

\[ I_4 = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_{b}^{\mu} x \frac{(x/\beta)^{\alpha+k}}{\alpha+k} \, dx \]

\[ = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{\beta^{-(\alpha+k)}}{(\alpha+k)(\alpha+k+2)} (b^{\alpha+k+2} - \tilde{\mu}^{\alpha+k+2}). \quad (3.17) \]

Note now that

\[ \text{Pr}(Y_{t+\Delta t} \geq \tilde{X}_{t+\Delta t}) = 1 - \frac{1}{6\tilde{\sigma}^2 \Gamma(\alpha)} (I_1 - (\tilde{\mu} - \tilde{\sigma} \sqrt{6}) I_2 + (\tilde{\mu} + \tilde{\sigma} \sqrt{6}) I_3 - I_4), \quad (3.18) \]

and after doing the algebra we get

\[ \text{Pr}(Y_{t+\Delta t} \geq \tilde{X}_{t+\Delta t}) \]

\[ = 1 - \frac{1}{6\tilde{\sigma}^2 \Gamma(\alpha)} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{\beta^{-(\alpha+k)}}{(\alpha+k)(\alpha+k+2)} (2\tilde{\mu}^{\alpha+k+1} - b^{\alpha+k+1}) \]

\[ + \frac{1}{6\tilde{\sigma}^2 \Gamma(\alpha)} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} (\tilde{\mu} - \tilde{\sigma} \sqrt{6}) \beta^{-(\alpha+k)} \tilde{\mu}^{\alpha+k+1} \]

\[ - \frac{1}{6\tilde{\sigma}^2 \Gamma(\alpha)} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} (\tilde{\mu} + \tilde{\sigma} \sqrt{6}) \beta^{-(\alpha+k)} (b^{\alpha+k+1} - \tilde{\mu}^{\alpha+k+1}). \quad (3.19) \]

This is a closed-form solution of the default probability with an infinite series representation. In practice, in order to program the formula in a computer we need
to sum up to a large $k$ using a “while” loop, but the factorial functions will explode with a large $k$ and will return NaNs (“not-a-number”). In MATLAB the factorial function with an argument above 170 will yield infinity, and in our calculations for the comparative statics the sums run from 0 to 170. Finally, to fully formalize the solution we need to prove convergence of the infinite sums and provide conditions for convergence. To do that we will prove the following lemma.

**Lemma 3.1**  

The series

$$
\sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{(\bar{\mu} - \bar{\sigma} \sqrt{6}) \beta^{-(\alpha+k)}}{(\alpha + k)(\alpha + k + 1)} \mu^{\alpha+k+1}
$$

converges.

**Proof**

We define the sequence

$$
z(k) = \frac{(-1)^k}{k!} \frac{(\bar{\mu} - \bar{\sigma} \sqrt{6}) \beta^{-(\alpha+k)}}{(\alpha + k)(\alpha + k + 1)} \mu^{\alpha+k+1}.
$$

We will use the ratio test to test the convergence of the series. To prove convergence of the series with the ratio test, we need to calculate the limit $\lim_{k \to \infty} z(k+1)/z(k)$. If the absolute value of the limit is less than 1, then the series converges. We now have

$$
\lim_{k \to \infty} \frac{z(k+1)}{z(k)} = \lim_{k \to \infty} \left( \frac{(-1)^{k+1}}{(k+1)!} \frac{(\bar{\mu} - \bar{\sigma} \sqrt{6}) \beta^{-(\alpha+k+1)}}{(\alpha + k + 1)(\alpha + k + 2)} \mu^{\alpha+k+2} \right) \times \left( \frac{(-1)^k}{k!} \frac{(\bar{\mu} - \bar{\sigma} \sqrt{6}) \beta^{-(\alpha+k)}}{(\alpha + k)(\alpha + k + 1)} \mu^{\alpha+k+1} \right)^{-1} = \lim_{k \to \infty} \frac{-1}{k+1} \frac{(\alpha + k) \beta^{-1}}{(\alpha + k + 2)} \mu = 0.
$$

The same test also holds for the convergence of the other sums.

\[\blacksquare\]

**4 PARETO CLAIMS AND NORMAL ASSET RETURNS**

Gamma claims can be thought of as claims emanating from an “average” portfolio of insurance risks. By “average” we mean well diversified with multiple lines of business that dampen the total volatility of the portfolio. If the insurance company is not well diversified (e.g., a property catastrophe reinsurer or a monoline mortgage (re)insurer), then gamma losses may not be sufficient to describe the portfolio. Usually Pareto distributions describe such high-risk portfolios. The cumulative distribution function of a Pareto random variable $Y_t$ is

$$
\Pr(Y_t \leq y) = 1 - \left( \frac{\lambda}{\lambda + y} \right)^\alpha,
$$

(4.1)
where $\lambda$ and $\alpha$ are scale and shape parameters. The parameter $\alpha$ is also called the Pareto index. The lower the value of $\alpha$, the higher the risk. For example, if $\alpha \leq 2$, the variance of a Pareto distribution is infinite. To estimate the default probability with Pareto-distributed claims with density $g(y)$, we use (3.6). We obtain

\[
\Pr(\widetilde{X}_{t+\Delta t} \leq 0) + \int_0^\infty \int_x^\infty f(x)g(y)\,dy\,dx
\]

\[
= \Pr(\widetilde{X}_{t+\Delta t} \leq 0) + \int_0^\infty (1 - \Pr(Y_t \leq x)) f(x)\,dx
\]

\[
= \Pr(\widetilde{X}_{t+\Delta t} \leq 0) + \int_0^\infty \left(\frac{\lambda}{\lambda + x}\right)^\alpha f(x)\,dx
\]

\[
= \Pr(\widetilde{X}_{t+\Delta t} \leq 0) + \frac{1}{6\sigma^2} \int_{\bar{\mu}}^{\tilde{\mu}} \left(\frac{\lambda}{\lambda + x}\right)^\alpha (x - \bar{\mu} + \tilde{\sigma} \sqrt{6}) \,dx
\]

\[
+ \frac{1}{6\sigma^2} \int_{\bar{\mu}}^{b} \left(\frac{\lambda}{\lambda + x}\right)^\alpha (\bar{\mu} + \bar{\sigma} \sqrt{6} - x) \,dx. \tag{4.2}
\]

We can calculate $\Pr(\widetilde{X}_{t+\Delta t} \leq 0)$ from the tables of the normal distribution. We now need to calculate the integrals:

\[
J_1 = \frac{1}{6\sigma^2} \int_{\bar{\mu}}^{\tilde{\mu}} \left(\frac{\lambda}{\lambda + x}\right)^\alpha (x - \bar{\mu} + \tilde{\sigma} \sqrt{6}) \,dx, \tag{4.3}
\]

\[
J_2 = \frac{1}{6\sigma^2} \int_{\bar{\mu}}^{b} \left(\frac{\lambda}{\lambda + x}\right)^\alpha (\bar{\mu} + \bar{\sigma} \sqrt{6} - x) \,dx. \tag{4.4}
\]

We obtain

\[
J_1 = \frac{1}{6\sigma^2} \int_{\bar{\mu}}^{\tilde{\mu}} \left(\frac{\lambda}{\lambda + x}\right)^\alpha (x - \bar{\mu} + \tilde{\sigma} \sqrt{6}) \,dx
\]

\[
= \frac{\lambda^\alpha}{6\sigma^2} \left(\int_{\bar{\mu}}^{\tilde{\mu}} \frac{x - \bar{\mu} + \tilde{\sigma} \sqrt{6}}{(\lambda + x)^\alpha} \,dx - (\bar{\mu} - \sigma \sqrt{6}) \int_{\bar{\mu}}^{\tilde{\mu}} \frac{1}{(\lambda + x)^\alpha} \,dx\right)
\]

\[
= \frac{\lambda^\alpha}{6\sigma^2} \left(\int_{\lambda}^{\tilde{\mu} + \lambda} \frac{x - \bar{\mu} + \tilde{\sigma} \sqrt{6}}{x^\alpha} \,dx - (\bar{\mu} - \tilde{\sigma} \sqrt{6}) \int_{\lambda}^{\bar{\mu} + \lambda} \frac{1}{x^\alpha} \,dx\right)
\]

\[
= \frac{\lambda^\alpha}{6\sigma^2} \left(\int_{\lambda}^{\tilde{\mu} + \lambda} x^{1-\alpha} \,dx - (\lambda + \bar{\mu} - \tilde{\sigma} \sqrt{6}) \int_{\lambda}^{\bar{\mu} + \lambda} x^{-\alpha} \,dx\right)
\]

\[
= \frac{\lambda^\alpha}{6\sigma^2} \left(\frac{(\bar{\mu} + \lambda)^{2-\alpha} - \lambda^{2-\alpha}}{2 - \alpha} - (\lambda + \bar{\mu} - \tilde{\sigma} \sqrt{6}) \left(\frac{\bar{\mu} + \lambda)^{1-\alpha} - \lambda^{1-\alpha}}{1 - \alpha}\right)\right). \tag{4.5}
\]

Note that meaningful values for $J_1$ apply only for $\alpha > 2$, which implies that the Pareto distribution should have defined both the first and the second moment. Likewise, for
the other integral we obtain

\[ J_2 = \frac{1}{6\sigma^2} \int_{\tilde{\mu}}^{b} \left( \frac{\lambda}{\lambda + x} \right)^{\alpha} (\tilde{\mu} + \tilde{\sigma} \sqrt{6} - x) \, dx \]

\[ = \frac{\lambda^\alpha}{6\sigma^2} \left( (\tilde{\mu} + \tilde{\sigma} \sqrt{6}) \int_{\tilde{\mu}}^{\lambda} \frac{1}{(\lambda + x)^\alpha} \, dx - \int_{\tilde{\mu}}^{b} \frac{x}{(\lambda + x)^\alpha} \, dx \right) \]

\[ = \frac{\lambda^\alpha}{6\sigma^2} \left( (\tilde{\mu} + \tilde{\sigma} \sqrt{6} + \lambda) \int_{\tilde{\mu} + \lambda}^{b + \lambda} x^{-\alpha} \, dx - \int_{\tilde{\mu} + \lambda}^{b + \lambda} \frac{x - \lambda}{x^{-\alpha}} \, dx \right) \]

\[ = \frac{\lambda^\alpha}{6\sigma^2} \left( (\tilde{\mu} + \tilde{\sigma} \sqrt{6} + \lambda) \frac{(b + \lambda)^{1-\alpha} - (\tilde{\mu} + \lambda)^{1-\alpha}}{1-\alpha} \right. \]

\[ \left. - \frac{(b + \lambda)^{2-\alpha} - (\tilde{\mu} + \lambda)^{2-\alpha}}{2-\alpha} \right). \] (4.6)

and meaningful values for \( J_2 \) apply only for \( \alpha > 2 \). These restrictions stem from the integral

\[ \int_{0}^{\infty} (1 - \Pr(Y_{t+dt} \leq x)) f(x) \, dx \]

since it resembles the calculation of the variance of \( Y_t \) from the cumulative distribution function. Therefore, the default probability is

\[ \Pr(Y_{t+dt} \geq \tilde{X}_{t+dt}) = \Pr(\tilde{X}_{t+dt} \leq 0) + J_1 + J_2. \] (4.7)

### 5 COMPARATIVE STATICS AND VALIDATION OF GAMMA AND PARETO LOSS MODELS

To validate our models we use the Monte Carlo estimator. For this we sample random numbers from the normally distributed random variable \( X_{t+dt} \) as well as the gamma and Pareto distributions. The form of the Monte Carlo estimator is

\[ \frac{1}{M} \sum_{n=1}^{M} \mathbb{1}_{\{Y_n > X_n\}}, \]

where \( M \) is the number of random numbers drawn, and \( Y_n, X_n \) are the random samples from the Pareto or gamma and from the normal distribution, respectively. Moreover, we plot the changes in the probability of default against the asset volatility \( \sigma \) and drift \( \mu \). For the second row of Figure 3 we draw 30 000 scenarios.
FIGURE 3  Validation and comparative statics for gamma losses.

We observe that the Monte Carlo estimator converges quickly after fewer than 100,000 scenarios, while the differences between the closed-form solution and the Monte Carlo estimate reduce as the number of scenarios increases. Overall the closed-form solution fares well with respect to brute force simulation.

An interesting comparative static is the nonmonotonic relationship between asset volatility $\sigma$ and the probability of default. There is an optimum volatility at which the probability of default is maximized, and afterward it reduces steadily. This is because a higher asset volatility increases the probability of the asset losses surpassing the claim losses. In addition to $\sigma$, we plot the closed-form and Monte Carlo estimates with respect to changes in the parameter $\mu$. We observe that the closed-form solution diverges when we increase $\mu$ above 0.7.

This divergence is explained by the fact that the triangular density on the two-dimensional domain changes position relative to the gamma density. Thus, as $\mu$ increases, with volatility $\sigma$ constant, there is more probability mass left out on the
FIGURE 4 Validation and comparative statics for Pareto losses.

Column (a) shows $P(Y_{t+1} > \bar{X}_{t+1})$ for the Monte Carlo estimate (solid line) and the closed-form solution (dashed line). Column (b) shows the difference between the Monte Carlo estimate and the closed-form solution. First row parameters: $\mu = 0.03, \sigma = 0.2, \lambda = 1, \alpha = 2.5$. Second row parameters: $\mu = 0.03, \sigma = 0.2, \lambda = 1$.

left side. However, for practical financial applications the value of $\mu$ is usually below 0.2, and thus the approximation will not develop a serious problem.

In Figure 4 we show validation results when the insurance losses are of Pareto type. For the second row we draw 150 000 scenarios.

Again we observe that increasing the number of scenarios yields more accurate results. More scenarios are required to reduce the volatility of the random seeds for Pareto losses than for gamma losses. Note that in both cases there is an upward bias in the Monte Carlo estimates compared with the closed-form solutions. This stems from the fact that the triangular approximation leaves some probability density outside the normal density when it is fitted. However, these differences are in the third decimal place.

In the comparative statics, when different parameters are changed the differences between the Monte Carlo estimates and the closed-form solutions are minimal. Therefore, we can see that the results of the closed-form solution compare very well with the Monte Carlo estimator if we increase the number of scenarios, or if we keep the number constant and change the parameters of the distribution. Overall, we can confirm that the triangular approximation fares quite well when used for calculating
two-dimensional probability problems. Again we must be aware of potential problems with large drift parameters, which are usually ruled out for financial returns.

6 FREQUENCY–SEVERITY MODELS

Most actuaries, especially in property (re)insurance, use frequency–severity models to model loss portfolios. One of the most widely used is the compound Poisson process. In this model, the frequency of losses is modeled by a Poisson process that counts the number of individual claim events at a particular point in time. In addition to the claim-counting process, there is a claim-amount process (severity), which describes the amount claimed for a particular loss event.

Therefore, there are two sources of uncertainty. The first is whether a claim has arrived. The second is the amount the (re)insurer has to pay if the claim arrives. Usually, we assume that the frequency of claims and the severity of the claims-amount processes are independent. Formally, this process is described by the following equation:

\[ L_{t+dt} = \sum_{n=0}^{N_{t+dt}} X_n \]  

(6.1)

(further details of the frequency–severity model can be found in Dickson (2010)).

The frequency process, \( N(t) \), is usually a Poisson process, while the severity process, \( X_n \), can be any distribution with support on the positive side of the real line. All random variables \( X_i, i = 1, \ldots, n \), are independent and identically distributed. The probability distribution of \( L_t \) is almost impossible to find in closed form due to the fact that it involves the \( N_t \)th convolution of the random variables \( X_n \), whose distribution functions can be quite cumbersome to work with. For example, the sum of lognormal distributions is not lognormal itself. Therefore, numerical techniques including simulation or some more nuanced approaches that take advantage of the form of the frequency distribution (such as Panjer’s recursion) have been developed.

However, if \( X_n \) is lognormal with parameters \( \mu \) and \( \sigma \), we can use the results of the triangular approximation to find the probability distribution function with the help of the subexponential property of the lognormal distribution (for more on the notion of subexponentiality, the reader is referred to Foss et al (2007)). Now, instead of a difficult integral, we can use analytic solutions for the convolution, and then we can work with the sum of probability weights of the frequency distribution, since the counting process is discrete. Formally, we have the following expression for the probability that the compound Poisson will exceed an arbitrary \( k \):

\[ \Pr(L_{t+dt} \geq k) = \Pr\left( \sum_{n=0}^{N_{t+dt}} X_n \geq k \right). \]  

(6.2)
By using the total probability theorem, we have the following:

\[
\Pr \left( \sum_{n=0}^{N_{t+dt}} X_n \geq k \right) = \sum_{i=0}^{\infty} \Pr \left( \sum_{n=0}^{N_{t+dt}} X_n \geq k \mid N_{t+dt} = i \right) \Pr(N_{t+dt} = i). \tag{6.3}
\]

We now analyze the severity part in the sum in (6.3) as follows:

\[
\Pr \left( \sum_{n=0}^{N_{t+dt}} X_n \geq k \mid N_{t+dt} = i \right) = \Pr \left( \sum_{n=0}^{i} X_n \geq k \mid N_{t+dt} = i \right) \sim \Pr \left( \max_{n=0,\ldots,i} X_n \geq k \mid N_{t+dt} = i \right)
\]

\[
= 1 - \Pr \left( \max_{n=0,\ldots,i} X_n < k \mid N_{t+dt} = i \right)
\]

\[
= 1 - (\Pr(X_n < k \mid N_{t+dt} = i))^i
\]

\[
= 1 - (\Pr(\ln X_n < \ln k \mid N_{t+dt} = i))^i. \tag{6.4}
\]

For a lognormal random variable $X_n$ with parameters $\mu$ and $\sigma$, the natural logarithm $\ln X_n$ is normally distributed with mean $\mu$ and standard deviation $\sigma$. We can now use the triangular approximation to the normal distribution to evaluate the probability $\Pr(\ln X_n < \ln k \mid N_{t+dt} = i)$. There are two cases that we need to consider: $\mu > \ln k$ and $\mu \leq \ln k$. We will work through each case separately.

**Case 1** ($\mu > \ln k$) In this case it is straightforward to work with the triangular approximation. The first branch of the triangular density is used, and we obtain

\[
\Pr(\ln X_n < \ln k \mid N_{t+dt} = i)
\]

\[
= \int_{\ln k}^{\mu - \sigma \sqrt{6}} \frac{z - \mu + \sigma \sqrt{6}}{6 \sigma^2} \, dz
\]

\[
= \frac{1}{6 \sigma^2} \left( \int_{\ln k}^{\mu - \sigma \sqrt{6}} z \, dz - (\mu - \sigma \sqrt{6}) \int_{\mu - \sigma \sqrt{6}}^{\ln k} \, dz \right)
\]

\[
= \frac{\ln^2 k - (\mu - \sigma \sqrt{6})^2}{12 \sigma^2} - \frac{(\mu - \sigma \sqrt{6}) \ln k - (\mu - \sigma \sqrt{6})^2}{6 \sigma^2}
\]

\[
= \frac{\ln^2 k - 2(\mu - \sigma \sqrt{6}) \ln k + (\mu - \sigma \sqrt{6})^2}{12 \sigma^2}
\]

\[
= \frac{(\ln k - (\mu - \sigma \sqrt{6}))^2}{12 \sigma^2}. \tag{6.5}
\]
Combining (6.5) with (6.3) yields

\[ \Pr(L_{t+dt} \geq k) \sim 1 - \sum_{i=0}^{\infty} \left( \frac{(\ln k - (\mu - \sigma \sqrt{6}))^2}{12\sigma^2} \right)^i \Pr(N_{t+dt} = i). \] (6.6)

We now have to prove convergence for the infinite series in (6.6). In actuarial practice, either a Poisson frequency process or a binomial or negative binomial distribution is used for the frequency of losses. For a binomial distribution, the upper limit of the sum is not infinite by default. Therefore, convergence is self-evident. Both the Poisson and negative binomial distributions belong to the so-called Panjer family of distributions. For these distributions, the Panjer recursion applies:

\[ \Pr(N_{t+dt} = i + 1) = \left( \gamma + \frac{\delta}{i+1} \right) \Pr(N_{t+dt} = i). \]

In order to prove convergence of the infinite series in (6.6) we will use the ratio test. This is shown in the next lemma.

**Lemma 6.1** The series

\[ \sum_{i=0}^{\infty} \left( \frac{(\ln k - (\mu - \sigma \sqrt{6}))^2}{12\sigma^2} \right)^i \Pr(N_{t+dt} = i) \]

converges for the Poisson process. For the negative binomial distribution we have convergence if \((\ln k - (\mu - \sigma \sqrt{6}))^2 < 12\sigma^2 / \gamma.\)

**Proof** We define the following sequence:

\[ z(i) = \left( \frac{(\ln k - (\mu - \sigma \sqrt{6}))^2}{12\sigma^2} \right)^i \Pr(N_{t+dt} = i). \]

To prove convergence of the infinite series, we need to calculate \(\lim_{i \to \infty} z(i + 1)/z(i).\) If the absolute value of the limit is less than 1, then the series converges. We now have

\[
\lim_{i \to \infty} \frac{z(i + 1)}{z(i)} = \lim_{i \to \infty} \frac{(\ln k - (\mu - \sigma \sqrt{6}))^2 / 12\sigma^2)^i+1 \Pr(N_{t+dt} = i + 1)}{(\ln k - (\mu - \sigma \sqrt{6}))^2 / 12\sigma^2)^i \Pr(N_{t+dt} = i)}
\]

\[
= \lim_{i \to \infty} \left( \frac{(\ln k - (\mu - \sigma \sqrt{6}))^2}{12\sigma^2} \right) \left( \gamma + \frac{\delta}{i+1} \right)
\]

\[
= \gamma \left( \frac{(\ln k - (\mu - \sigma \sqrt{6}))^2}{12\sigma^2} \right).
\]

For the Poisson distribution we have \(\gamma = 0\) and the result is self-evident. For the negative binomial we know that \(\gamma < 1;\) therefore, the sequence \(z(i)\) converges if

\[ \left| \gamma \left( \frac{(\ln k - (\mu - \sigma \sqrt{6}))^2}{12\sigma^2} \right) \right| < 1 \iff (\ln k - (\mu - \sigma \sqrt{6}))^2 < \frac{12\sigma^2}{\gamma}. \] (6.7)
For the Poisson distribution with parameter $\lambda$ we have a rather elegant expression for the default probability. From (6.6) we obtain

$$\Pr(L_{t+dt} \geq k) \sim 1 - \sum_{i=0}^{\infty} \left( \frac{(\ln k - (\mu - \sigma \sqrt{6}))^2}{12\sigma^2} \right)^i \frac{\lambda^i}{i!} = 1 - \exp(-\lambda) \sum_{i=0}^{\infty} \frac{1}{i!} \left( \lambda \frac{(\ln k - (\mu - \sigma \sqrt{6}))^2}{12\sigma^2} \right)^i = 1 - \exp \left( -\lambda \left( 1 - \frac{(\ln k - (\mu - \sigma \sqrt{6}))^2}{12\sigma^2} \right) \right). \quad \text{(6.8)}$$

Likewise, for the negative binomial with parameters $(r, p)$, the relationship (6.6) can be further simplified to

$$\Pr(L_{t+dt} \geq k) \sim 1 - \left( \frac{p}{1 - (1 - p)(\ln k - (\mu - \sigma \sqrt{6}))^2/12\sigma^2} \right)^{-r} \left( \frac{(\ln k - (\mu - \sigma \sqrt{6}))^2}{12\sigma^2} \right) < \frac{1}{p}. \quad \text{(6.10)}$$

The last condition is in fact the convergence condition (6.7) for the infinite series, since $\gamma = p$ for the Panjer recursion.

**Case 2 (\(\mu < \ln k\))** Working with the triangular density is a little more complicated, since for $\Pr(\ln X_n < \ln k)$ we have to integrate both branches of the triangular density:

$$\Pr(\ln X_n < \ln k \mid N_{t+dt} = i) = \int_{\mu - \sigma \sqrt{6}}^{\mu} \frac{z - \mu + \sigma \sqrt{6}}{6\sigma^2} \, dz + \int_{\mu}^{\ln k} \frac{\mu + \sigma \sqrt{6} - z}{6\sigma^2} \, dz = \frac{1}{2} + \frac{1}{6\sigma^2} \left( (\mu + \sigma \sqrt{6}) \int_{\mu}^{\ln k} dz - \int_{\mu}^{\ln k} z \, dz \right) = \frac{1}{2} + \frac{(\mu + \sigma \sqrt{6})(\ln k - \mu)}{6\sigma^2} = \frac{\ln^2 k - \mu^2}{12\sigma^2} - \frac{\mu(\mu + 2\sigma \sqrt{6})}{12\sigma^2}. \quad \text{(6.9)}$$

From (6.9) and (6.3) we have

$$\Pr(L_{t+dt} \geq k) \sim 1 - \sum_{i=0}^{\infty} \left( \frac{1}{2} + \frac{(\mu + \sigma \sqrt{6})\ln k - \ln^2 k}{12\sigma^2} - \frac{\mu(\mu + 2\sigma \sqrt{6})}{12\sigma^2} \right)^i \Pr(N_{t+dt} = i). \quad \text{(6.10)}$$
Again we will need to prove convergence for the sums in (6.10). For the Panjer class of distributions it is almost self-evident that the sums converge in the $\gamma = 0$ case, which is the Poisson distribution. Likewise, for the negative binomial, the series converges with the restriction that $|\Gamma| < 1/\gamma$ (see (6.11)). As in the first case, we can further simplify the sums in (6.10) for the Poisson distribution with parameter $\lambda$ and for the negative binomial with parameters $(r, p)$. For the Poisson we obtain

$$\Pr(L_{t+\delta t} \geq k) \sim 1 - \sum_{i=0}^{\infty} \left( \frac{1}{2} + \frac{2(\mu + \sigma \sqrt{6}) \ln k - \ln^2 k}{12\sigma^2} - \frac{\mu(\mu + 2\sigma \sqrt{6})}{12\sigma^2} \right)^i \exp(-\lambda) \frac{\lambda^i}{i!}$$

$$= 1 - \exp(-\lambda(1 - \Gamma)).$$

For the negative binomial we obtain

$$\Pr(L_{t+\delta t} \geq k) \sim 1 - \left( \frac{p}{1 - (1 - p)(1 - \Gamma)} \right),$$

$$|\Gamma| < \frac{1}{p}, \quad \Gamma = \frac{1}{2} + \frac{2(\mu + \sigma \sqrt{6}) \ln k - \ln^2 k}{12\sigma^2} - \frac{\mu(\mu + 2\sigma \sqrt{6})}{12\sigma^2}. \quad (6.11)$$

7 COMPARATIVE STATICS AND VALIDATION FOR THE FREQUENCY–SEVERITY LOSS MODEL

Before we proceed to the validation of the approximation of the frequency–severity model we need to discuss further the subexponential approximation in (6.4). Formally, the subexponential property is defined as

$$\lim_{k \to \infty} \Pr \left( \sum_{n=0}^{\infty} X_n \geq k \left| N_{t+\delta t} = i \right. \right) \sim \Pr \left( \max_{n=0,\ldots,i} X_n \geq k \left| N_{t+\delta t} = i \right. \right).$$

The subexponential property of the lognormal distribution implies that the maximum value of the summands dominates the sum of lognormal variables. For example, if there is a sample with many low-value lognormally distributed random numbers and an infrequent large one, the sum of the many low-valued numbers will not exceed the infrequent large number. Since the result for the frequency–severity model hinges on the subexponential property, it is useful to show how this property manifests for different parameters.

The reason for discussing this issue is that the utilization of the closed-form solution hinges on the use of parameters that are pertinent for the subexponential property, since it is a limit property. In Figure 5 we draw $10^6$ samples from lognormal distributions with different shape parameters $\sigma$ but the same scale parameter $\mu$. For each sample

Journal of Risk Model Validation

To subscribe to a Risk Journal visit subscriptions.risk.net/journals or email info@risk.net
we calculate the sum of all $10^6$ random numbers and we report the maximum of the sample.

To aid comparison we plot using logarithmic scales. We observe that the subexponential property manifests at high values of $\sigma$, while low values of $\sigma$ do not show the subexponential property. Of course, the subexponential property is present even for low $\sigma$, but it requires more samples in order for it to become evident in the graph. Moreover, as we show in the following figures, low values of $\sigma$ require high percentiles, $k$, for the triangular approximation to work. Figures 6–9 show the quality of approximation of (6.8) and (6.10) using the Poisson distribution with parameter $\lambda$ for the frequency distribution $N_{t+df}$. We draw 30,000 scenarios for the figures.

We observe that the approximation works relatively well for changing $\lambda$ for the frequency distribution. For low values of $\lambda$ there is a deviation of about 12.8%, which is attenuated as $\lambda$ increases, because an increasing $\lambda$ indicates a higher probability of loss. Therefore, the probability of exceeding a specific value converges to 1, and converges to the value of the Monte Carlo estimator. Changing $\sigma$ makes manifest the subexponential approximation of the random sum. Low values of $\sigma$ imply greater deviation from the Monte Carlo estimator. In order to see the effect of changing percentiles $k$, in Figure 7 we plot $\Pr(Y_{t+df} > k)$ for increasing values of $k$ using the same parameters as in Figure 6.

As before, observe that, as expected, the probability decreases with increasing $k$. However, for lower percentiles the approximation error is high, and thus on the graph it would look as if the probability were increasing. This effect is a result of the closed

---

**FIGURE 5** Samples of lognormal distributions (sum and maximum) for $\mu = 15$. 

![Graph showing lognormal distributions with sum and maximum highlighted.](image-url)
Column (a) shows $P(L_{t+\Delta t} > k)$ for the Monte Carlo estimate (solid line) and the closed-form solution (dashed line). Column (b) shows the difference between the Monte Carlo estimate and the closed-form solution. First row parameters: $\mu = 15$, $\sigma = 2$, $k = 1000$. Second row parameters: $\mu = 15$, $\lambda = 2$, $k = 1000$.

form converging to the true value from below. To further show the influence of $\sigma$ and the approximation quality compared with $k$, we redraw the previous figures, increasing $\sigma$ for $k = 2$. The results are shown in Figures 8 and 9.

We observe an almost perfect fit between the closed-form solution and the Monte Carlo estimator even at low $k$ ($k = 2$) for increasing $\lambda$. This is because at high $\sigma$ the subexponential property manifests itself even at very low percentiles. Note in the second row that at $k = 2$ and $\sigma = 2$ the closed-form solution produces larger error results, but the situation improves as $\sigma$ increases. In Figure 9 we plot the probability $Pr(L_{t+\Delta t} > k)$ for the parameters of Figure 8 with increasing values of $k$.

Again, Figure 9 reinforces the validity of the results by implying that the closed-form solution has a rather good fit for increasing $k$. However, some volatility exists as $k$ increases, since we keep the number of scenarios of the Monte Carlo estimator fixed for all $k$. Overall, we show that the closed-form solution works well for lower percentiles $k$ for high values of volatility $\sigma$.

8 TAIL BEHAVIOR OF THE FREQUENCY–SEVERITY MODEL

In this section we study some tail properties of the frequency–severity model, namely VaR. From the probability $Pr(L_{t+\Delta t} \geq k)$ we can infer VaR$_a$ at $a\%$ confidence by
FIGURE 7  Probability $P(L_{t+dt} > k)$ for $\mu = 15$, $\sigma = 2$ and $\lambda = 2$.

![Figure 7](image)

FIGURE 8  Validation and comparative statics for compound Poisson losses.

![Figure 8](image)

Column (a) shows $P(L_{t+dt} > k)$ for the Monte Carlo estimate (solid line) and the closed-form solution (dashed line). Column (b) shows the difference between the Monte Carlo estimate and the closed-form solution. First row parameters: $\mu = 15$, $\sigma = 15$, $\lambda = 2$, $k = 2$. Second row parameters: $\mu = 15$, $\lambda = 2$, $k = 2$. 

www.risk.net/journals

To subscribe to a Risk Journal visit subscriptions.risk.net/journals or email info@risk.net
FIGURE 9 Probability $\Pr(L_{t+dt} > k)$ for $\mu = 15$, $\sigma = 15$ and $\lambda = 2.$

solving the equation $\Pr(L_{t+dt} \geq \text{VaR}_{1-a}) = a$. We will carry out the calculation for the model with the Poisson frequency process, while it makes sense to concentrate on the branch for which $\ln k > \mu$. From the definition of VaR we obtain

$$\Pr(L_{t+dt} \geq \text{VaR}_{1-a}) = a$$

$$\iff 1 - \exp \left( -\lambda \left( 1 - \frac{1}{2} - \frac{2(\mu + \sigma \sqrt{6}) \ln \text{VaR}_{1-a} - \ln^2 \text{VaR}_{1-a}}{12\sigma^2} + \frac{\mu(\mu + 2\sigma \sqrt{6})}{12\sigma^2} \right) \right) = a$$

$$\iff \frac{1}{2} + \frac{2(\mu + \sigma \sqrt{6}) \ln \text{VaR}_{1-a} - \ln^2 \text{VaR}_{1-a}}{12\sigma^2} = 1 + \frac{1}{\lambda} \ln(1 - a)$$

$$\iff \frac{-(\ln \text{VaR}_{1-a} - (\mu + \sigma \sqrt{6}))^2 + 6\sigma^2}{12\sigma^2} = \frac{1}{2} + \frac{1}{\lambda} \ln(1 - a)$$

$$\iff (\ln \text{VaR}_{1-a} - (\mu + \sigma \sqrt{6}))^2 = -\frac{12\sigma^2}{\lambda} \ln(1 - a)$$

$$\iff \ln \text{VaR}_{1-a} = \mu - \sigma \sqrt{6} \pm \sqrt{-\frac{12\sigma^2}{\lambda} \ln(1 - a)}$$

$$\iff \ln \text{VaR}_{1-a} = \mu + \sigma \sqrt{6} \pm \sqrt{-\frac{12\sigma^2}{\lambda} \ln(1 - a)}$$

$$\iff \text{VaR}_{1-a} = \exp \left( \mu + \sigma \sqrt{6} \pm \sqrt{-\frac{12\sigma^2}{\lambda} \ln(1 - a)} \right).$$

(8.1)
Note that the VaR has two possible solutions. If we choose the solution with the positive square root, the VaR will be a decreasing function of the frequency \( \lambda \) and an increasing function of the volatility \( \sigma \); the former is counterintuitive. We therefore pick the negative square root. The VaR in (8.1) shows that the upper percentile is the maximum value of the approximated normal exponentiated into a lognormal \( \exp(\mu + \sigma \sqrt{6}) \) minus a dampening standard deviation effect. In numerical estimations, as we will show, under this specification the VaR does not behave well. Therefore, we need to augment the triangular approximation in order to capture percentiles above and below \( \mu \pm \sigma \sqrt{6} \).

### 8.1 Improving the tail behavior of the frequency–severity model

The triangular approximation in (2.5) cuts the normal distribution at \( \sqrt{6} \) times the standard deviation above the mean. The highest value that the triangular approximation can obtain is \( \mu + \sigma \sqrt{6} \), which implies for the lognormal distribution \( \exp(\mu + \sigma \sqrt{6}) \). In order to move beyond that point, we allow the triangular distribution to vary in length, ie, instead of having upper and lower bounds \( \pm \sigma \sqrt{6} \) times the mean, we allow an arbitrary amount \( w \) plus or minus the mean of the triangular density such that it matches percentiles and the mean of the normal density. Formally, this is written as

\[
 f(x) = \begin{cases} 
 \frac{x - \mu + w}{w(c - \mu + w)} & \mu - w \leq x \leq c, \\
 \frac{\mu + w - x}{w(\mu + w - c)} & c < x \leq \mu + w, \\
 0 & \text{otherwise,}
\end{cases}
\]  

(8.2)

The number \( w \) has to be chosen so that the function \( f(x) \) is a density, ie,

\[
 \int_{\mu - w}^{\mu + w} f(x) \, dx = 1 \iff \int_{\mu - w}^{c} \frac{x - \mu + w}{w(c - \mu + w)} \, dx + \int_{c}^{\mu + w} \frac{\mu + w - x}{w(\mu + w - c)} \, dx \\
 = \frac{c - \mu + w}{2w} + \frac{\mu + w - c}{2w} = 1. 
\]  

(8.3)

In (8.3) we can choose \( c = \mu \) such that the distribution is symmetrical around \( \mu \). It suffices to show that, for \( c = \mu \), \( E[X] = \mu \) holds. To prove this we need to calculate the following integral:

\[
 E[X] = \int_{\mu - w}^{\mu} \frac{x - \mu + w}{w^2} \, dx + \int_{\mu}^{\mu + w} \frac{\mu + w - x}{w^2} \, dx \\
 = \frac{\mu^3 - (\mu - w)^3}{3w^2} - \frac{(\mu - w)(\mu^2 - (\mu - w)^2)}{2w^2} \\
 + \frac{(\mu + w)((\mu + w)^2 - \mu^2)}{2w^2} - \frac{(\mu + w)^3 - \mu^3}{3w^2} 
\]
We have matched the symmetry and the mean of the normal density. Instead of matching the variance, we want to match upper percentiles, i.e., the VaR of the normal distribution \( \ln X_n \) with mean \( \mu \) and standard deviation \( \sigma \) at some confidence level \( u \). The value of the VaR is given by

\[
\text{VaR}_{1-u}(\ln X_n) = \mu + \sigma Z_{1-u},
\]

where \( Z_{1-u} \) is the inverse distribution function for a standard normal random variable evaluated at \( 1 - u \). To match the upper percentile using the triangular approximation we need the following:

\[
\Pr(\ln X_n \geq \text{VaR}_{1-u}(\ln X_n)) = u \\
\iff \int_{\text{VaR}_{1-u}(\ln X_n)}^{\mu+w} \frac{\mu+w-x}{w^2} \, dx = \frac{(\mu+w-\text{VaR}_{1-u}(\ln X_n))^2}{2w^2} = u \\
\iff \pm(\mu+w-\text{VaR}_{1-u}(\ln X_n)) = w\sqrt{2u} \tag{8.5a}
\]

\[
\implies \text{VaR}_{1-u}(\ln X_n) = \mu + w - w\sqrt{2u}. \tag{8.5b}
\]

In (8.5a) we pick the positive absolute value since the VaR \( 1-u \) has to be lower than the upper bound, \( \mu + w \). We now equate (8.4) with (8.5b), which yields

\[
\mu + \sigma Z_{1-u} = \mu + w - w\sqrt{2u} \implies w = \frac{\sigma Z_{1-u}}{1 - \sqrt{2u}}. \tag{8.6}
\]

We have thus extended the triangular approximation more than \( \sigma \sqrt{6} \) above the mean. Instead of \( \sigma \sqrt{6} \) we have \( w \), which is given in (8.6). It suffices now to use this new triangular approximation with (6.9) and (6.10) to calculate the VaR \( 1-a \) for \( L_{t+dt} \):\(^2\)

\[
\Pr(L_{t+dt} \geq \text{VaR}_{1-a}) = \Pr\left( \sum_{n=0}^{N_{t+dt}} X_n \geq \text{VaR}_{1-a} \right)
= \sum_{i=0}^{\infty} \Pr\left( \sum_{n=0}^{N_{t+dt}} X_n \geq \text{VaR}_{1-a} \mid N_{t+dt} = i \right) \Pr(N_{t+dt} = i)
\]

\(^2\) Note that the probabilities \( u \) and \( a \) need not be equal. The probability \( u \) is used to match upper percentiles of the triangular approximation, while \( a \) pertains to the convolution of lognormal variables of the compound Poisson.
\[
\sim \sum_{i=0}^{\infty} \Pr \left( \max_{n=0,\ldots,i} X_n \geq \text{VaR}_{1-a} \Big| N_{t+dt} = i \right) \Pr(N_{t+dt} = i)
\]

\[
= 1 - \sum_{i=0}^{\infty} \Pr \left( \max_{n=0,\ldots,i} X_n < \text{VaR}_{1-a} \Big| N_{t+dt} = i \right) \Pr(N_{t+dt} = i)
\]

\[
= 1 - \sum_{i=0}^{\infty} \Pr(\ln X_n < \ln \text{VaR}_{1-a})^i \Pr(N_{t+dt} = i)
\]

\[
= 1 - \sum_{i=0}^{\infty} \left( \int_{\mu-w}^{\mu} \frac{x - \mu + w}{w^2} \, dx + \int_{\mu}^{\ln \text{VaR}_{1-a}} \frac{\mu + w - x}{w^2} \, dx \right)^i \Pr(N_{t+dt} = i)
\]

Assuming a Poisson process for \(N_{t+dt}\), with parameter \(\lambda\), we obtain an elegant expression for the probability:

\[
\Pr(L_{t+dt} > \text{VaR}_{1-a})
\]

\[
= 1 - \sum_{i=0}^{\infty} \left( \frac{1}{2} + \frac{(\mu + w) \ln \text{VaR}_{1-a} - \mu}{w^2} \right) - \frac{\ln^2 \text{VaR}_{1-a} - \mu^2}{2w^2} \exp(-\lambda) \frac{\lambda^i}{i!}
\]

\[
= 1 - \exp \left( -\lambda \left( 1 - \left( \frac{1}{2} + \frac{(\mu + w) \ln \text{VaR}_{1-a} - \mu}{w^2} \right) - \frac{\ln^2 \text{VaR}_{1-a} - \mu^2}{2w^2} \right) \right)
\]

For a confidence level \((1 - a)\) and by skipping some tedious but easy algebra, we can show that \(\ln \text{VaR}_{1-a}\) solves the following second-order equation:

\[-\frac{1}{2} (\ln \text{VaR}_{1-a} - \mu)^2 + w (\ln \text{VaR}_{1-a} - \mu) = w^2 \left( \frac{1}{2} + \frac{\ln(1-a)}{\lambda} \right).\]

This equation is solved analytically to give the exact solution

\[
\text{VaR}_{1-a} = \exp \left( \mu + w \pm \sqrt{(w + \mu)^2 + 2 \left( -\frac{\mu^2}{2} - w\mu - w^2 \left( \frac{1}{2} + \frac{\ln(1-a)}{\lambda} \right) \right)} \right)
\]

(8.7)

From (8.7) we pick the negative root, since the derivative \(\partial \text{VaR}_{1-a} / \partial \lambda\) is positive for the negative sign outside the square root. In summary, in order to augment the VaR of \(L_{t+dt}\), having used the triangular approximation (3.5) we need to move beyond \(\mu + \sigma \sqrt{6}\) for the approximation of the normal density. We achieve this by matching the mean and upper percentiles using an augmented triangular approximation (8.2).
FIGURE 10  Validation and comparative statics for 95% VaR of compound Poisson losses.

Note that the matching of the percentile VaR\(_{1-u}(\ln X_n)\) is not necessarily equal to that of VaR\(_{1-a}\). This is because we would need a rather extreme percentile VaR\(_{1-u}(\ln X_n)\) to match VaR\(_{1-a}\), since VaR\(_{1-a}\) also includes uncertainty from the Poisson counting process. We show the comparative statics of the VaR in Figure 10.

In Figure 10 the VaR in (8.1) is called the thin-tailed VaR and is represented by the dash-dotted line; the VaR in (8.7) is called the fat-tailed VaR and is represented by the dashed line; the solid line represents the Monte Carlo estimate with 30,000 scenarios. In all comparative statics it is important to find a matching percentile, \(u\), such that the triangular approximation goes far enough into the tails. The usual rule is for increasing \(\sigma\) and \(\lambda\) to match the \(u = 1/(50\lambda)\) or the \(u = 1/(50\sigma)\); variations of the rule include \(u = 1/(30\lambda)\) or \(u = 1/(30\sigma)\). In parts (a) and (b) we observe that the fat-tailed VaR using the closed-form solution is very close to, or even above, the Monte Carlo estimate for high \(\sigma\) (\(\sigma = 15\)). For low \(\sigma\) (\(\sigma = 2\)) the fat-tailed closed-form VaR overestimates the Monte Carlo estimate up to \(\lambda = 20\), and then underestimates it.

This is because for low \(\sigma\) the subexponential property does not manifest in a very pronounced way, as we saw when we calculated the probabilities. This is also shown
FIGURE 11 Validation and comparative statics for VaR_{1−a} of compound Poisson losses.

Parameters: (a) \( \mu = 15, \sigma = 2, \lambda = 2, u = 1/200 \); (b) \( \mu = 15, \sigma = 15, \lambda = 2, u = 1/200 \); (c) \( \mu = 15, \sigma = 2, \lambda = 20, u = 1/3000 \); (d) \( \mu = 15, \sigma = 15, \lambda = 20, u = 1/2200 \).

in Figure 10(c), where increasing \( \sigma \) provides a very good fit. In Figure 10(d) we show the log-scaled VaR, and we see what the matching probability \( u \) must be so that the closed-form solution matches the Monte Carlo estimate. For both \( \sigma = 15 \) and \( \lambda = 20 \) the triangular approximation has to match around the 100 \( \% \)th percentile such that the fat-tailed VaR can match the Monte Carlo estimate.

Let us reiterate what the probability \( u \) versus the probability \( a = 5\% \) means: with \( u \) we widen the triangle to match the upper percentiles of the normal distribution, so that when we convolute the exponential of this normal distribution (ie, a lognormal) it is able to give accurate VaR estimates for the probability-weighed convoluted lognormal process (ie, the compound Poisson).

In all cases the thin-tailed VaR severely underestimates the Monte Carlo estimates. This observation makes the compound Poisson model with triangular approximation of the lognormal summands within \( \pm \sigma \sqrt{6} \) around the mean better suited for calculations, such as reserving, around the mean. For risk management we must use the fat-tailed version. Finally, in Figure 11 we show the VaR under changing confidence levels, \( 1 − a \). We used 200 000 scenarios for the Monte Carlo estimates.
It is evident in Figure 11 that both low $\sigma (\sigma = 2)$ and high $\sigma (\sigma = 15)$ require similar matching probability $u$ (high percentiles) to have our closed-form solutions close to the Monte Carlo estimate. This is mostly driven by the relatively low intensity, $\lambda$, of the Poisson process. However, for large $\lambda (\lambda = 20)$ the matching probability $u$ has to be further into the tails of the normal density. However, once a high $\lambda$ has been chosen, the matching percentile $u$ is lower for $\sigma = 15$ in order for the closed-form solution to match the Monte Carlo estimate. This has to do with the fact that the subexponential property manifests more clearly for high $\sigma$.

Overall, we see that, in general, the fat-tailed estimation works rather well for high $\sigma$ and for high percentiles, which makes our method really efficient for estimating VaRs. Reducing the probability $u$ further, we could use the closed-form solutions as upper bounds for VaR or for evaluating the confidence intervals of the Monte Carlo estimates. This is because in most numerical experiments the fat-tailed closed-form solution will exceed most Monte Carlo estimations. Therefore, the formulas are particularly useful for robustness checks of the Monte Carlo estimation of the VaR at high percentiles.

As in the case of the probability estimations, the thin-tailed VaR does not perform very well for both low $\sigma (\sigma = 2)$ and high $\sigma (\sigma = 15)$. For low $\sigma (\sigma = 2)$ and $\lambda = 2$ the thin-tailed VaR is able to capture the VaR rather well from the 50% VaR (median) close to the 90% VaR. Again, this reinforces our assertion that the thin-tailed formulas work well close to the mean and are more appropriate for pricing and reserving, while the fat-tailed version is more appropriate for risk management.

9 CONCLUSION

In this paper we dealt with extracting closed-form solutions using the triangular approximation of the normal distribution for probabilities that can be used for VaR and relevant risk measurement calculations. These closed-form solutions allow fast and easy calculation for several practical cases that emerge in risk management, such as capital charges, pricing and reserving. Whenever there are normal or lognormal distributions the triangular approximation can replace them and reduce the complexity of calculations of multidimensional integrals. Instead of relying on black box simulations, the risk manager can use simple formulas that consume no computing resources.

The advantage of closed-form solutions is that they make it much easier to understand the risk drivers, and allow easy comparative statics and much easier sensitivity analysis of the results. Closed-form solutions can serve as benchmarks for more complicated problems and in many cases can serve as the input for the starting values in optimization algorithms. Moreover, closed-form solutions economize
on computational resources, especially if there are large enterprise risk management systems that require thousands of simulation runs.

Unlike Monte Carlo, where solutions are probabilistic and require a confidence interval, our solutions are “exact” in the sense that there is no uncertainty around the result and no estimation error. Therefore, there is no need for confidence intervals or nuanced sampling methods that take into account rarer events. In a CORE i5 laptop with 4 GB of RAM, the time to perform a Monte Carlo estimation of the gamma loss model is 0.6–9 seconds, depending on the number of scenarios, while the closed-form solution takes around 0.01 seconds. For the Pareto loss model the times are 1–9 seconds (depending on the number of scenarios) for the Monte Carlo estimation, and 0.0003 seconds for the closed-form solution. Finally, for the frequency–severity model the execution times are around 0.002 seconds for the closed-form solution and around 2 seconds for the Monte Carlo estimation with 30,000 scenarios.

We also performed validation checks and cautioned practitioners on the utilization of the closed-form distribution function for the collective risk model, since the result hinges on a limit theorem for the sum of lognormal random variables. Finally, this paper serves as an introduction to the risk management practice of the triangular approximation. It is hoped that further research will allow more interesting results to emerge, and that these results will make life easier for risk managers in their everyday jobs.

DECLARATION OF INTEREST

The views expressed in this paper are those of the author and do not represent the views of the Bermuda Monetary Authority. The author reports no conflicts of interest. The author alone is responsible for the content and writing of the paper.

REFERENCES


To subscribe to a Risk Journal visit subscriptions.risk.net/journals or email info@risk.net