Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 23 (2007) 513 – 529
international journal of forecasting
www.elsevier.com/locate/ijforecast

# Forecasting and analyzing insurance companies' ratings

Tony Van Gestel [a,*], David Martens [b,*], Bart Baesens [c,a], Daniel Feremans [a], Johan Huysmans [b], Jan Vanthienen [b]

[a] *Credit Risk Modelling, Group Risk Management, Dexia Group, Square Meeus 1, 1000 Brussels, Belgium*
[b] *Department of Decision Sciences and Information Management, K.U.Leuven, Naamsestraat 69, B-3000 Leuven, Belgium*
[c] *University of Southampton, School of Management, Highfield Southampton, SO17 1BJ, United Kingdom*

**Abstract**

Insurance companies sell protection to policy holders against many types of risks: property damage or loss, health and casualties, financial losses, etc. In return for this risk protection, insurance companies receive a premium from the policy holder, which is used to cover expenses and the expected risk. For longer-term risk protections, part of the premiums are invested to get higher yields. Although the protection buyer mitigates the individual risk to the large and better diversified portfolio of the insurer, it does not mean that the risk is completely reduced since the insurer may default his obligations. Insurers need to have sufficient equity or buffer capital to meet their obligations in adverse conditions when their losses on the diversified portfolio exceed the expected losses. Ratings provide an assessment of the ability of the insurer to meet its obligations to policy holders and debt holders. In this paper, the relationship between financial ratios and the rating is analyzed for different types of insurance companies using advanced statistical techniques that are able to detect non-linear relationship. The resulting rating model approach is similar to the approach for a low default portfolio, which uses a common set of explanatory variables (such as capitalization, profitability, leverage and size) which is generally applicable for all insurance types, and is complemented with insurance type specific ratios. The resulting model is found to yield a good accuracy, with 75% of the model ratings differing at most one notch from the external rating.


*JEL classification:* G22; G33; C45; C51

*Keywords:* Credit scoring; Internal rating system; Insurance companies

## 1. Introduction

The primary business of insurance companies is to provide protection against financial risks for individuals, households, professionals and companies. Life insurers offer financial protection for relatives when the insurance buyer dies, and offer different kinds of policy schemes to offer additional income upon retirement. The latter business broadens the scope of the insurer to financial intermediation, where it collects and re-invests premiums to the benefit of the insurance buyer. Property and casualty insurers, also called non-life

* Corresponding authors.
*E-mail addresses:* Tony.Vangestel@dexia.com (T. Van Gestel), David.Martens@econ.kuleuven.be (D. Martens).

0169-2070/$ - see front matter © 2007 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2007.05.001

insurers, have broadened their initial policy package of fire and accident insurance to also cover health insurance, natural disasters, liabilities, etc. Composite companies combine the services of life and non-life insurers. Reinsurance companies sell protection to insurance companies. By insuring large portfolios across different countries and sectors, a better diversification is obtained. Financial guarantors provide protection for investors on corporate and municipal bonds and structured products. At the same time, a credit enhancement is delivered to the debt issuer, providing easier access to funding at a lower cost.

The business of the insurer consists of two main activities: the underwriting of insurance premiums to cover the financial risks, and investing the premium income. In return for the payment of the insurance premium, the protection buyer is insured against a potential loss. The policy holder typically buys insurance for infrequent, but potentially high, individual losses. Whereas the individual protection buyer wants to reduce its exposure to a concentrated risk, the insurer has a large portfolio on which the risk is more diversified. The premiums received from the policies are used to cover operating expenses, pay the insured loss under the contractual terms and accumulate reserves for the future. For longer term insurance contracts, parts of the currently received premiums are invested in stocks, bonds and other financial assets to increase the revenues. Although the insurance company has a large and diverse portfolio of policies, the yearly claims are volatile and are not exactly equal to the expected loss. Therefore, the insurer needs to have sufficient buffer capital to be able to meet its obligations in adverse circumstances. National and international capital requirements (e.g., Solvency II) exist to ensure sufficiently stable and adequate insurance companies.

The risk that an insurance company will not be able to meet its obligations with respect to its policy holders, is expressed by the claims payability rating or financial strength rating. The external long term ratings are typically alpha-numerically encoded[1] and are constructed using quantitative economic factors and their interactions, as well as more judgmental aspects and future projections. The ability of the insurer to meet its obligations to the debt holders is reflected by the issuer credit rating, which is different from the financial strength rating. In practice, the difference between the two ratings is limited in most cases. Credit ratings are especially interesting for investors for investment decisions and risk management.

Rating models are mathematical decision tools that determine, in an objective way, the risk of a counterpart based upon explanatory variables derived from the financial statements and, possibly, macroeconomic and market information. Credit scoring models have received a lot of attention in the literature concerning default or bankruptcy predictions of retail customers (Baesens et al., 2003; Thomas, Edelman, & Crook, 2002) and corporates (Altman, 1968; Ohlson, 1980). The credit scoring models that are generated aim to discriminate between risky and less risky counterparts: good counterparts receive high scores, while bad counterparts receive low scores. In cases where the number of defaults is too low to develop a scoring model or where one is interested in predicting external ratings, rating models have been developed for sovereigns (Van Gestel et al., 2006a), banks (Standard & Poor's, 2004; Van Gestel et al., 2006b) and corporates (Horrigan, 1966; Rosch, 2005). Internal rating models have received a lot of attention recently. The Basel II capital accord (Basel Committee on Banking Supervision, 2004) encourages banks to develop internal rating models for accurate internal monitoring of the credit risk. When only a limited number of documented defaults is available, as is the case with insurance companies, banks can develop statistical models that explain and predict external ratings (Basel Committee on Banking Supervision, 2005, 2004) §461). Apart from the advantage of a clearly understood and internally motivated risk assessment, internal models also allow rating counterparts that are not analyzed by an external rating agency. External ratings are believed to react rather slowly to changing credit quality (Association for Financial Professionals, 2002; Laster 2003). The main reason for this persistence is believed to be the through-the-cycle methodology that is used by rating agencies: to avoid ratings fluctuating in reaction to market cycle, ratings are changed only when the changes are likely to be long lasting (Altman & Rijken, 2004). As rating models are a key decision tool, many techniques have been developed to analyze ratings. In

[1] Moody's uses Aaa (best credit), Aa1, Aa2, …, C (worst credit before default), S&P and Fitch use AAA, AA+, AA, …, C, and A. M. Best uses A++, A+, …, F.

the pioneering paper of Horrigan (1966), ordinary linear least squares regression was applied. More recently, neural networks and Support Vector Machines have been reported to provide better accuracy (Van Gestel et al., 2006b).

In this paper, a white-box internal system for rating insurance companies is developed which is both financially intuitive, easily interpretable and optimally predictive for default. For this, an incremental approach is followed so as to find a trade-off between simple, linear techniques with excellent readability but restricted model flexibility and complexity, and advanced techniques with reduced readability but extended flexibility and generalization behavior. A single insurance rating model with a common part for all types is defined, and is further refined using several other components for each (combination of) the insurance types. Much attention is paid to the financial interpretation of the statistical analysis.

This paper is organized as follows. The model formulation is described in Section 2, with a review of the statistical approach provided in Section 3. Section 4 describes the data, followed by an analysis of the rating drivers in Section 5. In Section 6, the model performance is analyzed further.

## 2. Rating model

The following approach will be used to rate an insurance company, as illustrated in Fig. 1. In the first step, a basic idea of the stand-alone creditworthiness of the insurance company is obtained from the scoring system based on quantitative information like the operating/underwriting performance, capitalization, reserves, investment yield and liquidity.

The general score is computed via linear and non-linear effects based upon the general financial variables. The type specific score is added on top, where one uses either general ratios which are particularly important for the insurer type or ratios that are specifically reported for that type of insurance company. This yields the quantitative score. On top of the quantitative score, a judgmental qualitative score is

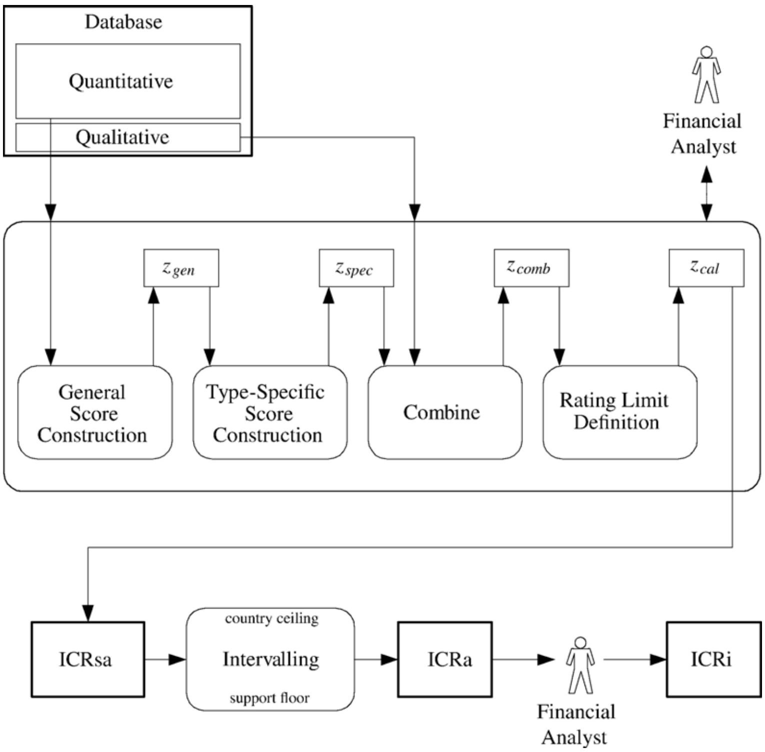


Fig. 1. The incremental process model for analyzing insurance companies' ratings. To deal with the limited data on the low default portfolio, a general model for all five insurance types is developed and further refined with a specific part for each of the (combinations of) insurance types.

added that takes into account elements like management quality and business risk. The resulting total score is reported on a 17 notch credit rating scale from AAA to CCC.

As a result, the analytical stand-alone Issue Credit Rating (ICRsa) is obtained. This ICRsa is restricted by the country ceiling[2] and mother support[3] when mapping it to the analytical Issuer Credit Rating (ICRa). Based upon the ICRa, the analysts assign the internal Issuer Credit Rating (ICRi) where it is possible to make motivated adjustments.

In this paper, the focus is on the relationships between the rating drivers and the stand-alone Long Term Rating. Therefore, the data will be cleaned from the country ceiling and support effects. The model has to satisfy several performance criteria in order to be a valid model:

(1) The model has to be stable, meaning that all coefficients in the model are well determined and the uncertainty in the estimated coefficients is sufficiently low with respect to the estimated value of the coefficients. Each variable should have a significant contribution in the model.
(2) The readability of the model is an important aspect, requiring a model to be relatively easy to interpret by the financial analysts.
(3) The model needs to accurately discriminate between solvent and non-solvent insurers. Assuming that the external rating is discriminative, the internal model rating should correspond very well to the external rating.

Each of these criteria will be addressed by the modelling techniques and verified in the statistical analysis.

[2] An insurance company cannot receive a rating higher than the rating of the country it is located in (or higher than the corresponding country ceiling limit).

[3] Because insurance companies are important companies for the economy, with typically large numbers of customers, a default of an insurance company may have quite an impact on the economy. Therefore, it is not unlikely that there will be support from mother companies (e.g., to defend their reputation) and/or from the government in case the insurance company comes into a financial distress (Patrino, Buckley, & Burke, 2001). This probability of the support is represented by the support rating, which represents a lower floor to the rating.

## 3. The incremental credit scoring model approach

### 3.1. Linear ordinal logistic regression

For binary classification problems, like bankruptcy prediction, ordinary least squares and logistic regression (Ohlson, 1980) are key techniques to build a discriminant function between two classes: class 1 (defaults) versus class 2 (non-defaults). Logistic regression is typically preferred because its model formulation is specific to a binary classification problem (defaults/non-defaults), and it is known to be more robust to deviations from multivariate Gaussian distributed classes. Furthermore, benchmark results indicate that logistic regression models exhibit better generalization behavior than least squares regression, although it has not yet been conclusively proved beyond reasonable doubt (Baesens et al., 2003; Thomas et al., 2002; Van Gestel et al., 2004). The ordinal logistic regression (OLR) model (McCullagh, 1980) is an extension of the binary logistic regression model for ordinal multi-class categorization problems, e.g., class nr. 1 (very good), class nr. 2 (good), class nr. 3 (medium), class nr. 4 (bad) and class nr. 5 (very bad). Since the results do not depend on the numerical coding of the classes, and a probabilistic interpretation is provided that indicates how sure the model is on a rating decision (as opposed to least squares regression), ordinal logistic regression is a reference technique to model external ratings.

In the cumulative OLR model, the cumulative probability of the rating $y$ is given by:

$$P(y \le i) = \frac{1}{1+\exp(-\theta_i + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n)}, \quad i = 1, \ldots, m, \tag{1}$$

with vector $x=[x_1, x_2, \ldots, x_n]^T$ containing $n$ explanatory variables $x_1$, $x_2$, …, $x_n$, and the corresponding coefficient vector $\beta=[\beta_1, \beta_2, \ldots, \beta_n]^T$. Because $P(y \le m)=1$, the parameter $\theta m$ is equal to $\infty$. The latent variable $z_L$ is the linear combination of the explanatory variables $xi$, ($i=1$, …, $n$):

$$z_L = -\beta_1 x_1 - \beta_2 x_2 - \ldots \beta_n x_n = -\beta^T x, \tag{2}$$

and summarizes the financial information of the risk entity. From the cumulative probabilities $P(y \le i)$, with

$i=1, .., m$, one obtains the rating probabilities $P(y=i)$ as

$$P(y = 1) = P(y \leq 1),$$
$$P(y = i) = P(y \leq i) - P(y \leq i - 1) \quad \textit{for} \quad 1 < i < m,$$
$$P(y = m) = 1 - P(y \leq m - 1).$$

Fig. 5c–d illustrate the rating probabilities as a function of the latent variable.

Given a training dataset $D=\{x_i, y_i\}_{i=1}^N$ of $N$ data points, the parameters $\theta_1, \theta_2, \ldots, \theta_m$ and $\beta_1, \beta_2, \ldots, \beta_n$ are estimated by maximizing the likelihood of observing the data. Apart from the optimal negative log likelihood and the optimal parameters, as a result of the optimization, the standard error of the parameters is also obtained. By comparing the ratio $\beta_i/\text{st.err.}(\beta_i)$, the statistical significance of the estimate is assessed from its $p$-value of the hypothesis test $H_0$ ($\beta_i=0$). The lower the $p$-value, the more reliable is the estimate and the more important is the rating driver $i$ to explaining the rating. Here, a $p$-value of less than 5% will be required.

The importance of a variable is also reported in terms of improvement in the model fit (deviance). The model fit follows as a result of the maximum likelihood optimization. The larger the improvement, the more important the variable is in the model. As the dataset is rather small, we feel confident with a difference of 5 or more corresponding to a strong level of confidence (Jeffreys, 1961).

As a readability constraint, it is imposed that the signs of the coefficients should be in line with the expectations of the financial analysts so as to increase the understandability of the model. Variables for which the coefficient has the opposite sign will not be selected. For example, an increase in profit margin should yield an increase in the credit rating, hence making the expected sign for this variable +. If the coefficient is negative, it will not be selected in the final model, as it is not in line with existing domain knowledge.

### 3.2. Intrinsically linear ordinal logistic regression

In the linear model, a ratio $x$ influences the latent variable $z$ in a linear way. However, it seems reasonable that a change of a ratio with 5% should not always have the same influence on the score (Box & Cox, 1964; Van Gestel et al., 2006b). Therefore, non-linear transformations of the independent variables ($x_i \mapsto f_i(x_i)$) will be considered (Box & Cox, 1964). Applying the transformation to ratios $m+1, \ldots, n$, the $z$-score (2) becomes

$$z_{\text{IL}} = -\beta_1 x_1 - \ldots - \beta_m x_m - \beta_{m+1} f_{m+1}(x_{m+1}) - \ldots - \beta_n f_n(x_n). \quad (3)$$

This model is called intrinsically linear, in the sense that after applying the non-linear transformation to the explanatory variables, a linear model is being fit (Box & Cox, 1964). A non-linear transformation of the explanatory variables is applied only when it is reasonable from both a financial and a statistical point of view, as will be illustrated in the next section.

Sigmoid transformations are typically used to floor the influence of large negative and positive values. By tailoring the shape of the sigmoid transformation, stress zones are introduced in which the rating becomes more sensitive to variable changes than for other variable values. The non-linear transformation is the hyperbolic tangent transformation

$$x \mapsto f(x) = \tanh(x \times a + b). \quad (4)$$

The hyperparameters $a$ and $b$ are estimated via a grid search mechanism (Van Gestel et al., 2004). For readability and model simplicity, a transformation is only applied when the model fit improves significantly (Jeffreys, 1961) and when it is financially intuitive. Of course, the coefficients $\beta$ have to remain stable and their signs intuitive. Examples of univariate non-linear transformations are provided in Section 5.2, further in the text.

### 3.3. Non-linear Support Vector Machines

More advanced techniques include artificial neural networks (ANN) and Support Vector Machines (SVM), which are able to capture general multivariate non-linearities in the data. In contrast to the univariate transformations (4), more complex relationships can be captured. ANNs are known to be universal approximators, meaning that they can approximate any continuous real function arbitrarily closely by a multi-layer perceptron with just one hidden layer (Bishop, 1995). However, ANNs suffer from local optima and the need for architectural design choices (number of layers, number of neurons) based on trial and error. SVMs, on the other hand, do not suffer from these disadvantages, while retaining the ability to model non-linearities (Baesens et al., 2003; Suykens, Van Gestel, De Brabanter, De Moor, & Vanderwalle, 2002; Van Gestel et al., 2004; Vapnik, 1998).

Although the SVM model may provide a high level of accuracy compared to more classical techniques, the comprehensibility of this 'black-box' model is more complex. Rule extraction is a technique to help in understanding the inside of the SVM (Martens, Baesens, Van Gestel, & Vanthienen, in press). Here, an incremental approach is followed in which the SVM terms are estimated on top of the intrinsically linear model:

$$z_{\mathrm{IL+SVM}} = \begin{array}{l} \overbrace{\overbrace{-\beta_1 x_1 - \ldots - \beta_m x_m}^{\text{linear part}} \; \overbrace{-\beta_{m+1} f(x_{m+1}) - \ldots - \beta_n f(x_n)}^{\text{non-linear transformations}}}^{\text{intrinsicaly linear part}} \\ \underbrace{-w_1 \varphi_1(x) - \ldots - w_p \varphi_p(x)}_{\text{SVM terms}} \end{array} \quad (5)$$

The remaining non-linear relationships are captured by the SVM terms $w_i \varphi_i(x)$, where $x = [x_{\mathrm{SVM},1}, \ldots, x_{\mathrm{SVM},n}]^T$ is the vector with the explanatory variables for the SVM part. The function $\varphi$ is a non-linear mapping that is often determined implicitly in terms of a RBF kernel function $K$, with parameters determined by a grid search procedure (Van Gestel et al., 2004). Note that we use Nyström sampling with a sample size of 50, so as to counter the heavy computational and memory requirements (Suykens et al., 2002).

Different approaches exist for estimating the parameters ($w$, $\beta$) from the training data. One way is to perform a joint estimation of the parameters of the (intrinsically) linear part and of the SVM terms all together. This approach has the advantage that the estimation is done in the full multivariate set-up in the space spanned by all the regressors, yielding an optimal solution. A disadvantage of this approach is that SVM terms typically have better explanatory power than the intrinsically linear terms, where the (small) improvement in model fit corresponds to a more difficult model interpretation. Here, the parameters $\beta$ are estimated first, assuming that $\boldsymbol{w} = 0$, and in a second step the $w$ parameters are optimized with $\beta$ fixed from the previous step. The variables $x$ have to be intuitive and the coefficients $w$ have to be stable. In a least squares regression set-up, this corresponds to modelling the residuals with a non-linear model.

## 4. Database description

The ISIS source was used as the main data provider for the candidate explanatory variable. The Standard & Poor's (S&P) database contains the Issuer Credit Rating (ICR) and Financial Strength Rating (FSR) of all insurance companies ranging from 1996 to 2003. Among the main rating agencies, Fitch, Moody's and S&P, the ratings of S&P were selected because it provides the broadest coverage[4] on insurance companies. The ICR is taken as the dependent variable, because of its higher availability and good representation of the credit risk. In cases where the ICR is not available, the FSR is used instead. The target variable is encoded from 1 (highest rating, AAA) to 17 (lowest rating, C or lower). The correspondence between the ICR and FSR in the provided database is very close, with a 0-notch difference of 94.0% and an average difference of 0.03 notches.

Candidate explanatory variables are chosen using the expertise of the financial analysts and based upon available literature, see for instance A.M. Best Company (2003), Laster (2003), Patrino et al. (2001). These candidate explanatory variables can be divided into the following seven categories: capital adequacy, debt and leverage, liquidity, performance, cash flow, profitability and size[5]. Some of the included ratios are applicable to all types of insurance companies, while others are specific to one particular type (or combination of types) of insurers. Sometimes a ratio is not meaningful for one type. Some variables can only be computed for a specific type because of the specificities in the financial reporting. For each ratio, several variants are included, as they provide relevant information. Based on the information from the years $t-0$, $t-1$, …, $t-4$, the variables $x_t, x_{t-1}, \ldots, x_{t-4}$ are used to define the following inputs:

(1) 5-year average (av): $x_{5y} = \dfrac{x_t + x_{t-1} + \ldots + x_{t-4}}{5}$

(2) Last available value ($t_0$): $x_t$

[4] S&P started to assign ratings to insurance companies in 1971, and nowadays analyzes more than 4000 insurance entities in more than 70 countries.

[5] A full overview of the candidate set of variables can be found at www.econ.kuleuven.be/public/ndbaf65.

(3) Relative trend (rtr): $x_{\mathrm{rtr}} = \frac{x_t - x_{t-4}}{4 \times x_{t-4}}$

(4) Absolute trend (atr): $x_{\mathrm{atr}} = \frac{x_t - x_{t-4}}{4}$

Merging the ISIS financial variable database and the S&P rating database, a database containing 2084 insurer-year observations was obtained. Amongst these observations, there are many dependent observations. Support and dependence cause hard non-linear effects (due to the support floor effect) that would drastically perturb the regression analysis between the financial variables and the rating. Because the aim is to construct a financial scoring, insurers with a too high a dependence (based upon shareholder concentration) were removed from the database. A similar effect occurs due to the country ceiling: observations with strong ratios may have a lower country rating when the country rating is lower than the intrinsic rating of the insurer. Because this effect cannot be explained by the financial ratios, insurers in weak countries with a rating equal or close to the country rating are removed from the modelling database.

The resulting database was preprocessed for the statistical analysis. Observations and ratios with too many missing values were removed from the dataset. Removed ratios include insurance debtors/surplus, intercompany investments/total assets and number of employees, among others. If possible, missing values were replaced by the value of the year before or, if this was not available, by the median. Outliers are treated by reducing them to the $3\sigma$-borders in a procedure similar to the calculation of the Winsorized mean (Van der Vaart, 1998). Size variables are transformed via the log transformation. The resulting database consists of about 4 to 5 years of information on 149 insurance companies, which corresponds to a total of 436 insurer-year entries. The number of observations per rating and type is reported in Fig. 2.

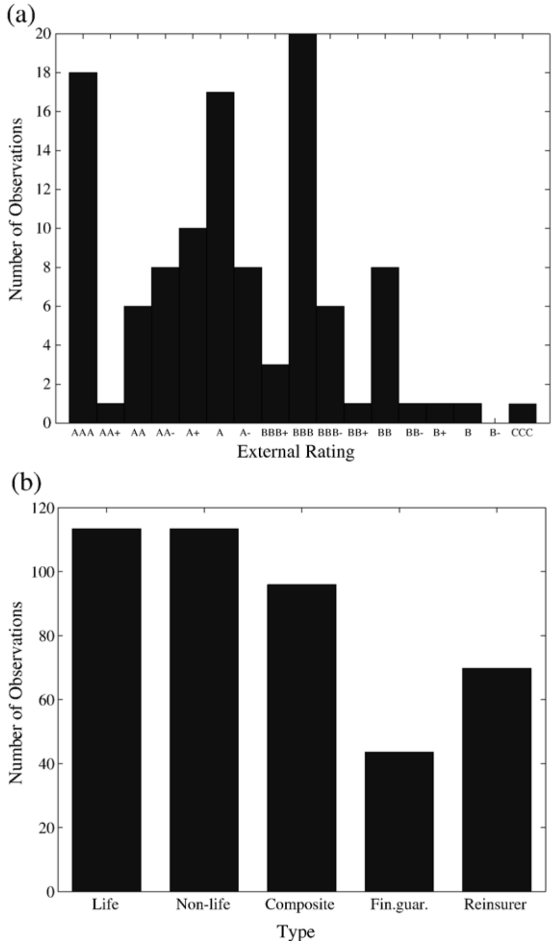


Fig. 2. Number of rating observations per rating category (a) and per type (b).

## 5. Analysis of rating drivers

### *5.1. Ordinal logistic regression model*

The database of candidate explanatory variables was analyzed by testing multiple ratio combinations in forward, backward and stepwise variable selection. These results were then discussed with the financial analysts. Surprisingly, the performance of these trial models was very low, the models did not succeed to have a good performance on the different types of insurance companies. Hence, the construction of a one-fits-all model was not possible.

Instead of developing different models for the 5 insurer types, it was decided to develop a common model, but with specific terms for each insurer type (or combinations). The resulting formulation for the linear model is the following

$$
\begin{aligned}
z_{\mathrm{L}} = & -\beta_1 x_1 - \beta_2 x_2 - \beta_3 x_3 - I(\mathrm{life})\beta_4 x_4 \\
& - I(\mathrm{non-life})\beta_5 x_5 - I(\mathrm{reins.})\beta_6 x_6 \\
& - I(\mathrm{fin.guar.})\beta_7 x_7 - I(\mathrm{comp.})\beta_8 x_8, \qquad (6)
\end{aligned}
$$

where for the sake of notational simplicity only 1 term is denoted per type. The indicator function $I(t)$ is equal to one for type $t$ and 0 otherwise. The advantage of developing such a model is that the number of parameters that need to be estimated remains low, because of the common part $\beta_1$, $\beta_2$, $\beta_3$ in Eq. (6) and because only one set of $\theta$ parameters has to be estimated instead of 5. Given that there are 17 possible ratings requiring the estimation of 16 $\theta$ parameters $\theta_1$, …, $\theta_{16}$, this is an important reduction in the number of parameters that need to be estimated.

Using stepwise regression, the selected variables are reported in Table 1. The in-sample performance is equal to 48% on 0 notch, 70% of the model ratings are within 1 notch of the external rating, and 83% of the model ratings are within two notches (0–2 notches difference) of the external rating.

Table 1

Selected variables for the linear model, the intrinsically linear model, and the intrinsically linear model with SVM terms

| Variable | Insurance type | Variable type | Variant | Sign | Linear | | | Intrinsically linear | | | SVM |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | I/O | $p$-value (%) | ΔDev | I/O | $p$-value (%) | ΔDev | I/O |
| Net investment income/net premium written | All | Performance | av | + | L | 0.4 | 10.00 | NL | 0.0 | 32.56 | SVM |
| Expense ratio | All | Performance | av | − | L | 0.1 | 12.13 | 0 | 0.7 | 7.45 | 0 |
| Elementary Capital Adequacy Ratio | All | Capital Adequacy | av | + | L | 0.1 | 12.02 | 0 | 0.0 | 23.12 | 0 |
| Net technical reserves/Net premium written | All | Capital Adequacy | atr | + | 0 | | | 0 | 1.2 | 6.34 | SVM |
| Profit before tax/Net premium written | All | Profitability | av | + | L | 0.0 | 22.28 | NL | 0.0 | 17.76 | 0 |
| Liquid A/Illiquid A | All | Liquidity | av | + | L | 0.5 | 8.04 | 0 | 0.0 | 14.76 | SVM |
| Debt/Equity | All | Debt & Leverage | av | − | L | 0.0 | 68.63 | 0 | 0.1 | 12.24 | SVM |
| Gross premium written | All | Size | av | + | L | 0.0 | 23.00 | 0 | 0.0 | 266.49 | 0 |
| Listed | All | Other | $t_0$ | − | L | 0.0 | 16.23 | 0 | 0.0 | 44.70 | 0 |
| Gross premium written/surplus | L+NL+F+R | Capital Adequacy | av | − | L | 0.1 | 10.75 | 0 | 0.0 | 27.43 | SVM |
| Investment yield | L+F+R | Performance | av | + | L | 0.1 | 11.27 | NL | 0.0 | 103.86 | SVM |
| S&P's Loss ratio | F | Performance | av | − | L | 0.2 | 10.03 | 0 | 0.0 | 25.79 | 0 |
| Net premium earned/Gross premium written | F | Performance | $t_0$ | + | L | 0.0 | 19.35 | 0 | 0.0 | 24.71 | 0 |
| Combined ratio | F | Performance | av | − | L | 0.0 | 16.56 | 0 | 0.0 | 40.86 | 0 |
| Investment yield | F | Performance | atr | + | L | 0.0 | 62.33 | 0 | 0.2 | 10.39 | 0 |
| Net unpaid losses/net claims | NL | Performance | av | − | L | 0.4 | 8.21 | NL | 0.0 | 12.74 | 0 |
| Combined ratio | NL | Performance | av | − | L | 0.1 | 10.67 | 0 | 0.0 | 15.72 | 0 |
| Net claims/surplus | NL | Performance | $t_0$ | − | L | 1.5 | 5.85 | 0 | 0.0 | 15.73 | 0 |
| Total cash flow ratio | NL | Performance | av | + | L | 0.0 | 13.42 | NL | 0.0 | 27.01 | 0 |
| Dummy life | L | Other | $t_0$ | − | L | 1.7 | 5.73 | 0 | 0.0 | 98.42 | SVM |
| Profit margin | L | Profitability | $t_0$ | + | L | 0.0 | 33.38 | 0 | 0.0 | 13.29 | 0 |
| Return on Average Assets | L | Profitability | rtr | + | L | 1.7 | 5.74 | NL | 0.1 | 11.64 | 0 |
| Liquid A/Total A | L | Liquidity | rtr | + | L | 0.0 | 39.36 | 0 | 0.2 | 9.76 | 0 |
| Debt/Gross premium written | L | Debt & Leverage | atr | − | L | 0.0 | 1.7 | NL | 0.0 | 19.06 | 0 |
| Elementary Capital Adequacy Ratio | R | Capital Adequacy | atr | + | L | 0.0 | 21.05 | 0 | 0.5 | 8.01 | 0 |
| Net unpaid losses/net technical reserves | R | Capital Adequacy | atr | − | L | 0.1 | 10.04 | 0 | 3.8 | 4.39 | 0 |
| Percentage reinsurance | F | Other | $t_0$ | + | L | 0.0 | 119.64 | 0 | 0.0 | 77.22 | SVM |

The following variables are used in the general score that is applied to for all 5 types of insurance companies:

Elementary capital adequacy: The capital adequacy ratio compares the capital buffer (surplus) to the risks the insurer has taken. These risks originate from the underwritten policies and the investments. Capital adequacy is a very important rating driver when sufficient details are available (Puccia, Osborne, & Anthony, 2005).

Debt to capital: The higher the debt to capital ratio, the more sensitive the company is to profit volatility and the more risky the counterpart is.

Profit before tax/net premiums written: This ratio reports the ability to generate profit. The before tax profit allows better comparison across countries with possibly different tax schemes. A non-linear transformation of this ratio is applied to penalize insurers with a gross profitability of less than 5%. Higher profitabilities are considered not to add value to the creditworthiness.

Gross premiums written: As for corporates and banks, size is an important rating driver. The bigger the insurance company, the broader the customer base, and scale-effects have a positive impact. While total assets is typically chosen as the size factor for banks and corporates, for insurance companies the gross premiums written was selected.

Liquid assets/illiquid assets: The liquidity reflects the ability to absorb short term changes in claims, liabilities, etc., and is measured by the ratio of liquid assets to illiquid assets.

Net investment income/net premiums written: The ratio of net investment income before taxes compared to net premiums written gives an idea of the success of the investment strategy. In the intrinsically linear model, values below 25% are considered very bad, while values above 60% are very good.

Expense ratio: The efficiency of the insurer's underwriting business is measured by the expense ratio (costs to acquire and manage policies/ net premiums written). The worse the ratio, the higher the expenses to acquire the premium income.

Net technical reserves/net premiums written: Net technical reserves compared to net premiums written measures the amount of reserves in relation to net underwriting risks. Insufficient reserves for prior accident years might result in reserves strengthening requirements that will weight on future profitability.

Flag listed/non-listed: The qualitative variable indicates whether the company is stock-listed or not. The indicator indicates that there is a difference in the perception of the risk of a stock-listed company and a non-stock-listed.

For life companies, the following variables are found to be important:

Investment yield: Life, non-life and other insurance companies all collect insurance premiums and will invest a portion of it in the financial markets, while seeking to obtain a profit in excess of the claims that have to be paid. Life insurers are able to calculate future claim demands more precisely than non-life insurers, by using statistical mortality tables. Hence, they can invest a larger portion of the premium income in long-term investments. An indicator that measures the company's investment performance, i.c. investment yield, is well able to distinguish between better and poorer performing life insurance companies.

Liquid assets/total assets: In line with the investment yield, the liquid assets/total assets ratio indicates the portion of investments that is invested in liquid assets, which can easily be converted into cash when claims emerge.

Profit margin: This is a profitability measure that compares the profit before taxes to turnover (gross premium written). Although such ratios are generally important, this variable was selected for life only as a specific emphasis on comparing profitability amongst life insurers.

Debt/gross premiums written: This ratio compares the debt level to the gross turnover.

Return on average assets: This ratio gives an indication of the profitability of the total balance sheet assets that are used to generate premium and investment income. In contrast to the return on equity, this ratio is not influenced by changes in leverage.

The dummy-variable for life adjusts for the difference in the (weighted) mean, and should not be considered as an explanatory variable. The non-life specific variables are:

Gross premiums written/surplus: For the non-life insurance company, for which the future path of claim demands is much more uncertain, this ratio is intended to give a first indication of the extent to which risk (as presented by premium income) is covered by capital (surplus).

Net claims/surplus: This measure is similar to the above ratio. The risk is measured (up to a proportional scaling) by the realized net claims demands (after reinsurance), which is again compared to the surplus.

Combined ratio: For a better understanding, the combined ratio has to be decomposed into its building blocks:

The expense ratio is a measure of the efficiency of the insurer's underwriting business. It compares the net premium income, after reinsurance, with the costs that have been incurred to acquire and manage the policies.

The loss ratio is a measure of the insurer's risk picking ability. It compares net premiums earned, i.e. after reinsurance and after taking into account changes in the net unearned premiums reserves, to net claim demands.

The combined ratio is a key summary indicator which distinguishes between more and less efficient non-life insurance companies. It was not selected for life companies because the loss ratio part is less discriminant because the risk picking is less straightforward.

Net unpaid losses/net claims: This ratio measures the amount of losses that still need to be repaid (in comparison to the net claims). When this ratio is high, it means that there is a high short term debt to the policy holders.

Total cash flow ratio: The underwriting cash flow (net premiums written − underwriting expenses) + net investment income − taxes is compared with net premiums written. This ratio compares the main cash flow items in the analysis and was found to be discriminant for non-life companies.

For composite companies, no specific variables were found to be significant.

Given that reinsurance companies are a special category that are active in either life or non-life insurance (or both), it should not come as a surprise that some of the variables that are relevant for life and non-life insurers, such as investment yield and gross premiums written/surplus, will also prove to be explanatory for this type of insurer. The capital adequacy ratio, however, deserves further attention: while this ratio is already included in the common part, it is given a higher weight for reinsurance companies; hence the reason why it is also included in the type specific part. While reinsurers themselves can also make an appeal on reinsurance, referred to as retrocession, it might be arguable that this is more costly than primary reinsurance, and that as a result, reinsurers will generally retain a larger portion of business written, making them riskier, especially to risk correlation and concentration. From this arises the need to have a capital base that is sufficient to cover the potential losses. In addition, the change in net unpaid losses/net technical reserves was found to be important in capturing changes in the proportion of 'real' reserves. Although net unpaid losses are part of the technical reserves according to the accountancy, an increase of this figure with a constant net technical reserves may indicate a reduction of the 'real' reserves of the reinsurer.

Financial guarantors are a specific type of insurance companies. Therefore, the specific scorecard is proportionally the most important. Among the specific ratios are the investment yield and gross premium written/surplus. Specific for financial guarantors alone are:

Combined ratio: Although this ratio was also selected for non-life companies, the coefficient for financial guarantors is larger.

Loss ratio: This ratio is a measure of the cost of risk (net claims/net premiums earned). Because financial guarantors insure against financial losses, the selection process of which assets to insure and for which cost, is a primary indicator of the success of the company.

Net premium earned/gross premium written: The most recent value of this indicator compares the net premiums earned (total premiums earned − reinsurance − cost of losses incurred) to the gross premiums written. It measures the current net premium income, corrected for reinsurance and losses.

Investment yield: Financial guarantors also have relatively reliable default and loss statistics to determine their capital requirements, making investment yield an important ratio.

Reinsurance: The percentage reinsurance is an indication of the risk mitigation.

The incremental approach leads to the models described in Table 1, which shows the relevant variables; the insurance type, which indicates for each of the variables whether it takes part in the common part of the model, or is relevant for particular insurance parts only; the variable type; the input variant (average, most recent value, absolute or relative trend); and their expected sign. Columns 6 and 11 indicate the (non) selected variables for the linear (L when selected) and intrinsically linear model (NL when selected) respectively. Also included is the *p*-value and deviance difference when adding/removing each one of the variables. The final column reports the ratios used in the SVM terms, denoted by SVM when the variable is selected.

### 5.2. Intrinsically linear ordinal logistic regression model

For each ratio $x_i$ the optimal non-linear transformation $f(x_i; a_i, b_i)$ is potentially introduced in to the model. Sigmoid transformations are applied to all trend variables (both relative and absolute), once again assuring that transformations are only introduced when the model fit is significantly improved thereby. After each non-linear variable transformation, the validity of the current input set is assessed. When applicable, the set is updated by removing and/or adding variables. At the end of this process, non-linear transformations on additional non-selected variables are also considered.

This results in the non-linear transformations of the following inputs: Investment yield (av), Profit before tax/Net premium written (av), Debt/Gross premium written (atr), Total cash flow ratio (av), Net investment income/net premium written (av), Profit margin ($t_0$), Return on average assets (rtr) and Net unpaid losses/net claims ($t_0$). We discuss, for illustrative purposes, four important non-linearities, as also highlighted in Fig. 3.

The transformation of investment yield is a sigmoid transformation with cut-off values at 0 and 5%; values more than 5% do not attribute to a better rating because, despite the average, it may indicate a higher investment risk. The total cash flow ratio compares the underwriting cash flow and net investment income minus taxes to the net premiums written. When the value is above 100%, the variable is intuitively considered very good, and statistically a saturation occurs: a further improvement does not yield a rating improvement. Decreasing from 100%, and especially 90%, a lower ratio increasingly penalizes the creditworthiness. Changes in debt as a percentage of gross premium written become important when the annual increase is, on a 5-year average, more than 2%. From values more than 10% onwards, no additional penalization is applied. It is interesting to observe that values below 0% are not additionally beneficial than an increase of 1%. The profit before tax/net premium written starts to penalize counterparts when the ratio has values lower than 5–10% to values below −10%.

Testing for the stability of the estimated coefficients, no ratios of the linear model dropped out. Testing for the addition of other variables, the ratio net technical reserves/Net premium written (atr) was found to be additionally significant from a statistical perspective. Given that net technical reserves is an important indicator for an insurance company, the trend is a financially meaningful indicator, and it was decided to use this ratio in the model.

Details of the model obtained are reported in Table 1. The in-sample accuracy improves to 52%, 75% and 86% on 0, 0–1 and 0–2 notches difference, respectively. The performance improvement is observed for all insurer types.

### 5.3. SVM terms

Given the intrinsically linear model, the non-linear model is constructed by applying a partial regression on top of the estimated intrinsically linear model, yielding the latent variable $z_{\mathrm{IL+SVM}}$. The selected SVM inputs are investment yield, gross premium

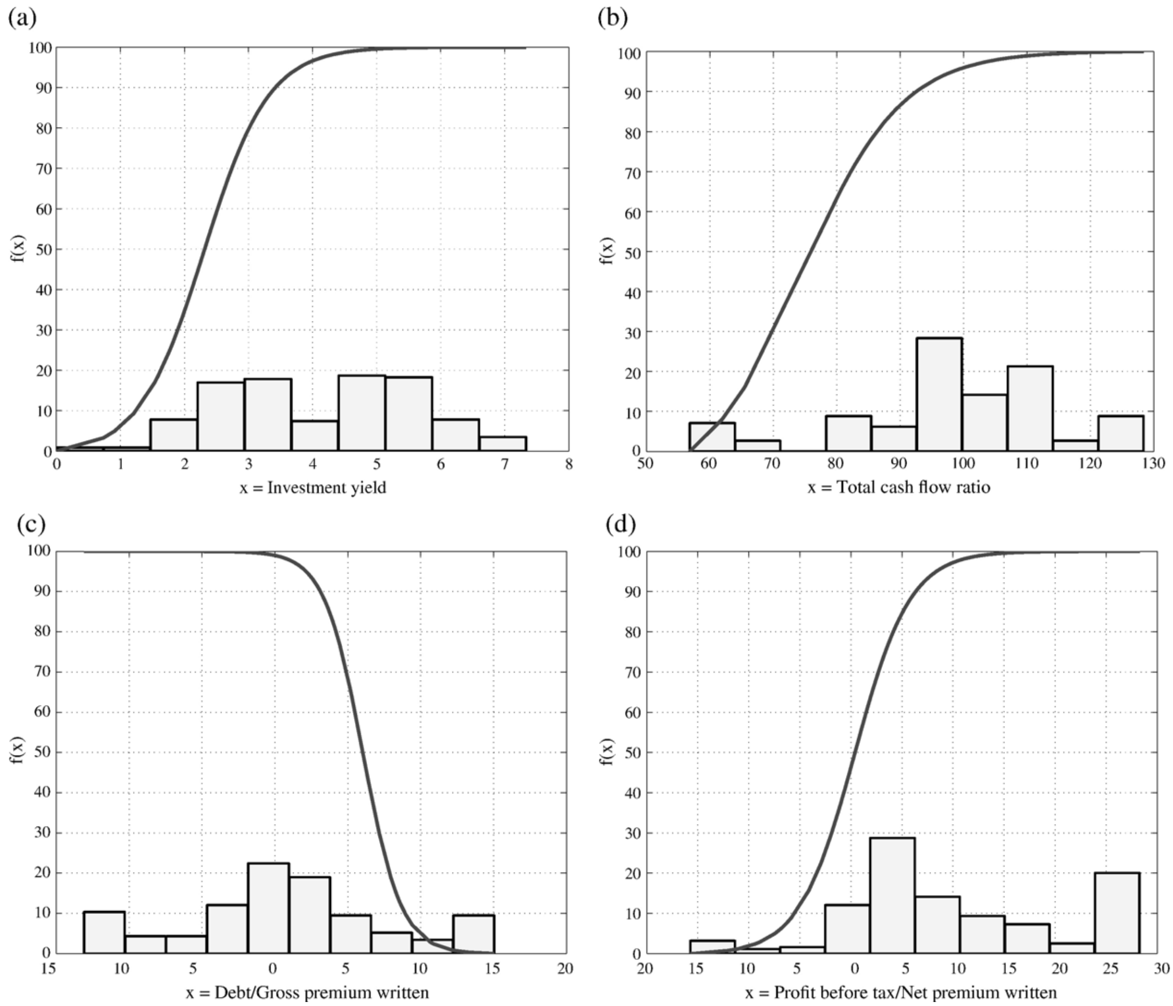


Fig. 3. Visualisation of the univariate non-linear transformations applied to variables in the intrinsically linear model. The histogram of the ratio/variable distribution is reported by the histograms, as denoted by the white bars.

written/surplus, debt/equity, liquid assets/illiquid assets, dummy life, percentage reinsurance, net investment income/net premium written and net technical reserves/net premium written.

The addition of SVM terms allows us to increase the performance on 0 notch to 53%, to 76% on 0–1 and to 88% on 0–2 notches difference, as shown in Fig. 4.

## 5.4. Qualitative variables

The developed score function $z_{\mathrm{IL+SVM}}$ only takes into account financial variables. However, it is plausible to assume that qualitative indicators like quality of management, regulatory supervision, risk type and profile and quality of risk management also impact the rating (Grunert, Norden, & Weber, 2005).

For the sake of completeness, it is mentioned that a second, judgmental database was constructed. It consists of 110 qualitative scorecards of insurance companies with the variables market position, risk profile, financial risk profile, and disclosure quality. Each variable was represented on a 5 value scale: very strong, strong, medium (neutral), weak and very weak. Based upon the analysis of the difference between the model rating and the external ratings, a composite

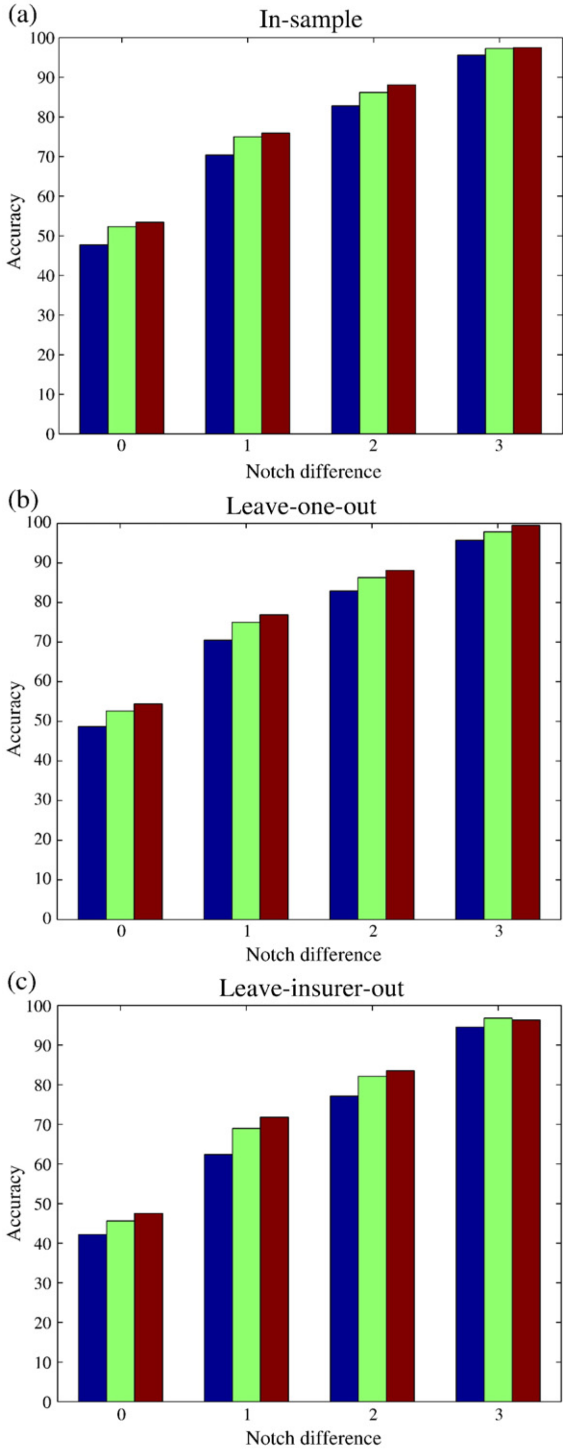


Fig. 4. In-sample (a), leave-one-out (b) and leave-insurer-out (c) performances of the linear (■), intrinsically linear (■), and intrinsically linear model with SVM terms (■).

qualitative indicator was defined using these four judgmental indicators.

The combination of the judgmental variables with the quantitative model is not straightforward, as the sample sizes are different. By means of partial regression, the $\beta$- and $\theta$-parameters of the quantitative model are fixed, and on top of that, the coefficient $\beta$ of the qualitative scorecard is obtained via maximizing the likelihood of the observations in the qualitative database. The estimated coefficient of the qualitative scorecard has a $p$-value of 0.01%. The latter indicates that the estimated coefficient is highly stable.

## 6. Model evaluation and calibration

### 6.1. Model evaluation

As the external ratings are ordinal measures of risk, no metric assumption (e.g., the long term default rate) has been assigned. This lack of a clear metric makes it difficult to determine the misclassification cost: the rating determines not only the default risk, but also various other aspects like pricing and liquidity. The most common way of reporting the model performance is to report the confusion matrix and ordinal correlation coefficients. In this case, the gamma-coefficient for ordinal correlation (Goodman & Kruskal, 1972) is reported in Table 3. Intuitive summary tables, in terms of notch differences are extracted from the large confusion matrix and are reported in Table 2.

Due to the limited size of the dataset, the model accuracy performance will be determined using leave-one-out and leave-insurer-out cross-validation, which approximate the out-of-sample performance. The model parameters are estimated on the full dataset, except for one insurer-year observation that is used for validation. This is repeated 436 times, such that each insurer-year observation has been used exactly once as a validation observation. The final discrimination performance is the average over the 436 validation performances. This approach has the disadvantage that some part of the insurer information is already included in the training dataset. Therefore, we also

Table 2
Average in-sample, leave-one-out and leave-insurer-out performances for the linear model (L), the intrinsically linear model (IL), and the intrinsically linear model with SVM terms (IL+SVM)

| Model | Insurance type | Nb | In-sample | | | | Leave-one-out | | | | Leave-insurer-out | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 0 notch (%) | 0–1 notch (%) | 0–2 notch (%) | 0–3 notch (%) | 0 notch (%) | 0–1 notch (%) | 0–2 notch (%) | 0–3 notch (%) | 0 notch (%) | 0–1 notch (%) | 0–2 notch (%) | 0–3 notch (%) |
| L | All | 436 | 47.7 | 70.4 | 82.8 | 95.6 | 45.6 | 67.2 | 81.2 | 95.2 | 42.2 | 62.4 | 77.1 | 94.5 |
| | Life | 113 | 37.2 | 64.6 | 73.5 | 93.8 | 33.6 | 58.4 | 69.9 | 92.0 | 29.2 | 52.2 | 62.0 | 89.4 |
| | Non-Life | 113 | 39.8 | 65.5 | 84.1 | 95.6 | 38.1 | 63.7 | 82.3 | 95.6 | 35.4 | 62.0 | 80.5 | 95.6 |
| | Composite | 96 | 57.3 | 68.8 | 78.1 | 95.8 | 57.3 | 66.7 | 77.1 | 95.8 | 54.2 | 61.5 | 74.0 | 95.8 |
| | Fin. Guar. | 45 | 93.3 | 97.8 | 100.0 | 100.0 | 86.7 | 93.3 | 100.0 | 100.0 | 80.0 | 84.4 | 95.6 | 100.0 |
| | Reinsurer | 96 | 34.8 | 72.5 | 91.3 | 95.7 | 34.8 | 71.0 | 91.3 | 95.7 | 33.3 | 66.7 | 88.4 | 95.7 |
| IL | All | 436 | 52.3 | 75.0 | 86.2 | 97.3 | 50.2 | 72.5 | 84.9 | 96.8 | 45.6 | 69.0 | 82.1 | 96.8 |
| | Life | 113 | 40.7 | 5.0 | 76.1 | 95.6 | 39.8 | 62.8 | 75.2 | 94.7 | 34.5 | 58.4 | 70.8 | 94.7 |
| | Non-Life | 113 | 45.1 | 72.6 | 85.0 | 95.6 | 43.4 | 70.8 | 93.2 | 95.6 | 36.3 | 67.3 | 80.5 | 95.6 |
| | Composite | 96 | 61.5 | 75.0 | 87.5 | 99.0 | 61.5 | 72.9 | 86.5 | 99.0 | 58.3 | 68.8 | 82.3 | 99.0 |
| | Fin. Guar. | 45 | 97.8 | 100.0 | 100.0 | 100.0 | 88.9 | 91.1 | 97.8 | 100.0 | 84.4 | 86.7 | 97.8 | 100.0 |
| | Reinsurer | 96 | 40.6 | 78.3 | 94.2 | 98.6 | 37.7 | 78.3 | 92.8 | 97.1 | 36.2 | 78.3 | 92.5 | 97.1 |
| IL+ SVM | All | 436 | 53.4 | 75.9 | 88.1 | 97.5 | 51.6 | 74.1 | 85.8 | 97.3 | 47.5 | 71.8 | 83.5 | 96.3 |
| | Life | 113 | 44.3 | 66.4 | 81.4 | 95.6 | 43.4 | 62.8 | 77.9 | 95.6 | 37.2 | 59.3 | 73.5 | 95.6 |
| | Non-Life | 113 | 46.0 | 73.5 | 85.8 | 96.5 | 43.4 | 71.7 | 83.2 | 95.6 | 39.8 | 68.1 | 80.5 | 92.9 |
| | Composite | 96 | 62.5 | 79.2 | 88.5 | 99.0 | 61.5 | 78.1 | 86.5 | 99.0 | 59.4 | 78.1 | 86.5 | 99.0 |
| | Fin. Guar. | 45 | 93.3 | 95.6 | 100.0 | 100.0 | 88.9 | 93.3 | 97.8 | 100.0 | 84.4 | 86.7 | 97.8 | 100.0 |
| | Reinsurer | 96 | 42.0 | 78.3 | 94.2 | 98.6 | 40.6 | 78.3 | 94.2 | 98.6 | 36.2 | 79.7 | 91.3 | 97.1 |

The performance measures provided are the 0, 0–1, 0–2 and 0–3 notch differences, differentiated over the insurance types. Also provided is the number of data instances (Nb) for each type.

Table 3
Out-of-sample backtesting results for the implemented model *Model* (intrinsically linear with SVM terms), the model complemented with analyst judgment and potential overruling *Dexia*, and external rating models *Moody's*, *S&P* and *Fitch*, for data from the period 2005 to 2006

| Rating comparison | Nb | Absolute cumulative difference | | | Notch difference | | | | | | | Correlation | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0 (%) | 0–1 (%) | 0–2 (%) | ≤−3 (%) | −2 (%) | −1 (%) | 0 (%) | 1 (%) | 2 (%) | ≥3 (%) | Gamma | *p*-value (%) |
| Dexia — Model | 128 | 51 | 80 | 96 | 2 | 3 | 4 | 51 | 25 | 13 | 2 | 0.92 | 0 |
| Dexia — Moody's | 64 | 66 | 89 | 98 | 2 | 9 | 17 | 66 | 6 | 0 | 0 | 0.95 | 0 |
| Dexia — S&P | 89 | 57 | 91 | 97 | 3 | 6 | 11 | 57 | 22 | 0 | 0 | 0.94 | 0 |
| Dexia — Fitch | 67 | 49 | 91 | 99 | 1 | 4 | 7 | 49 | 34 | 3 | 0 | 0.93 | 0 |
| Dexia — External Average | 91 | 70 | 92 | 97 | 3 | 4 | 9 | 70 | 13 | 0 | 0 | 0.96 | 0 |
| Model — Moody's | 63 | 41 | 70 | 87 | 13 | 16 | 22 | 41 | 6 | 2 | 0 | 0.78 | 0 |
| Model — S&P | 88 | 43 | 76 | 86 | 14 | 9 | 22 | 43 | 11 | 1 | 0 | 0.82 | 0 |
| Model — Fitch | 66 | 36 | 71 | 91 | 9 | 14 | 14 | 36 | 21 | 6 | 0 | 0.75 | 0 |
| Model — External Average | 90 | 43 | 74 | 89 | 11 | 13 | 22 | 43 | 9 | 1 | 0 | 0.83 | 0 |
| Moody's — S&P | 64 | 63 | 91 | 95 | 0 | 2 | 9 | 63 | 19 | 3 | 5 | 0.91 | 0 |
| Moody's — Fitch | 59 | 39 | 88 | 98 | 0 | 2 | 3 | 39 | 46 | 8 | 2 | 0.93 | 0 |
| Moody's — External Average | 65 | 74 | 95 | 100 | 0 | 0 | 5 | 74 | 17 | 5 | 0 | 0.96 | 0 |
| S&P — Fitch | 66 | 50 | 88 | 100 | 0 | 2 | 12 | 50 | 26 | 11 | 0 | 0.92 | 0 |
| S&P — External Average | 90 | 81 | 100 | 100 | 0 | 0 | 11 | 81 | 8 | 0 | 0 | 0.99 | 0 |
| Fitch — External Average | 68 | 57 | 99 | 100 | 0 | 1 | 37 | 57 | 4 | 0 | 0 | 0.98 | 0 |

The notch differences and their cumulated values are provided, as is the Gamma-coefficient for ordinal correlation. As the absolute cumulative notch differences for over 3 notches all are 100%, these values are omitted.

report the cross-validation performance where, in turn, we leave out all insurer-year combinations relating to the same insurance company, and use them for validation. This basically provides the model accuracy performance for insurance companies on which the model had not yet been trained. The Basel II capital accord explicitly recognizes the requirement for out-of-sample and out-of-time testing (Basel Committee on Banking Supervision, 2004, §420). The performances of these models are reported using in-sample, leave-one-out and leave-insurer-out measurements in Fig. 4. More detailed accuracies for each of the insurance types, as well as the number of instances for each type (Nb), are provided in Table 2.

Basel II requires that the performance of the models be backtested on an annual basis, on both the out-of-sample and out-of-time observations of the past years. As these results are already available for the period 2005 and 2006 for the implemented model, we are able to complement the in-sample and cross-validation results with these out-of-sample backtesting results, shown in Table 3. Note that the Dexia ratings are the ratings after analyst judgment and potential overruling. From this table, we can conclude that our model's predicted ratings (denoted by Model) coincide well with the external ratings, shown by the high Gamma values with low $p$-values, and the high percentages for the notch differences. We can also observe that the results between our model ratings and the external ratings are comparable to those between the external ratings themselves. As we used the S&P ratings to train our model, it is not surprising that the model fits best

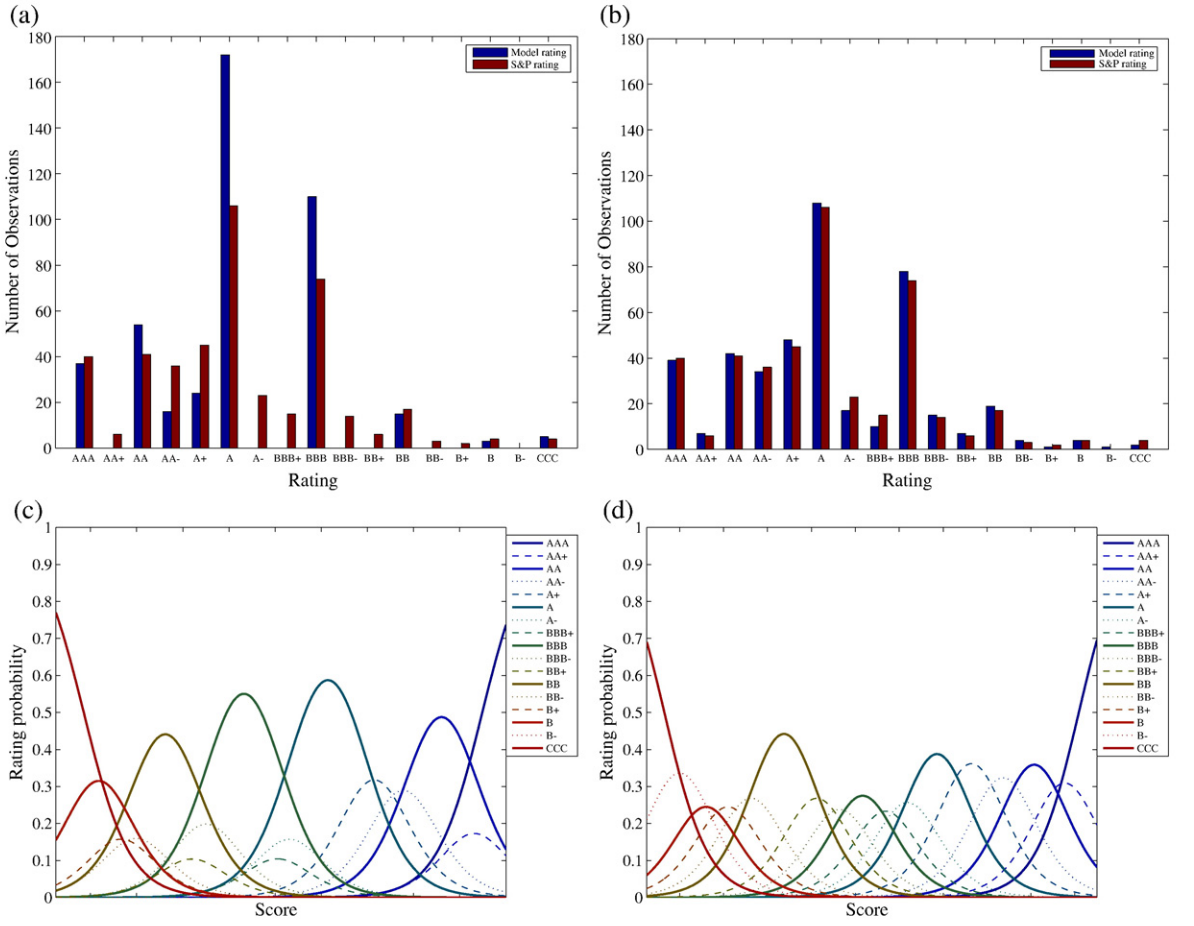


Fig. 5. Histograms and probabilities of predicted ratings, before (a, c) and after (b, d) the model calibration by modifying $\vartheta$.

with the S&P external ratings. Also note the high conformity amongst the ratings of the external rating agencies.

*6.2. Rating calibration*

Given the OLR model with score parameters $w$ and $b$ and class probability parameters $\theta$, the output of the model is the different rating probabilities $P$ (rating = rating$_i|z$) for $i=1,\ldots,n$. The decoding problem is to assign the actual class label class$_i$. The most popular approach is to assign the rating with the highest probability

$$\text{rating}_i = \text{argmax}_{i=1,\ldots,n} P(\text{rating} = \text{rating}_i|z).$$

A disadvantage of this approach is that a rating $i$ that does not often occur in the database may never have the highest probability, as shown by Fig. 5(c). The resulting model rating distribution indicates too high a number of predicted A and BBB ratings, which can be explained by the fact that S&P considers + and − modifiers as transition ratings. Given that the discrimination in the A and BBB zone is not perfect, the optimization problem of the ordinal logistic model is such that it will favor the assignment of a A or BBB rating instead of a A− or BBB+. Although this may be statistically optimal, such rating jumps may not be either practical or intuitive from a financial perspective. One possible solution would involve reducing the number of training observations in the A and BBB range. As this would involve reducing the size of an already small dataset, we will opt for another solution.

The decoding scheme can be changed in a number of ways. A first way is to change the parameters $\vartheta = [\theta_1, \theta_2, \ldots \theta_m]$ such that the rating changes continuously (with a maximum one notch distance per rating change) as a function of the $z$-score. Alternatively, one could also define upper and lower limits of the $z$-score for each rating class. Since the first option has the advantage that one can also report rating probabilities, as shown in Fig. 5(d), which provide additional information to the financial analysts on how sure the model is of the assigned rating, it is decided here to change the parameters $\vartheta$. The parameters $\vartheta$ are changed such that a continuous change of the rating as a function of the $z$-score is obtained and that the histogram of the predicted ratings closely matches the histogram of the training database, as shown in Fig. 5(b). Of course, this also impacts the model performance, as can be seen from Fig. 6. The model performance decreases on 0 notches due to the limited number of A and BBB predictions, but increases on 2 notches as the rating prediction is much smoother.

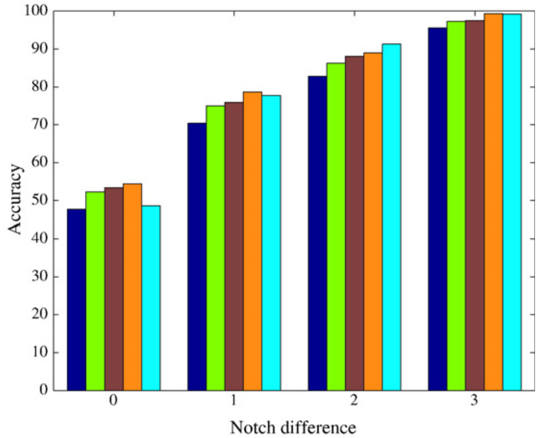


Fig. 6. In-sample performances of the linear model (■), intrinsically linear model (■), intrinsically linear model with SVM terms (■), model with qualitative variables (■), and model with rating distribution adjustment (■).

## 7. Conclusion

Rating models have typically been built to manage credit risk and to support investment decisions. Most rating models have been constructed for corporates, retail customers, countries and banks. The analysis of insurance ratings allows us to understand rating drivers for insurance companies: size, capitalization, leverage and profitability are the main drivers. On top of this, specific rating drivers focusing on type specific information were found to be important: investment yield for life; combined ratio for non-life; capitalization for reinsurance; and investment yield, loss ratio and income compared to turnover for financial guarantors. It was found that having one general model for all insurance types did not yield sufficient discrimination, but that the combination of general ratios with specific ratios does yield sufficient discrimination. Some ratios enter in the model via a non-linear transformation that allows us to indicate stress zones of the ratio in the total risk evaluation. The resulting performance of the model, using quantitative variables only, is equal to 53.4% on 0, 75.9% on 0–1 and 88.1% on 0–2 notch differences, respectively. The final model, using both quantitative and qualitative variables, as well as rating distribution adjustment, achieves an accuracy of 48.6% on 0, 77.7% on 0–1, and 91.3% on 0–2 notch differences.

## References


Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, *23*, 589−609.

Altman, E., & Rijken, H. (2004). How rating agencies achieve rating stability. *Journal of Banking and Finance*, *28*(11), 2679−2714.

A.M. Best Company, November (2003). *Understanding BCAR*. Tech. rep.

Association for Financial Professionals (2002). *Ratings agencies survey: Accuracy, timeliness, and regulation.* Available from www.afponline.org

Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, *54*(6), 627−635.

Basel Committee on Banking Supervision (2004). *International convergence of capital measurement and capital standards.* BIS.

Basel Committee on Banking Supervision (2005). Validation of lowefault portfolios in the Basel II framework. *Basel Committee Newsletter*, *Vol. 6*.

Bishop, C. (1995). *Neural networks for pattern recognition.* Oxford University Press.

Box, G., & Cox, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, *26*, 211−243.

Goodman, L. A., & Kruskal, W. H. (1972). Measures for association for cross-classification IV. *Journal of the American Statistical Association*, *67*, 415−421.

Grunert, J., Norden, L., & Weber, M. (2005). The role of non-financial factors in internal credit ratings. *Journal of Banking and Finance*, *29*(2), 509−531.

Horrigan, J. (1966). The determination of long-term credit standing with financial ratios. *Journal of Accounting Research*, *4*, 44−62.

Jeffreys, H. (1961). *Theory of probability.* Oxford University Press.

Laster, D. (2003). *Insurance Company Ratings.* Sigma 4, Swiss Re.

Martens, D., Baesens, B., Van Gestel, T., & Vanthienen, J. (in press). Comprehensible credit scoring models using rule extraction from Support Vector Machines. *European Journal of Operational Research*.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B, Methodological*, *42*(2), 109−142.

Ohlson, J. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, *18*, 109−131.

Patrino, P., Buckley, K., & Burke, J. (2001). *Fitch's approach to "Group" insurance ratings.* Special report. Fitch.

Puccia, M., Osborne, G., & Anthony, D. (2005). *Insurance criteria: Analysis of insurer capital adequacy enhanced.* Tech. rep., Standard & Poor's.

Rosch, D. (2005). An empirical comparison of default risk forecasts from alternative credit rating philosophies. *International Journal of Forecasting*, *21*(1), 37−51.

Standard & Poor's (2004). *FI criteria, bank rating analysis methodology profile.* Research report.

Suykens, J., Van Gestel, T., De Brabanter, J., De Moor, B., & Vandewalle, J. (2002). *Least squares support vector machines.* Singapore: World Scientific.

Thomas, L., Edelman, D., & Crook, J. (Eds.)(2002). *Credit scoring and its applications*. SIAM.

Van der Vaart, A. (1998). *Asymptotic statistics.* Cambridge University Press.

Van Gestel, T., Baesens, B., Van Dijcke, P., Garcia, J., Suykens, J., & Vanthienen, J. (2006a). A process model to develop an internal rating system: credit ratings. *Decision Support Systems*, *42*(2), 1131−1151.

Van Gestel, T., Baesens, B., Van Dijcke, P., Suykens, J., Garcia, J., & Alderweireld, T. (2006b). Linear and non-linear credit scoring by combining logistic regression and Support Vector Machines. *Journal of Credit Risk*, *1*(4), 31−60.

Van Gestel, T., Suykens, J., Baesens, B., Viaene, S., Vanthienen, J., Dedene, G., et al. (2004). Benchmarking least squares Support Vector Machine classifiers. *Machine Learning*, *54*, 5−32.

Vapnik, V. (1998). *Statistical learning theory.* New-York: John Wiley.