MODEL SELECTION BASED ON LORENZ AND CONCENTRATION CURVES, GINI INDICES AND CONVEX ORDER

Michel Denuit
Institute of Statistics, Biostatistics and Actuarial Science
Louvain Institute of Data Analysis and Modeling, UC Louvain
Louvain-la-Neuve, Belgium

Dominik Sznajder
Leuven Statistics Research Center, KU Leuven
Leuven, Belgium

Julien Trufin
Department of Mathematics, ULB
Bruxelles, Belgium

January 6, 2020

Abstract

In order to determine an appropriate amount of premium, statistical goodness-of-fit criteria must be supplemented with actuarial ones when assessing performance of a given candidate pure premium. In this paper, concentration curves and Lorenz curves are shown to provide actuaries with effective tools to evaluate whether a premium is appropriate or to compare two competing alternatives. The idea is to compare the premium income for sub-portfolios gathering low risks (identified as low by means of the premiums under consideration) to the true one, or equivalently, to the actual losses. Numerical illustrations performed on hypothetical data and real ones demonstrate the usefulness of the proposed approach.

Keywords: Pricing, risk classification, concentration curve, Lorenz curve, GLM, GBM, trees, neural networks.

1 Introduction

Nowadays, actuaries resort to advanced statistical tools to be able to accurately assess the risk profile of the policyholders. The cost-based technical premium (as opposed to the demand-based commercial one, actually charged to policyholders) needs to correctly evaluate the risk for each policyholder according to his or her individual characteristics. This paper entirely confines to technical premiums.

If the data are subdivided into groups determined by many features, actuaries are often faced with sparsely populated risk classes so that simple averages become suspect and regression models are needed. Regression models predict a response variable from a function of features (or explanatory variables) and parameters. By connecting the different risk profiles, they can deal with highly segmented problems resulting from the massive amount of information about the policyholders that has now become available to the insurers.

Actuarial pricing models are generally calibrated so that a measure of the goodness-of-fit is optimized (deviance or log-likelihood, in most cases). Goodness-of-fit criteria impose adherence to past, observed data. They include in-sample errors and out-of-sample errors, the latter being also known as predictive performance criteria. However, actuaries also need insurance-related metrics in addition to purely statistical ones, especially for comparing different pricing models. In insurance applications, more economic criteria must be used to decide whether a model is worth to be implemented or not. A very accurate, and thus costly to maintain pricing model offering limited gains compared to the existing, simple one, will certainly be abandoned. The notion of lift proposed by Meyers and Cummings (2009) proves to be useful in that respect. In brief, the lift measures the model's ability to prevent adverse selection. Precisely, it quantifies the model's ability to charge each insured an actuarially fair rate, thereby minimizing the potential for loosing business attracted by competitors using finer price lists.

In this paper, we aim to evaluate performance of a candidate premium based on the two following aspects:

- the variability of the resulting premium amounts, as larger premium differentials induce more lift.
- the ability of the premium income to match the true one for increasing risk profiles.

The first objective can be formalized with the help of the convex order that can be characterized by means of the Lorenz curves. This concept is often used in applied probability to compare the variability inherent to probability distributions, beyond standard deviations. The second objective is assessed by means of concentration curves. Considering the sub-portfolio gathering a given percentage of policies with the smallest amounts of premium (i.e. those policyholders susceptible to be targeted by a competitor due to their low-risk profile), the concentration curve evaluates the proportion of the total premium income corresponding to this group of contracts and compares it to the corresponding losses. The respective positions of the graphs of the Lorenz and concentration curves allow the actuary to accurately assess performance of the premium under consideration.

The present paper builds on the previous contributions by Gourieroux (1992, see also Gourieroux and Jasiak, 2011) and Frees et al. (2011, 2013). These authors defined performance curves, first in credit scoring and then for general insurance loss modeling. Compared to these previous works, we introduce here new metrics (such as ABC and ICC defined in the next sections) and expand on methodological aspects using stochastic orderings and dependence concepts.

The remainder of the paper is organized as follows. Section 2 introduces the notation, recalls some probabilistic concepts, and states technical assumptions used throughout this paper. In Section 3, we define the concentration curve and explain why this tool, supplemented with the Lorenz curve, is appropriate to assess performance of a given candidate premium. We show there that the unknown pure premium can be replaced with the actual observations so that this curve can easily be estimated from the actual data. Several useful properties of the concentration curves are also stated there. Section 4 is devoted to the comparison of the relative performances of two candidate premiums that differ either in the way they have been produced (GLM versus GBM, for instance) or in the information they contain. The variability of the premium amounts as well as the strength of their association with the response are both taken into account. These aspects are translated into mathematical terms with the help of the convex order, expectation dependence and concordance order. Section 5 illustrates the theoretical results with some hypothetical situations, by showing the impact of each characteristics on performance, such as variability or correlation with the true premium. In Section 6, we perform a case study using motor insurance data to demonstrate the usefulness of the proposed approach. The final Section 7 briefly concludes the paper.

2 True and working pure premiums

2.1 Regression function

Consider a response Y and a set of features $X_1, \ldots, X_p$ gathered in the vector $\boldsymbol{X}$. The dependence structure inside the random vector $(Y, X_1, \ldots, X_p)$ is exploited to extract the information contained in $\boldsymbol{X}$ about Y. In actuarial pricing, the aim is to evaluate the pure premium as accurately as possible. This means that the target is the conditional expectation $\mu(\boldsymbol{X}) = \mathrm{E}[Y|\boldsymbol{X}]$ of the response Y (claim number or claim amount) given the available information $\boldsymbol{X}$. Henceforth, $\mu(\boldsymbol{X})$ is referred to as the true (pure) premium.

Notice that the function $\boldsymbol{x} \mapsto \mu(\boldsymbol{x}) = \mathrm{E}[Y|\boldsymbol{X} = \boldsymbol{x}]$ is generally unknown to the actuary, and may exhibit a complex behavior in $\boldsymbol{x}$. This is why this function is approximated by a (working, or actual) premium $\boldsymbol{x} \mapsto \pi(\boldsymbol{x})$ with a relatively simple structure compared to the unknown regression function $\boldsymbol{x} \mapsto \mu(\boldsymbol{x})$. This means that the merits of a given pricing tool can be assessed using the pair $\big(\mu(\boldsymbol{X}), \pi(\boldsymbol{X})\big)$ so that we are back to the bivariate case even if there were thousands of features comprised in $\boldsymbol{X}$.

Measuring lift is an important component of model validation: once a predictive model has been built, it is essential to determine its performance of predicting the true premium $\mu(\boldsymbol{X})$ given the available features $\boldsymbol{X}$. Notice that the response Y itself does not play a direct role in the determination of the premium (beyond the definition of $\mu(\boldsymbol{X})$ and the calibration of the supervised regression model delivering the prices π), as departures $Y - \mu(\boldsymbol{X})$ cancel out when averaged over a sufficiently large portfolio (this is the very essence of insurance). The premium $\pi(\boldsymbol{X})$ has to be as close as possible to the true premium $\mu(\boldsymbol{X})$. We refer the reader to Meyer and Cummings (2009) for more details about the very aim of ratemaking, which is not to predict the actual losses Y but to create accurate estimates of $\mu(\boldsymbol{X})$, which is unobserved.

Remark 2.1. For many models considered in insurance studies, the features $\boldsymbol{X}$ are combined to form a score S (i.e. a real-valued function of the features $X_1, \ldots, X_p$) and the premium π is a monotonic, say increasing, function of S. Formally,

$$\pi(\boldsymbol{X}) = h(S)$$

for some increasing function h. Some models may include several scores, each one capturing a particular facet of the cost transferred to the insurer. This is the case for instance with zero-augmented regression models where the probability mass in zero and the expected cost when claims have been filed are both modeled by a specific score. As these scores are all functions of the available features, we keep in this paper the notation $\pi(\boldsymbol{X})$ to remain as general as possible.

2.2 Technical assumptions

To ease the exposition, we assume that premium $\pi(\boldsymbol{X})$ under consideration, as well as the conditional expectation $\mu(\boldsymbol{X})$ are continuous random variables admitting probability density functions. This is generally the case when there is at least one continuous feature contained in the available information $\boldsymbol{X}$ and the function π is a continuously increasing function of a real score S. However, this may rule out predictions based on discrete features only, as well as piecewise constant predictors, e.g., a single tree. Indeed, then $\pi(\boldsymbol{X})$ takes only a limited number of values. As actuarial pricing is nowadays based on more sophisticated models (trees being combined into random forests, for instance), this continuity assumption does not really restrict the generality of the approach.

Furthermore, we assume that the predictor is correct on average, that is,

$$\mathrm{E}[\pi(\boldsymbol{X})] = \mathrm{E}[\mu(\boldsymbol{X})] = \mathrm{E}[Y]. \tag{2.1}$$

We also assume that both the response Y and predictor π are non-negative and that $\mathrm{E}[Y] < \infty$. These assumptions are retained throughout this paper.

Notice that this paper is entirely dedicated to technical premiums. This makes the continuity assumption for $\pi(\boldsymbol{X})$ quite reasonable since nowadays actuaries have access to much more features and resort to advanced statistical tools to be able to accurately assess the risk profile of the policyholders. However, while the technical premium needs to correctly evaluate the risk for each policyholder according to individual characteristics, the commercial premium actually charged to policyholders aims to optimize insurer's profit, must comply with the rules imposed by the local supervisory authority and is subject to IT limitations and constraints from sales channels and marketing department. The simplifications of the technical price leading to the commercial one may result in a discrete pricing scheme. Specifically, a continuous feature such as the age of the policyholder is often categorized in the commercial price list while it is treated as a continuous variable in the technical premium by means of Generalized Additive Models for instance. Hence, in practice, the results derived in the following are most likely not applicable for a commercial price list (most likely discrete) while they should be for a technical price list (most likely continuous). Now, even if the theoretical results obtained in Sections 3 and 4 are only proved for continuous premiums, the key metrics "Integrated Concentration Curves" and "Area Between the Curves" introduced in Sections 4.2 and 4.3 still make sense for discrete premiums. We refer the interested reader to Henckaerts et al. (2018) for a lucid exposition of the possible simplification to the technical price list in order to arrive at the commercial price list using categorical rating factors. Section 2.4 in Yitzhaki and Schechtman (2013) explains the adjustment needed for discrete distributions. Notice however that when the price list is so simple that it results in a limited number of risk classes with a sufficient exposure, the experience can be aggregated at that level and compared to expected values using simple actual versus expected graphs.

2.3 Notation

We denote as

$$F_\pi(t) = \mathrm{P}[\pi(\boldsymbol{X}) \leq t], \;\; t \geq 0,$$

the distribution function of $\pi(\boldsymbol{X})$, as f_π the corresponding probability density function,

$$F_\pi(t) = \int_0^t f_\pi(s)\mathrm{d}s, \;\; t \geq 0,$$

and as F_π^{-1} the associated quantile function (or Value-at-Risk) defined as the generalized inverse of F_π, i.e.

$$F_\pi^{-1}(\alpha) = \inf\{t | F_\pi(t) \geq \alpha\} \text{ for a probability level } \alpha.$$

Our continuity assumption ensures that the identity

$$F_\pi\Big(F_\pi^{-1}(\alpha)\Big) = \alpha \text{ holds true for all probability levels } \alpha.$$

2.4 Convex order

Clearly, the more $\pi(\boldsymbol{X})$ is dispersed, the more information it contains about the true premium. The constant predictor $\pi(\boldsymbol{X}) = \mathrm{E}[Y]$, the least dispersed one, does not bring any information about the relative riskiness of the different policies. Thus, comparing the underlying variability appears to be important in the problem under study.

The convex order is an effective probabilistic tool to assess the dispersion of random variables, beyond simple indicators such as standard deviations. Recall from Denuit et al. (2005, Section 3.4) that given two random variables Z and T, T is said to be smaller than Z in the convex order (denoted as $T \preceq_{\mathrm{cx}} Z$) if

$$\mathrm{E}[g(T)] \leq \mathrm{E}[g(Z)] \text{ for all convex functions } g, \tag{2.2}$$

provided the expectations exist.

An important characterization of the convex order is by construction on the same probability space using conditional expectations. Precisely, the random variables Z and T satisfy $T \preceq_{\mathrm{cx}} Z$ if, and only if, there exist two random variables $\widetilde{Z}$ and $\widetilde{T}$, defined on the same probability space, such that $\widetilde{T}$ is distributed as T, $\widetilde{Z}$ is distributed as Z, and $\{\widetilde{T}, \widetilde{Z}\}$ is a martingale, that is, $\mathrm{E}[\widetilde{Z}|\widetilde{T}] = \widetilde{T}$ holds almost surely. This directly shows that increasing the number of features is beneficial as $\boldsymbol{X}_1 \subseteq \boldsymbol{X}_2 \Rightarrow \mu(\boldsymbol{X}_1) \preceq_{\mathrm{cx}} \mu(\boldsymbol{X}_2)$. Switching from $\boldsymbol{X}_1$ to the richer information $\boldsymbol{X}_2$ thus produces more dispersed premiums μ.

3 Performance curves

3.1 Definition

Following Frees et al. (2011, 2013), we propose to base performance measure of a predictor on its concentration and Lorenz curves, whose definitions are recalled next.

Definition 3.1. The concentration curve of the true premium $\mu(\boldsymbol{X})$ with respect to the working premium π based on the information contained in the vector $\boldsymbol{X}$ is defined as

$$\alpha \mapsto \mathrm{CC}[\mu(\boldsymbol{X}), \pi(\boldsymbol{X}); \alpha] = \frac{\mathrm{E}\big[\mu(\boldsymbol{X})\mathrm{I}[\pi(\boldsymbol{X}) \leq F_\pi^{-1}(\alpha)]\big]}{\mathrm{E}[\mu(\boldsymbol{X})]}.$$

Compared to Frees et al. (2011, 2013) who considered a ratio of two premiums, here the premium $\pi(\boldsymbol{X})$ itself enters the definition of a concentration curve. This extends the performance measures for a predictor $\pi(\boldsymbol{X})$ for a binary response $Y \in \{0, 1\}$ proposed by Gourieroux and Jasiak (2007, Chapter 4, see also Gourieroux, 1992). For an exhaustive review of the properties of a concentration curve we refer the interested reader to the book by Yitzhaki and Schechtman (2013).

Let us now explain the intuitive meaning of the concentration curve of $\mu(\boldsymbol{X})$. Considering Definition 3.1, we see that $\mathrm{CC}\big[\mu(\boldsymbol{X}), \pi(\boldsymbol{X}); \alpha\big]$ represents the proportion of the total true premium income corresponding to the sub-portfolio $\pi(\boldsymbol{X}) \leq F_\pi^{-1}(\alpha)$, i.e. to the $100\alpha\%$ of contracts with the smallest premium π. As these policies have a low-risk profile, they are at risk of leaving the portfolio, being attracted by a competitor. It is therefore important not to over-charge this group of policyholders. Hence the importance of the concentration curve to assess the appropriateness of the premium π under consideration.

A concentration curve alone is not enough to assess performance of π. The reason is motivated by the following comparison. Gourieroux and Jasiak (2007) were interested in credit scoring, that is, in the initial selection of applicants based on their propensity to reimburse their loan. Hence, the actual values of π did not matter, only the rank they induce and the threshold defining acceptance or rejection of the application. In pricing however, we also have to consider premium amounts $\pi(\boldsymbol{X})$ together with true premium income based on $\mu(\boldsymbol{X})$. This is why we also use the Lorenz curve of the predictor, in addition to the concentration curve of the response with respect to the predictor.

Definition 3.2. The Lorenz curve LC associated with the predictor $\pi(\boldsymbol{X})$ is defined as

$$\begin{aligned} \alpha \mapsto \mathrm{LC}[\pi(\boldsymbol{X}); \alpha] &= \mathrm{CC}[\pi(\boldsymbol{X}), \pi(\boldsymbol{X}); \alpha] \\ &= \frac{\mathrm{E}\big[\pi(\boldsymbol{X})\mathrm{I}[\pi(\boldsymbol{X}) \leq F_\pi^{-1}(\alpha)]\big]}{\mathrm{E}[\pi(\boldsymbol{X})]}. \end{aligned}$$

A Lorenz curve is thus strictly related to dispersion (or variability) by definition. From Denuit et al. (2005, Property 3.4.41), we know that increasing the predictor $\pi(\boldsymbol{X})$ in the convex order moves its Lorenz curve lower.

Remark 3.3. If an actuary has access to the true premium $\mu(\boldsymbol{X})$ based on the information contained in $\boldsymbol{X}$ then there is no need to distinguish CC from LC. This is because if $\pi(\boldsymbol{X}) = \mu(\boldsymbol{X})$ then

$$\mathrm{LC}[\pi(\boldsymbol{X});\alpha] = \mathrm{CC}[\mu(\boldsymbol{X}),\pi(\boldsymbol{X});\alpha]$$

for all probability levels α.

In other words, the two performance curves reduce to the Lorenz curve of $\mu(\boldsymbol{X})$. This means that the sub-portfolio corresponding to $\pi(\boldsymbol{X}) \leq F_\pi^{-1}(\alpha)$ is in equilibrium, as the premium matches the conditional expectation on average. A large difference between the two performance curves thus suggests that the predictor under consideration poorly approximates the true technical premium.

This explains why many empirical studies only use the Lorenz curve of $\pi(\boldsymbol{X})$ to evaluate performance of a predictive model. They confuse the predictor π with the conditional expectation whereas it is very likely that $\pi(\boldsymbol{X})$ only approximates $\mu(\boldsymbol{X})$, being different from it in reality. Because of this difference, we must resort to the pair of curves $\mathrm{CC}[\mu(\boldsymbol{X}),\pi(\boldsymbol{X});\alpha]$ and $\mathrm{LC}[\pi(\boldsymbol{X});\alpha]$ to evaluate performance of a pricing model.

3.2 Estimation

The true premium is not observed in reality, which poses a problem of estimating $\mathrm{CC}\big[\mu(\boldsymbol{X}),\pi(\boldsymbol{X});\alpha\big]$. It turns out that the following identity holds for every probability level α

$$\mathrm{CC}\big[\mu(\boldsymbol{X}),\pi(\boldsymbol{X});\alpha\big] = \mathrm{CC}[Y,\pi(\boldsymbol{X});\alpha] = \frac{\mathrm{E}\big[Y\mathrm{I}[\pi(\boldsymbol{X}) \leq F_\pi^{-1}(\alpha)]\big]}{\mathrm{E}[Y]}. \tag{3.1}$$

Identity (3.1) comes from the law of total expectation:

$$\mathrm{E}\Big[\mathrm{E}\big[Y\big|\boldsymbol{X}\big]\mathrm{I}[\pi(\boldsymbol{X}) \leq t]\Big] = \mathrm{E}\Big[\mathrm{E}\big[Y\mathrm{I}[\pi(\boldsymbol{X}) \leq t]\big|\boldsymbol{X}\big]\Big] = \mathrm{E}\big[Y\mathrm{I}[\pi(\boldsymbol{X}) \leq t]\big].$$

Hence, a concentration curve can also be interpreted as the proportion of the total losses Y attributable to the sub-portfolio gathering a given proportion α of policies with the lowest predictions.

Formula (3.1) shows that we can equivalently replace the pure premium $\mu(\boldsymbol{X})$ with the response Y in the concentration curve. This property is of utmost importance as the actuary is allowed to replace the true, unobserved premium $\mu(\boldsymbol{X})$ with the actual response values in the evaluation of performance of the predictor under consideration. Therefore, assuming the samples $(Y_i,\boldsymbol{X}_i)$, $i = 1,\ldots,n$, to be independent and identically distributed, the concentration curve of the pure premium can be estimated as follows:

$$\begin{aligned}\widehat{\mathrm{CC}}\big[\mu(\boldsymbol{X}),\pi(\boldsymbol{X});\alpha\big] &= \widehat{\mathrm{CC}}[Y,\pi(\boldsymbol{X});\alpha] \\ &= \frac{1}{n\overline{Y}}\sum_{i|\widehat{\pi}(\boldsymbol{X}_i)\leq\widehat{F}_\pi^{-1}(\alpha)} Y_i \\ &= \frac{\sum_{i|\widehat{\pi}(\boldsymbol{X}_i)\leq\widehat{F}_\pi^{-1}(\alpha)} Y_i}{\sum_{i=1}^n Y_i}.\end{aligned}$$

Here, $\widehat{\pi}$ denotes the estimated predictor. For example, for a GLM premium model it denotes the predicted values based on the estimated regression parameters. $\widehat{F}_\pi$ denotes the empirical distribution function of the resulting predictions, i.e.

$$\widehat{F}_\pi(t) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}[\widehat{\pi}(\boldsymbol{X}_i) \leq t].$$

The empirical counterpart $\widehat{\mathrm{CC}}$ to the population concentration curve CC can be interpreted as follows: a ratio of the total loss produced by those policies with estimated predictor $\widehat{\pi}$ below its empirical quantile at level α and the aggregate loss of the entire portfolio. It means that $\widehat{\mathrm{CC}}$ expresses this total sub-portfolio loss in relative terms, as a percentage of the aggregate loss at the entire portfolio level.

The empirical version of the Lorenz curve is obtained as

$$\widehat{\mathrm{LC}}\big(\pi(\boldsymbol{X});\alpha\big) = \frac{\sum_{i|\widehat{\pi}(\boldsymbol{X}_i)\leq \widehat{F}_\pi^{-1}(\alpha)} \widehat{\pi}(\boldsymbol{X}_i)}{\sum_{i=1}^{n} \widehat{\pi}(\boldsymbol{X}_i)}.$$

In words, $\widehat{\mathrm{LC}}$ is the percentage of the total premium income corresponding to the $100\alpha\%$ smaller premiums when the latter are computed using a predictor π. If the global balance $\sum_{i=1}^{n} \widehat{\pi}(\boldsymbol{X}_i) = \sum_{i=1}^{n} Y_i$ holds true (which is the case with GLMs as long as canonical link functions are used, for instance) then $\widehat{\mathrm{LC}}$ also expresses this proportion with respect to the aggregate loss at the entire portfolio level.

3.3 From premiums to ranks

Notice that

$$\pi(\boldsymbol{X}) \leq F_\pi^{-1}(\alpha) \Leftrightarrow F_\pi\big(\pi(\boldsymbol{X})\big) \leq \alpha.$$

This means that it is enough to consider the ranking induced by the predictor, that is, we are free to replace every predictor $\pi(\boldsymbol{X})$ with the corresponding rank

$$\Pi = F_\pi\big(\pi(\boldsymbol{X})\big)$$

obeying the unit uniform distribution. The intuitive meaning of Π is as follows: Π is the rank of a policyholder, once all contracts have been ordered according to their corresponding premiums (in ascending order). In credit scoring applications, only Π matters together with the acceptance threshold. In insurance pricing, the actual values $\pi(\boldsymbol{X})$ are also important and this information is captured by the Lorenz curve.

We can also express concentration curve in terms of a dependence relation between the losses and the modeled premium. In particular, we can write the numerator of the concentration curve as follows: denoting as C the copula of the pair (Y,Π), we have

$$\begin{aligned}\mathrm{E}\big[Y\mathrm{I}[\pi(\boldsymbol{X})\leq F_\pi^{-1}(\alpha)]\big] &= \mathrm{E}\big[Y\mathrm{I}[\Pi\leq\alpha]\big]\\ &= \int_0^\infty\int_0^1 y\mathrm{I}[z\leq\alpha]\mathrm{d}F_{Y,\Pi}(y,z)\\ &= \int_0^\infty\int_0^1 y\mathrm{I}[z\leq\alpha]\mathrm{d}C(F_Y(y),z)\\ &= \int_0^1 F_Y^{-1}(u)\cdot C_1'(u,\alpha)\mathrm{d}u\end{aligned}$$

where C_1' is the partial derivative of C with respect to its first argument. In fact C is also the copula of the pair $(Y,\pi(\boldsymbol{X}))$.

3.4 Properties

Let us now briefly review some important properties of the concentration and Lorenz curves. Recall that the Lorenz curve inherits the properties of the concentration curve as a special case.

3.4.1 Monotonicity

The concentration curve is based on the function $t\mapsto\mathrm{E}\big[Y\mathrm{I}[\pi(\boldsymbol{X})\leq t]\big]/\mathrm{E}[Y]$ evaluated at quantiles of $\pi(\boldsymbol{X})$. This function is clearly non-decreasing, starting from $(0,0)$ to reach $(1,1)$. Therefore, $\alpha\mapsto\mathrm{CC}[Y,\pi(\boldsymbol{X});\alpha]$ is non-decreasing and satisfies

$$\lim_{\alpha\to0}\mathrm{CC}[Y,\pi(\boldsymbol{X});\alpha]=0 \text{ and } \lim_{\alpha\to1}\mathrm{CC}[Y,\pi(\boldsymbol{X});\alpha]=1.$$

3.4.2 Line of independence/equality

In the particular case where the premium brings no information about the response, in the sense that Y and $\pi(\boldsymbol{X})$ are mutually independent, then the concentration curve is the 45-degree line, often referred to as the line of independence in the literature. Formally, if Y and $\pi(\boldsymbol{X})$ are independent then the following identity holds true:

$$\mathrm{CC}[\mu(\boldsymbol{X}),\pi(\boldsymbol{X});\alpha]=\frac{\mathrm{E}[Y]\mathrm{P}\big[\pi(\boldsymbol{X})\leq F_\pi^{-1}(\alpha)\big]}{\mathrm{E}[Y]}=\alpha.$$

Let us now study the position of the concentration curve with respect to the 45-degree line. If π brings a lot of information about the true premium $\mu(\boldsymbol{X})$, this means that these random variables are strongly related and the concentration curve should be far from the line of independence. Furthermore, the shape of the concentration curve depends on the kind of relationship between $\mu(\boldsymbol{X})$ and $\pi(\boldsymbol{X})$. The next result shows that under weak positive dependence, every concentration curve lies below the independence line.

Property 3.4. *If $\mu(\boldsymbol{X})$ is positively expectation dependent on $\pi(\boldsymbol{X})$, that is, if the inequality*

$$E[\mu(\boldsymbol{X})]\geq E\big[\mu(\boldsymbol{X})\big|\pi(\boldsymbol{X})\leq t\big]$$

holds for all t, then the concentration curve lies below the 45-degree line, i.e.

$$CC[\mu(\boldsymbol{X}), \pi(\boldsymbol{X}); \alpha] \leq \alpha \text{ for all probability levels } \alpha.$$

Proof. It suffices to write

$$\begin{aligned}\frac{\mathrm{E}\big[\mu(\boldsymbol{X})\mathrm{I}[\pi(\boldsymbol{X}) \leq t]\big]}{\mathrm{E}[Y]} &= \frac{\mathrm{P}[\pi(\boldsymbol{X}) \leq t]\mathrm{E}\big[\mu(\boldsymbol{X})\big|\pi(\boldsymbol{X}) \leq t\big]}{\mathrm{E}[Y]} \\ &\leq \mathrm{P}[\pi(\boldsymbol{X}) \leq t].\end{aligned}$$

The announced then follows by replacing t with $F_\pi^{-1}(\alpha)$. □

Considering $\mathrm{LC}[\pi(\boldsymbol{X}); \alpha]$, the 45-degree line now refers to another limit case, called the line of equality. Assume that π is constant, that is, $\pi(\boldsymbol{X}) = \mathrm{E}[Y]$. This may be due to the fact that none of the feature contained in $\boldsymbol{X}$ is related to the response Y (hence the link with the line of independence). Then, we have

$$\pi(\boldsymbol{X}) = \mathrm{E}[Y] \Rightarrow \mathrm{LC}[\pi(\boldsymbol{X}); \alpha] = \alpha$$

so that this particular case corresponds to the 45-degree line. Notice that we must consider this limit case with some care because the constant predictor is not a continuous random variable (which conflicts with the technical assumptions we stated in Section 2 and invalidates some of the formulas contained in this paper).

3.4.3 Convexity

The next result states a positive dependence condition under which the concentration curve is convex. Again, the shape of the curve depends on the kind of relationship existing between the response and the predictor. We refer the reader to Yitzhaki and Schechtman (2013) for a proof.

Property 3.5. *The concentration curve $\alpha \mapsto CC[\mu(\boldsymbol{X}), \pi(\boldsymbol{X}); \alpha]$ is convex if, and only if, $\mu(\boldsymbol{X})$ is positively regression dependent on $\pi(\boldsymbol{X})$, that is, if the function*

$$t \mapsto E\big[\mu(\boldsymbol{X})\big|\pi(\boldsymbol{X}) = t\big] \tag{3.2}$$

is non-decreasing.

Let us briefly comment on condition (3.2). In general, the condition $\pi(\boldsymbol{X}) = t$ does not fix the value of $\mu(\boldsymbol{X})$. To realize this, consider the classical log-linear GLM premium

$$\pi(\boldsymbol{X}) = \exp\left(\beta_0 + \sum_{j=1}^{p} \beta_j X_j\right).$$

In this case, $\pi(\boldsymbol{X}) = t \Leftrightarrow \sum_{j=1}^{p} \beta_j X_j = \ln t - \beta_0$ so the condition (3.2) rewrites

$$\mathrm{E}\left[\mu(\boldsymbol{X})\Big| \sum_{j=1}^{p} \beta_j X_j = \ln t - \beta_0\right].$$

As features X_j's may arbitrarily enter $\mu(\boldsymbol{X})$, not only through the linear combination $\sum_{j=1}^{p} \beta_j X_j$, $\mu(\boldsymbol{X})$ remains random conditionally on the GLM score.

Coming back to the response Y, the convexity of the concentration curve $\alpha \mapsto \mathrm{CC}[Y, \pi(\boldsymbol{X}); \alpha]$ ensures that the increments of the function

$$\mathrm{CC}[Y, \pi(\boldsymbol{X}); \alpha + \Delta] - \mathrm{CC}[Y, \pi(\boldsymbol{X}); \alpha] = \frac{\mathrm{E}\big[Y \mathrm{I}[F_\pi^{-1}(\alpha) < \pi(\boldsymbol{X}) \leq F_\pi^{-1}(\alpha + \Delta)]\big]}{\mathrm{E}[Y]}$$

are non-decreasing in α, for every positive Δ such that $\alpha + \Delta \leq 1$. This means that the sub-portfolios created by isolating a proportion Δ of policies with premiums contained between $F_\pi^{-1}(\alpha)$ and $F_\pi^{-1}(\alpha + \Delta)$ bring an increasing share of the losses on average as α increases.

This property is in relation with lift charts as described in Tevet (2013). To draw such graphs, the data set is sorted based on the values of $\pi(\boldsymbol{X})$. The data are then bucketed into equally populated classes based on quantiles. Within each bucket, the average predicted loss is calculated with the help of π as well as the actual loss cost Y. The average predicted and average actual loss costs are then graphed for each class. To assess the reasonableness of π, the analyst checks whether the actual response monotonically increases as we move to higher buckets (by definition, this will be the case for the predictions). Property 3.5 precisely identifies the condition required to observe an increasing trend in such a lift chart.

As positive regression dependence implies positive expectation dependence (see for instance Wright (1987)), the condition of Property 3.5 ensures that the concentration curve is non-decreasing and convex, starting from (0,0) to end at (1,1) with a graph everywhere below the 45-degree line. Notice that every Lorenz curve is convex by virtue of Property 3.5.

3.5 Measuring goodness-of-lift

Henceforth, we assess the performances of a predictor $\pi(\boldsymbol{X})$ by means of the respective positions of the two curves

$$\alpha \mapsto \mathrm{LC}[\pi(\boldsymbol{X}); \alpha] \text{ and } \alpha \mapsto \mathrm{CC}[\mu(\boldsymbol{X}), \pi(\boldsymbol{X}); \alpha].$$

The first one represents the share of premiums collected from the $100\alpha\%$ of policies from the portfolio with the lowest $\pi(\boldsymbol{X})$ values. The second one gives the corresponding share of the true premium that should have been collected from this sub-portfolio. As the total expected income of π and μ match the total expected loss by (2.1), the two ratios are directly comparable. Indeed, we can compare two percentages: the one obtained with $\pi(\cdot)$ with the true one corresponding to $\mu(\cdot)$. As actuaries, we would like that the graph of CC is as close as possible to the graph of LC. In other words, the smaller the area between the two curves the better.

4 Comparison of the performances of two predictors

4.1 Concentration and Lorenz curves

Assume that we have two predictors π_1 and π_2. Both attempt to predict the unknown pure premium $\mu(\boldsymbol{X})$. Many methods to obtain such predictors have been proposed, ranging from the empirical bucket averages, through the classical GLMs to tree-based methods or neural networks.

Notice that the predictors may differ in their functional form (π_1 instead of π_2) and/or in the information ($\boldsymbol{X}_1$ instead of $\boldsymbol{X}_2$) on which they are based. There are two important aspects when evaluating performance of two predictors. First, their respective variability, i.e., their ability to identify different risk profiles. Second, their correlation with $\mu(\boldsymbol{X})$, i.e., the amount of information they bring in about the true premium.

The convex order appears to be an appropriate tool to measure the degree of lift induced by a candidate premium π. This probabilistic tool indeed assesses the discrepancy between the cheapest and costliest risk profiles identified by the model. In that respect, replacing π with a more variable one, based on the convex order, appears to be a promising strategy. Considering the particular case $g(x) = x^2$ in (2.2), we see that

$$\pi_2(\boldsymbol{X}_2) \preceq_{\text{cx}} \pi_1(\boldsymbol{X}_1) \Rightarrow \text{Var}[\pi_2(\boldsymbol{X}_2)] \leq \text{Var}[\pi_1(\boldsymbol{X}_1)]. \tag{4.1}$$

This explains why $\preceq_{\text{cx}}$ is a variability order: it only applies to random variables with the same expected value and compares their respective dispersions. The convex order is a more sophisticated comparison than only focusing on the variances, yet (4.1) indicates that it agrees with this approach. Henceforth, we can interpret $\pi_2(\boldsymbol{X}_2) \preceq_{\text{cx}} \pi_1(\boldsymbol{X}_1)$ as "$\pi_1(\boldsymbol{X}_1)$ is more variable than $\pi_2(\boldsymbol{X}_2)$", keeping in mind that the variability in question extends beyond the simple comparison of standard deviations.

The next definition extends the performance curves of Gourieroux and Jasiak (2007, Definition 4.5) to more general losses, beyond binary responses.

Definition 4.1. The premium $\pi_1(\boldsymbol{X}_1)$ is more discriminatory than $\pi_2(\boldsymbol{X}_2)$ if, and only if,

$$\pi_2(\boldsymbol{X}_2) \preceq_{\text{cx}} \pi_1(\boldsymbol{X}_1) \Leftrightarrow \text{LC}[\pi_1(\boldsymbol{X}_1); \alpha] \leq \text{LC}[\pi_2(\boldsymbol{X}_2); \alpha] \text{ for all } \alpha$$

and the inequality

$$\text{CC}[\mu(\boldsymbol{X}), \pi_1(\boldsymbol{X}_1); \alpha] \leq \text{CC}[\mu(\boldsymbol{X}), \pi_2(\boldsymbol{X}_2); \alpha]$$

holds for all probability levels α.

Ordering of the concentration curves means that the share of the total premium income corresponding to the sub-portfolios gathering the $100\alpha\%$ of the policies with the smallest premiums is smaller for π_1 compared to π_2. This condition can be rewritten as follows. Denote the respective distribution functions of the two predictors π_1 and π_2 as F_{π_1} and F_{π_2}. Both F_{π_k}'s are assumed to be continuous and strictly increasing, $k = 1, 2$. Define the scores $\Pi_1 = F_{\pi_1}(\pi_1(\boldsymbol{X}_1))$ and $\Pi_2 = F_{\pi_2}(\pi_2(\boldsymbol{X}_2))$ that are both uniformly distributed over the unit interval $[0, 1]$. Then,

$$\text{E}[Y|\pi_1(\boldsymbol{X}_1) \leq F_{\pi_1}^{-1}(\alpha)] \leq \text{E}[Y|\pi_2(\boldsymbol{X}_2) \leq F_{\pi_2}^{-1}(\alpha)]$$

$$\Leftrightarrow \text{E}[Y|\Pi_1 \leq \alpha] \leq \text{E}[Y|\Pi_2 \leq \alpha].$$

This amounts to requiring that Y is more positively expectation dependent on Π_1 than on Π_2, in the sense that the reduction in the expectation resulting from the knowledge that $\Pi_k \leq \alpha$ is larger for Π_1 compared to Π_2. This kind of condition also appears in Muliere and Petrone (1992), Denuit (2010) or in Shaked et al. (2012), for instance.

The next result gives a sufficient condition for comparing competing predictors in terms of their convex relation and concordance order with the response. Recall from Denuit et al. (2005) that two random variables are said to be concordant if they tend to be both large together or small together. The concordance order expresses the idea that large and small values tend to be more often associated under the distribution that dominates the other one. Precisely, let us consider two random couples (Z_1, Z_2) and (V_1, V_2) with the same marginal distributions, i.e. $\mathrm{P}[Z_k \le t] = \mathrm{P}[V_k \le t] = F_k(t)$ for $k \in \{1, 2\}$. If

$$\mathrm{P}[Z_1 \le t_1, Z_2 \le t_2] \le \mathrm{P}[V_1 \le t_1, V_2 \le t_2] \text{ for all } t_1 \text{ and } t_2, \tag{4.2}$$

or, equivalently, if

$$\mathrm{P}[Z_1 > t_1, Z_2 > t_2] \le \mathrm{P}[V_1 > t_1, V_2 > t_2] \text{ for all } t_1 \text{ and } t_2, \tag{4.3}$$

then (Z_1, Z_2) is said to be less concordant than (V_1, V_2). This is henceforth denoted as $(Z_1, Z_2) \preceq_{\text{conc}} (V_1, V_2)$. The intuitive meaning of a ranking with respect to $\preceq_{\text{conc}}$ is clear from (4.2). Indeed, $\mathrm{P}[Z_1 \le t_1, Z_2 \le t_2]$ and $\mathrm{P}[V_1 \le t_1, V_2 \le t_2]$ read as "Z_1 and Z_2 are both small" and "V_1 and V_2 are both small", respectively (small meaning that Z_1, resp. V_1, is smaller than the threshold t_1 and Z_2, resp. V_2, is smaller than the threshold t_2). So, (4.2) means that when $(Z_1, Z_2) \preceq_{\text{conc}} (V_1, V_2)$ holds, the probability that V_1 and V_2 are both small is larger than the corresponding probability for Z_1 and Z_2. Similarly from (4.3), $(Z_1, Z_2) \preceq_{\text{conc}} (V_1, V_2)$ also ensures that the probability that Z_1 and Z_2 are both large is smaller than the corresponding probability for V_1 and V_2. This corresponds to the intuitive content of "(V_1, V_2) being more positively dependent than (Z_1, Z_2)". Moreover, Pearson's correlation coefficient as well as Kendall's and Spearman's rank correlation coefficients all agree with the concordance order. This reinforces the intuitive meaning of $\preceq_{\text{conc}}$ as a tool to compare the strength of the dependence.

We are now ready to relate the concordance order to the discriminatory power of a predictor.

Property 4.2. *If*

$$\pi_2(\boldsymbol{X}_2) \preceq_{cx} \pi_1(\boldsymbol{X}_1) \textit{ and } (Y, \Pi_2) \preceq_{conc} (Y, \Pi_1)$$

then predictor $\pi_1(\boldsymbol{X}_1)$ *is more discriminatory than predictor* $\pi_2(\boldsymbol{X}_2)$ *for response* Y*.*

Proof. The result follows from

$$\mathrm{E}[Y|\Pi_2 \le \alpha] - \mathrm{E}[Y|\Pi_1 \le \alpha] = \frac{1}{\alpha}\int_0^\infty \Big(\mathrm{P}[Y \le y, \Pi_1 \le \alpha] - \mathrm{P}[Y \le y, \Pi_2 \le \alpha]\Big)\mathrm{d}y$$

which is indeed positive if (Y, Π_1) is more concordant than (Y, Π_2) as the integrand is non-negative for all values of y in such a case. □

Thus, we see that $\pi_1(\boldsymbol{X}_1)$ is more discriminatory than $\pi_2(\boldsymbol{X}_2)$ for the response Y if $\pi_1(\boldsymbol{X}_1)$ is simultaneously more variable (in the sense of the convex order) and more correlated (in the sense of positive expectation dependence or the stronger concordance order) with the response Y than $\pi_2(\boldsymbol{X}_2)$.

4.2 Integrated concentration and Lorenz curves

The preference relation proposed in Definition 4.1 only forms a partial ranking. Two predictors might well be incomparable because their respective concentration or Lorenz curves intersect: one predictor is better for low risks, and worse for high risks, for example. In such a case, we can base the comparison on the integral of the concentration curves. This amounts to considering the integrated concentration curve defined as

$$\begin{aligned}
\mathrm{ICC}[\mu(\boldsymbol{X}),\pi(\boldsymbol{X});\alpha] &= \int_0^\alpha \mathrm{CC}[\mu(\boldsymbol{X}),\pi(\boldsymbol{X});\xi]\mathrm{d}\xi \\
&= \int_0^\alpha \frac{\mathrm{E}\big[\mu(\boldsymbol{X})\mathrm{I}[\Pi\le\xi]\big]}{\mathrm{E}[Y]}\mathrm{d}\xi \\
&= \frac{\mathrm{E}\big[\mu(\boldsymbol{X})(\alpha-\Pi)_+\big]}{\mathrm{E}[Y]} \\
&= \frac{\mathrm{Cov}\big[\mu(\boldsymbol{X}),(\alpha-\Pi)_+\big]}{\mathrm{E}[Y]} + \mathrm{E}\big[(\alpha-\Pi)_+\big].
\end{aligned}$$

The first term is driven by the correlation between the response and the predictor whereas the second one is just a constant as Π is unit uniformly distributed:

$$\mathrm{E}\big[(\alpha-\Pi)_+\big] = \int_0^\alpha(\alpha-\xi)\mathrm{d}\xi = \frac{\alpha^2}{2}.$$

By ICC, we mean the integral of the concentration curve over the whole interval $[0,1]$, i.e.

$$\begin{aligned}
\mathrm{ICC} &= \mathrm{ICC}[\mu(\boldsymbol{X}),\pi(\boldsymbol{X});1] \\
&= \frac{\mathrm{Cov}\big[\mu(\boldsymbol{X}),1-\Pi\big]}{\mathrm{E}[Y]} + \frac{1}{2} \\
&= \frac{1}{2} - \frac{\mathrm{Cov}\big[\mu(\boldsymbol{X}),\Pi\big]}{\mathrm{E}[Y]}.
\end{aligned}$$

Again, as

$$\begin{aligned}
\mathrm{E}\big[Y(\alpha-\Pi)_+\big] &= \mathrm{E}\big[\mathrm{E}[Y(\alpha-\Pi)_+|\boldsymbol{X}]\big] \\
&= \mathrm{E}\big[\mu(\boldsymbol{X})(\alpha-\Pi)_+\big]
\end{aligned}$$

we are allowed to replace $\mu(\boldsymbol{X})$ with Y in the definition of the integrated concentration curve. This means that we can use it to measure performance of $\pi(\boldsymbol{X})$ in predicting the unknown pure premium $\mu(\boldsymbol{X})$.

Let us now provide an intuitive interpretation for ICC. We still consider the $100\alpha\%$ of policies with the smallest π values, as $(\alpha-\Pi)_+=0$ for $\alpha\le\Pi$. Now, ICC is based on the covariance between $\mu(\boldsymbol{X})$ and $(\alpha-\Pi)_+$. The idea is that, the smaller Π with respect to α (i.e., the larger $(\alpha-\Pi)_+$) the smaller the true premium should be. Hence, a positive relationship between $\pi(\boldsymbol{X})$ and $\mu(\boldsymbol{X})$ translates into a negative covariance between $\mu(\boldsymbol{X})$ and $(\alpha-\Pi)_+$. And the more negative the covariance term entering the decomposition of ICC, the better the corresponding candidate premium.

Proceeding in a similar way with the Lorenz curve, we define the integrated Lorenz curve as

$$
\begin{aligned}
\mathrm{ILC}[\pi(\boldsymbol{X});\alpha] &= \int_0^\alpha \frac{\mathrm{E}\big[\pi(\boldsymbol{X})\mathrm{I}[\Pi \le \xi]\big]}{\mathrm{E}[\pi(\boldsymbol{X})]} d\xi \\
&= \frac{\mathrm{E}\big[\pi(\boldsymbol{X})(\alpha-\Pi)_+\big]}{\mathrm{E}[\pi(\boldsymbol{X})]} \\
&= \frac{\mathrm{Cov}\big[\pi(\boldsymbol{X}),(\alpha-\Pi)_+\big]}{\mathrm{E}[\pi(\boldsymbol{X})]} + \mathrm{E}\big[(\alpha-\Pi)_+\big].
\end{aligned}
$$

4.3 Some useful insurance metrics

4.3.1 Gini coefficient

Many empirical studies use Gini coefficients based on the Lorenz curve of the predictor. Let us explain the reason for using this indicator. Notice that a similar discussion can be found in Section 2.1 in Yitzhaki (2003). Recall that the Gini mean difference is one possible measure of variability defined for a non-negative continuous random variable Z as

$$
\mathrm{Gini}[Z] = \mathrm{E}\big[|Z_1 - Z_2|\big] = \mathrm{E}\big[\max\{Z_1, Z_2\}\big] - \mathrm{E}\big[\min\{Z_1, Z_2\}\big]
$$

where Z_1 and Z_2 are independent and distributed as Z. It represents the average absolute difference between two observations distributed as Z. This is closely related to the variance, which can equivalently be expressed as

$$
\mathrm{Var}[Z] = \frac{1}{2}\mathrm{E}\big[(Z_1 - Z_2)^2\big].
$$

Hence, Gini mean difference is based on an absolute difference whereas variance uses a quadratic one.

If Z is continuous then it can be shown that

$$
\mathrm{Gini}[Z] = 4\mathrm{Cov}\big[Z, F_Z(Z)\big].
$$

Thus, Gini mean difference measures the association between a random variable and its rank. In other words, considering a sequence of observations ranked in ascending order, Gini mean difference quantifies the relationship between the actual value of Z and its position in the sequence. Clearly, the more variability in Z, the larger its actual value when its rank is high (i.e. it appears among the largest observations) whereas a lower Gini mean difference indicates that the observations are more concentrated around their central value.

This formula relates the Gini mean difference to the Lorenz curve. Starting from the lower absolute deviation of Z_1, defined as $\mathrm{E}\big[(t - Z_1)\mathrm{I}[Z_1 \le t]\big]$, let us replace t with Z_2 to get

$$
\mathrm{E}\big[(Z_2 - Z_1)\mathrm{I}[Z_1 \le Z_2]\big] = \frac{1}{2}\mathrm{E}\big[|Z_1 - Z_2|\big] = \frac{1}{2}\mathrm{Gini}[Z] = 2\mathrm{Cov}\big[Z, F_Z(Z)\big].
$$

Now, the area between the identity line and the Lorenz curve for Z is $\frac{1}{\mathrm{E}[Z]}\mathrm{Cov}[Z, F_Z(Z)]$. Therefore, the higher the Gini mean difference, the further the Lorenz curve of the predictor from the 45-degree line (i.e. from the Lorenz curve of the uninformative predictor constantly equal to $\mathrm{E}[Y]$). This is why candidate premiums with larger Gini mean difference tend to be preferred.

The Gini coefficient is the Gini mean difference divided by twice the mean. It is also known as the concentration ratio and represents the area between the 45-degree line and the actual Lorenz curve divided by the area between the 45-degree line and the Lorenz curve that yields the maximal value that this index can have.

4.3.2 Area Between the Curves (ABC)

Gini mean difference measures the area between the 45-degree line and the Lorenz curve. As we have explained earlier it is better to consider both curves simultaneously: the Lorenz curve of the predictor and the concentration curve of the response with respect to the predictor. Therefore, we are interested in a distance between the two curves $\mathrm{CC}[\mu(\boldsymbol{X}),\pi(\boldsymbol{X});\alpha]$ and $\mathrm{LC}[\pi(\boldsymbol{X});\alpha]$, knowing that they coincide when $\pi(\boldsymbol{X})=\mu(\boldsymbol{X})$. This is why the area between the two curves CC and LC turns out to be a better indicator of the performance of a given predictor. This area between the curves, ABC in short, is given by

$$\begin{aligned}
\mathrm{ABC}[\pi(\boldsymbol{X})] &= \int_0^1 \Big(\mathrm{CC}[Y,\pi(\boldsymbol{X});\alpha]-\mathrm{LC}[\pi(\boldsymbol{X});\alpha]\Big)\mathrm{d}\alpha \\
&= \frac{1}{\mathrm{E}[\pi(\boldsymbol{X})]}\int_0^1 \Big(\mathrm{E}\big[Y\mathrm{I}[\Pi\le\alpha]\big]-\mathrm{E}\big[\pi(\boldsymbol{X})\mathrm{I}[\Pi\le\alpha]\big]\Big)\mathrm{d}\alpha \\
&= \frac{1}{\mathrm{E}[\pi(\boldsymbol{X})]}\int_0^1\int_0^\infty \Big(\mathrm{P}\big[\pi(\boldsymbol{X})\le y,\Pi\le\alpha\big]-\mathrm{P}\big[Y\le y,\Pi\le\alpha\big]\Big)\mathrm{d}y\mathrm{d}\alpha \\
&= \frac{1}{\mathrm{E}[\pi(\boldsymbol{X})]}\Big(\mathrm{Cov}\big[\pi(\boldsymbol{X}),\Pi\big]-\mathrm{Cov}\big[Y,\Pi\big]\Big) \qquad (4.4)
\end{aligned}$$

where we recognize the difference between the Gini mean difference of the predictor $\pi(\boldsymbol{X})$ (up to the factor 4) and the Gini covariance of the response Y and the predictor $\pi(\boldsymbol{X})$. Let us notice that (4.4) can be rewritten as

$$\mathrm{ABC}[\pi(\boldsymbol{X})]=\frac{1}{\mathrm{E}[\pi(\boldsymbol{X})]}\mathrm{Cov}\big[\pi(\boldsymbol{X})-Y,\Pi\big].$$

Hence, if we think about $\pi(\boldsymbol{X})-Y$ as the profit associated with a policy, then ABC may be interpreted to be proportional to the covariance between profits and the rank of premiums. This interpretation is similar to the one given in Frees et al. (2013) for the Gini index.

Measures like the proposed ICC and ABC share some similarities with proper scoring rules used to assess the quality of probabilistic forecasts (see e.g. Gneiting and Raftery, 2007). As explained in Section 2.1 however, the aim of insurance pricing is to estimate conditional expectations, not to predict the responses, whereas scoring rules target the predictive distribution. This being stressed, there are also connections between the two concepts. For instance, the continuous ranked probability score (or CRPS, see (20) in Gneiting and Raftery, 2007) involves Gini mean difference, as does ABC.

Remark 4.3. Notice that for $(\mu(\boldsymbol{X}), \pi(\boldsymbol{X}))$ comonotonic, we have $\mathrm{CC}[\mu(\boldsymbol{X}), \pi(\boldsymbol{X}); \alpha] = \mathrm{LC}[\pi(\boldsymbol{X}); \alpha]$ for all $\alpha \in (0,1)$ if and only if $\mu(\boldsymbol{X}) = \pi(\boldsymbol{X})$. Indeed, for $(\mu(\boldsymbol{X}), \pi(\boldsymbol{X}))$ comonotonic, we can write

$$\mu(\boldsymbol{X}) = F_\mu^{-1}(U) \quad \text{and} \quad \pi(\boldsymbol{X}) = F_\pi^{-1}(U)$$

where U is uniformly distributed over the unit interval $[0,1]$ and F_μ^{-1} is the quantile function associated to the distribution function F_μ of $\mu(\boldsymbol{X})$. Therefore, since we require $\mathrm{E}[\mu(\boldsymbol{X})] = \mathrm{E}[\pi(\boldsymbol{X})]$, we get $\mathrm{CC}[\mu(\boldsymbol{X}), \pi(\boldsymbol{X}); \alpha] = \mathrm{LC}[\pi(\boldsymbol{X}); \alpha]$ for all $\alpha \in (0,1)$ if and only if

$$\begin{aligned}
&\mathrm{E}\big[\mu(\boldsymbol{X})\mathrm{I}[\pi(\boldsymbol{X}) \leq F_\pi^{-1}(\alpha)]\big] = \mathrm{E}\big[\pi(\boldsymbol{X})\mathrm{I}[\pi(\boldsymbol{X}) \leq F_\pi^{-1}(\alpha)]\big] \\
&\Leftrightarrow \mathrm{E}\big[F_\mu^{-1}(U)\mathrm{I}[F_\pi^{-1}(U) \leq F_\pi^{-1}(\alpha)]\big] = \mathrm{E}\big[F_\pi^{-1}(U)\mathrm{I}[F_\pi^{-1}(U) \leq F_\pi^{-1}(\alpha)]\big] \\
&\Leftrightarrow \int_0^\alpha F_\mu^{-1}(u)du = \int_0^\alpha F_\pi^{-1}(u)du. \qquad (4.5)
\end{aligned}$$

Thus, since (4.5) must be fulfilled for all $\alpha \in (0,1)$, it comes $F_\mu^{-1}(u) = F_\pi^{-1}(u)$ for all $u \in (0,1)$ and so $\mu(\boldsymbol{X}) = \pi(\boldsymbol{X})$.

5 Numerical examples

5.1 Assumptions

Consider $\pi(\boldsymbol{X})$ obeying the Gamma distribution with mean μ and variance σ^2, henceforth denoted as $\mathcal{G}am(\mu, \sigma^2)$. We take $\mu = 1$. Such predictors are known to be ordered in the $\preceq_{\mathrm{cx}}$-sense with σ.

Let us also consider two distributions for the conditional expectation $\mu(\boldsymbol{X})$:

- a Gamma distribution with unit mean;
- a LogNormal distribution with unit mean, i.e. $\ln \mu(\boldsymbol{X})$ is Normally distributed with mean $-\sigma_Y^2/2$ and variance σ_Y^2, which is henceforth denoted as $\mathcal{LN}or(-\sigma_Y^2/2, \sigma_Y)$.

The condition (2.1) is fulfilled in both cases as $\mathrm{E}[Y] = \mathrm{E}[\mu(\boldsymbol{X})] = \mathrm{E}[\pi(\boldsymbol{X})] = 1$ holds true. Notice that the response may be discrete (such as the number of claims considered in our case study, for instance) as the continuity assumption only concerns $\mu(\boldsymbol{X})$ and $\pi(\boldsymbol{X})$.

In addition to the parameters σ and σ_Y governing the variability of the predictor and of the true premium respectively, we also consider different dependence structures linking these two random variables, described by means of copulas. First, we consider that the copula linking $\pi(\boldsymbol{X})$ to $\mu(\boldsymbol{X})$ is monotonically $\preceq_{\mathrm{conc}}$-increasing with its parameter. This is for instance the case for Frank and Clayton copulas considered in this section. Recall from Denuit et al. (2005) that the Clayton copula is given by

$$C_\theta(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta}, \ \theta > 0,$$

whereas Frank's copula is given by

$$C_\theta(u,v) = -\frac{1}{\theta}\ln\left(1 + \frac{(\exp(-\theta u) - 1)(\exp(-\theta v) - 1)}{\exp(-\theta) - 1}\right), \ \theta \neq 0.$$

These two copulas express positive dependence (for positive values of the parameter in Frank's case) so that the results established in the previous sections hold true.

Parameter θ can be interpreted as a measure of strength of the dependence between $\mu(\boldsymbol{X})$ and $\pi(\boldsymbol{X})$. In order to make the dependence parameter θ more palatable, we use the corresponding Kendall's rank correlation coefficient, or Kendall's tau. For Clayton's copula, Kendall's tau is simply equal to $\frac{\theta}{\theta+2}$ whereas for Frank's copula, Kendall's tau is an increasing function of θ (its expression involving the Debye function).

5.2 Variability

Gamma family of distributions is known to be ordered in the $\preceq_{\text{cx}}$-sense with respect to variance if we keep the mean fixed. Let us consider several variance levels and fix the copula to be Clayton with Kendall's tau parameter equal to 0.5. The following table summarizes the results presented in Figure 5.1.

Line type	$\pi(\boldsymbol{X})$	$\mu(\boldsymbol{X})$	Copula C	ABC
medium dash	$\mathcal{G}am(1,1)$	$\mathcal{G}am(1,2)$	Clayton($\tau = 0.5$)	6.33%
short dash	$\mathcal{G}am(1,1)$	$\mathcal{G}am(1,1)$	Clayton($\tau = 0.5$)	9.66%
dotted	$\mathcal{G}am(1,1)$	$\mathcal{G}am(1,0.5)$	Clayton($\tau = 0.5$)	13.08%

We can see that the concentration curves are non-crossing as a result of the convex order among the different distributions of the conditional expectations. Furthermore, the smaller the variance of $\mu(\boldsymbol{X})$ the further away the concentration curve from the Lorenz curve which leads to a decreasing of the ABC value with the variance of $\mu(\boldsymbol{X})$. Hence, this example highlights the fact that when we have identically distributed predictors that perform similarly in terms of dependence with the true premium, the ABC metric will favor the case where the true premium is the most variable (in the convex order sense). Similarly, for a given true premium and predictors performing the same way in terms of dependence with the true premium, the ABC metric will favor the predictor that is the less variable in terms of the convex order.

The situation where $\mu(\boldsymbol{X}) \preceq_{\text{cx}} \pi(\boldsymbol{X})$ may be due to overfitting. This can happen for instance, if π integrates random noise. Indeed, assume that only the first q features, $q < p$, $X_1, \ldots, X_q$ matters and that $X_{q+1}, \ldots, X_p$ are independent, zero-mean random variables, independent of $X_1, \ldots, X_q$. Then, the true score $\beta_0 + \sum_{j=1}^{q} \beta_j X_j$ is dominated by $\beta_0 + \sum_{j=1}^{p} \beta_j X_j$ in the convex sense.

On the contrary, the situation where $\pi(\boldsymbol{X}) \preceq_{\text{cx}} \mu(\boldsymbol{X})$ may be due to underfitting, which is the case when $\mu(\boldsymbol{X}) = \mathrm{E}[Y|X_1, \ldots, X_q, X_{q+1}, \ldots, X_p]$ and $\pi(\boldsymbol{X}) = \mathrm{E}[Y|X_1, \ldots, X_q]$ for instance. Indeed, as mentioned in Section 2.4, increasing the number of features produce more dispersed premiums.

5.3 Dependence

If we now keep the conditional distribution and the predictor distribution fixed we can single out the effect of the strength of the dependence by considering copula for several parameter values corresponding to different Kendall's tau coefficients. The following table summarizes the setup of Figure 5.2:

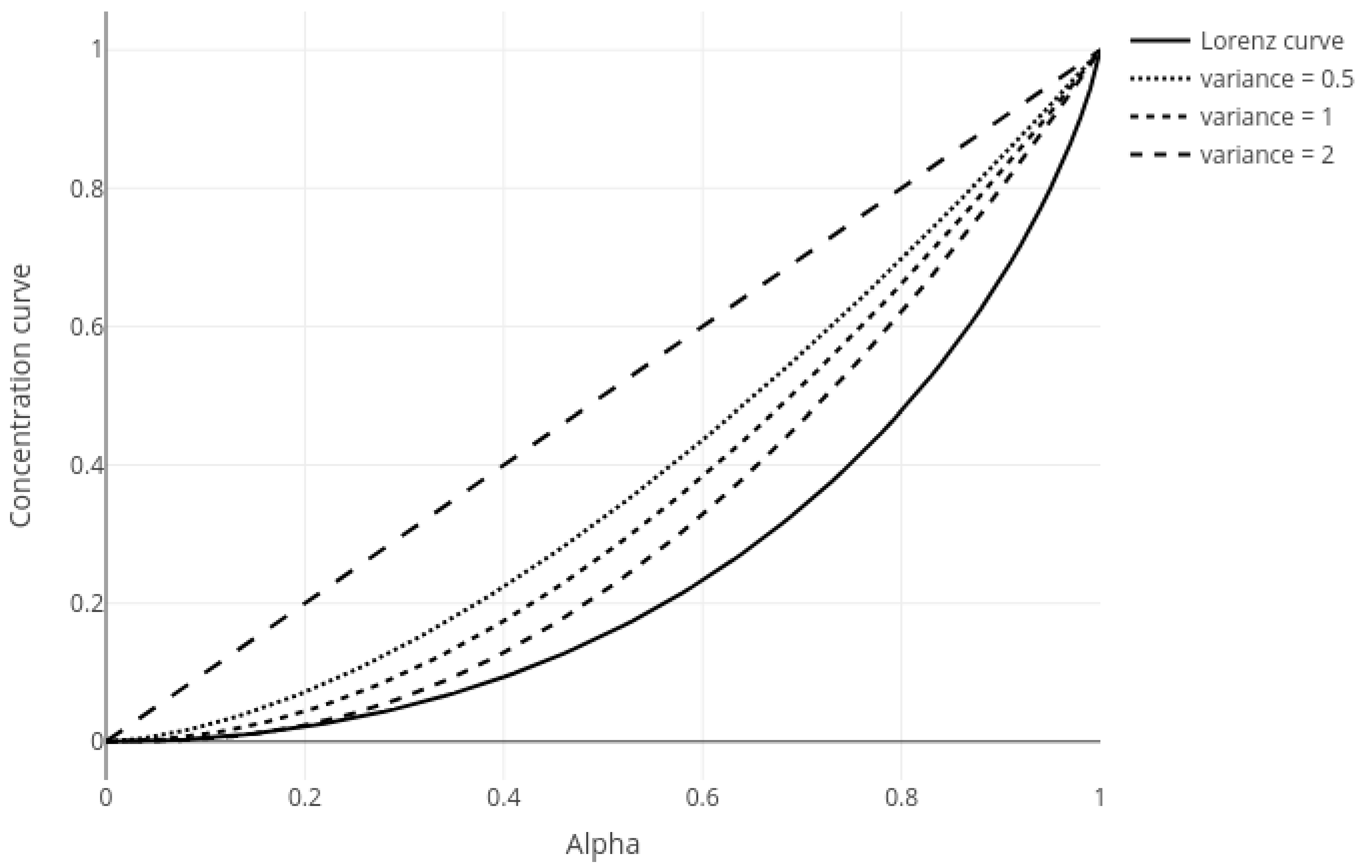

Figure 5.1: Lorenz curve and several concentration curves for different variances of conditional expectation distribution.

Line type	$\pi(\boldsymbol{X})$	$\mu(\boldsymbol{X})$	C	ABC
medium dash	$\mathcal{G}am(1,1)$	$\mathcal{G}am(1,1)$	Clayton($\tau = 0.75$)	3.46%
short dash	$\mathcal{G}am(1,1)$	$\mathcal{G}am(1,1)$	Clayton($\tau = 0.50$)	9.66%
dotted	$\mathcal{G}am(1,1)$	$\mathcal{G}am(1,1)$	Clayton($\tau = 0.25$)	17.04%

Similarly to the effect of the variance we observe that the weaker the dependence the further away the concentration curve from the Lorenz curve. This can be formally explained as follows. Increasing Kendall's tau results in a random pair $(\mu(\boldsymbol{X}), \pi(\boldsymbol{X}))$ larger in the sense of $\preceq_{\text{conc}}$ when both components are linked with Clayton copula. We then know from Property 4.2 that the concentration curve gets lower, and thus closer to the Lorenz curve. Increasing Kendall's tau means that $\pi(\boldsymbol{X})$ becomes more informative about the true premium $\mu(\boldsymbol{X})$. Also, the ABC values decreases when the association is stronger, as measured by Kendall's tau. This relation is depicted in Figure 5.2.

5.4 Distribution

We have seen cases where for ordered distributions or copulas the order of the concentration curves is preserved. However, this order is not guaranteed in arbitrary cases. In Figure 5.3

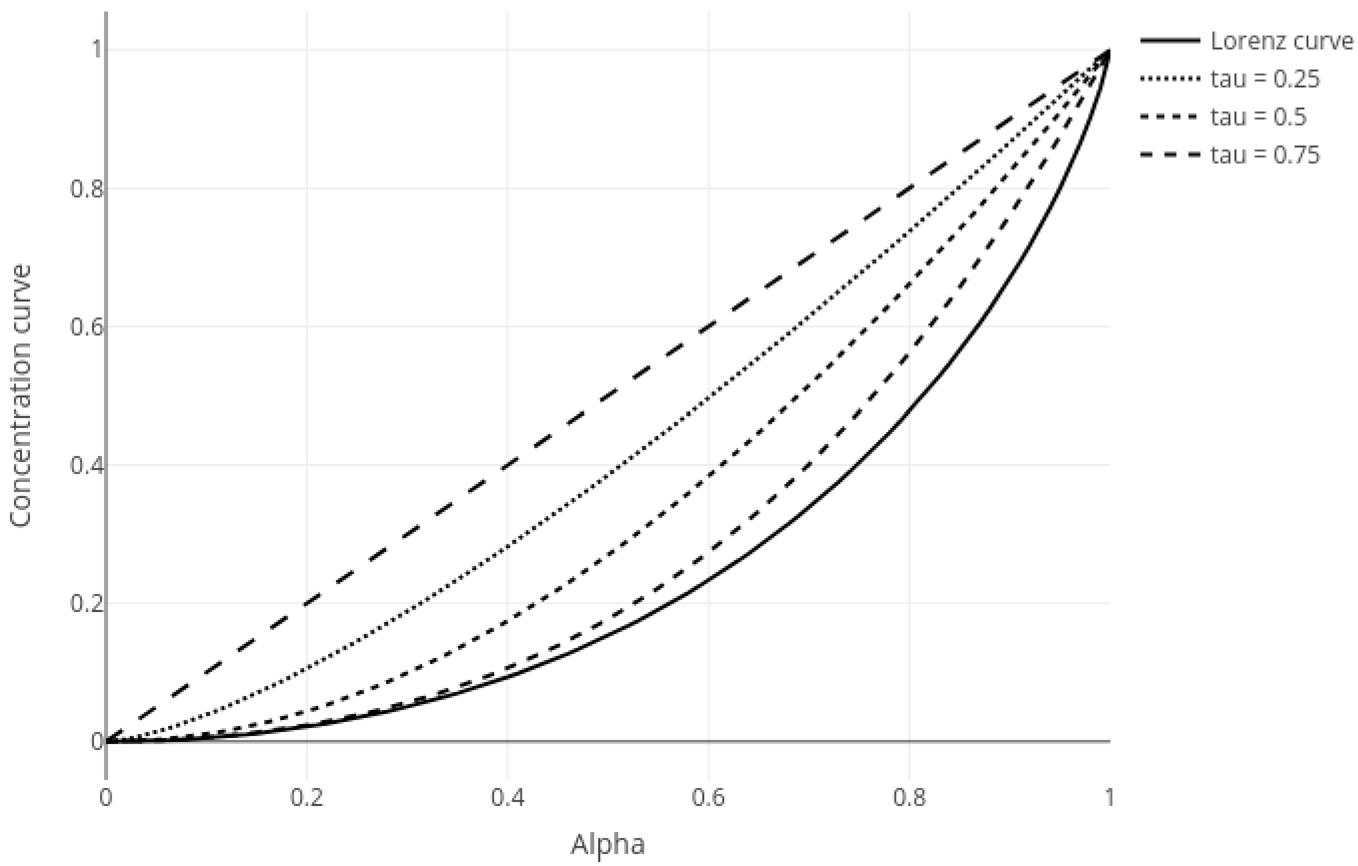

Figure 5.2: Lorenz curve and several concentration curves for different tau parameters of the copula.

we can see a situation where the predictor obeys the reference Gamma distribution whereas the conditional expectations follow LogNormal distributions with different parameters. We also fix the copula to be Clayton with Kendall's tau equal to 0.5. The corresponding values are listed in the following table:

Line type	$\pi(\boldsymbol{X})$	$\mu(\boldsymbol{X})$	C	ABC
medium dash	$\mathcal{G}am(1,1)$	$\mathcal{LN}or\left(-\frac{(1.25\sqrt{\ln 2})^2}{2}, 1.25\sqrt{\ln 2}\right)$	Clayton($\tau = 0.5$)	9.20%
short dash	$\mathcal{G}am(1,1)$	$\mathcal{LN}or\left(-\frac{\ln 2}{2}, \sqrt{\ln 2}\right)$	Clayton($\tau = 0.5$)	11.58%
dotted	$\mathcal{G}am(1,1)$	$\mathcal{G}am(1,1)$	Clayton($\tau = 0.5$)	9.66%

When the variances of the two distributions are equal (to 1) then the LogNormal concentration curve (short dash) lies further away from the Lorenz curve than the Gamma one (dotted). In these two cases, the dependences with $\pi(\boldsymbol{X})$ being also equal, the ABC value favors the case where the distributions of $\pi(\boldsymbol{X})$ and $\mu(\boldsymbol{X})$ are similar.

Now, when we increase σ_Y by factor 1.25 (red case) then the concentration curves cross around point 0.34. Thus, different distributions lead to different curvatures of their concentration curves and by adjusting their variability or the strength of copula dependence we can make them cross at an arbitrary α.

Let us notice that the ABC value is the smallest one in the red case, which is no surprising in light of Section 5.2 since this is the case where the variance of $\mu(\boldsymbol{X})$ is the largest one.

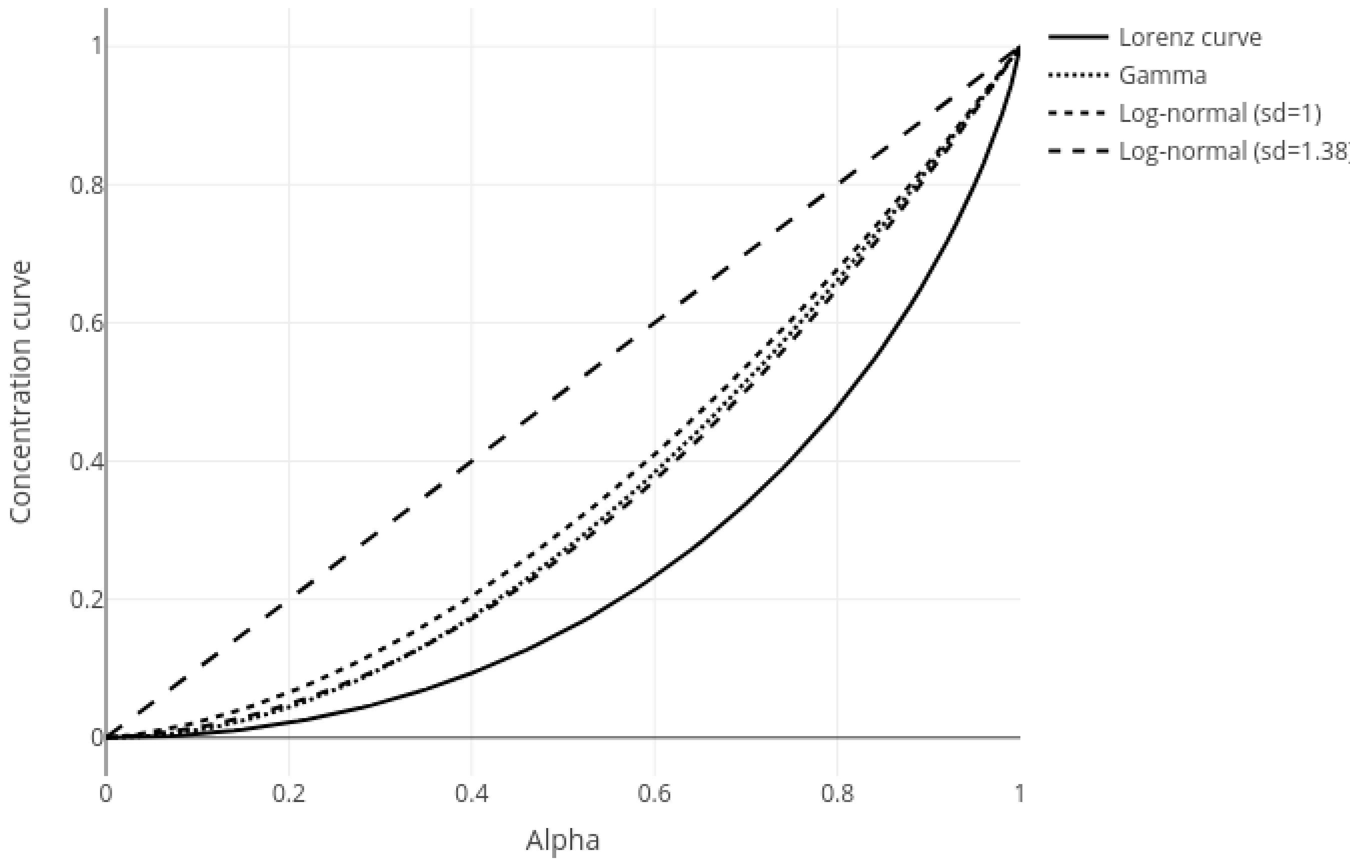

Figure 5.3: Lorenz curve and several concentration curves for different distributions of conditional expectation.

5.5 Crossing copulas

Similarly to the previous example, we can also use different copulas to obtain crossing concentration curves. To this end, let us consider a Clayton copula C_1 and a Frank copula C_2 as in Example 2.3 in Denuit and Mesfioui (2013). In this case, we know that there exists a function f such that $C_1(u,v) - C_2(u,v) \leq 0$ if $v \leq f(u)$ and $C_1(u,v) - C_2(u,v) \geq 0$ if $v \geq f(u)$. Hence, these two copulas are not ordered according to the concordance order.

In Figure 5.4 we look at the same distribution of the predictor and the conditional expectation but different copulas: Clayton and Frank copulas with the same Kendall's tau equal to 0.5. The setup is summarized in the following table:

Line type	$\pi(\boldsymbol{X})$	$\mu(\boldsymbol{X})$	C	ABC
short dash	$\mathcal{G}am(1,1)$	$\mathcal{G}am(1,1)$	Frank($\tau = 0.5$)	7.79%
dotted	$\mathcal{G}am(1,1)$	$\mathcal{G}am(1,1)$	Clayton($\tau = 0.5$)	9.66%

The curves cross (around point 0.35) which is an intuitive result as Clayton copula has stronger dependence in the lower quadrant than Frank copula and, if overall dependence is equal for both then, in the upper quadrant the opposite holds. The stronger the dependence the closer the concentration curve is to the Lorenz curve. Thus, Clayton copula lies closer to the Lorenz curve for the small values and further away for the large ones.

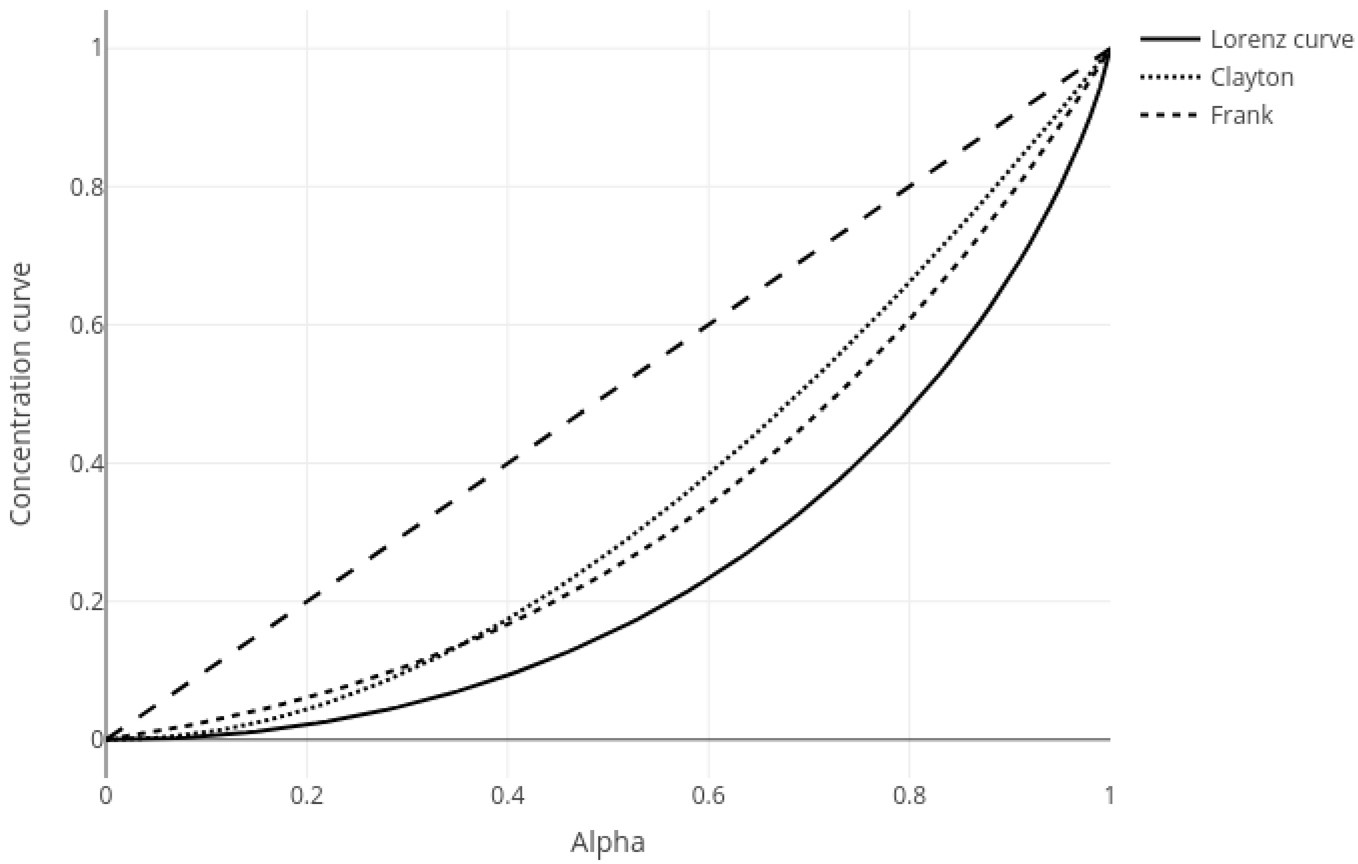

Figure 5.4: Lorenz curve and two concentration curves for different copulas.

5.6 Non-regression dependent copula impact

Let us now consider a copula that does not exhibit positive quadrant dependence but only positively expectation dependence. To this end, let us proceed as in Egozcue et al. (2011) and mix two copulas expressing quadrant dependence of opposite signs. For instance, considering the Frechet-Hoeffding upper and lower bound copulas, we employ the mixture

$$C(u,v) = (1-\theta)\min\{u,v\} + \theta \max\{0, u+v-1\} \tag{5.1}$$

as in Example 2.1 of Denuit and Mesfioui (2017). We know from Egozcue et al. (2011) that this mixture expresses positive expectation dependence if, and only if, $\theta \leq \frac{1}{2}$. Alternatively, the Frechet-Hoeffding lower bound copula may be replaced with another copula expressing negative quadrant dependence (such as the Farlie-Gumbel-Morgenstern, or FGM copula with negative dependence parameter, see Egozcue et al., 2011).

In Figure 5.5 we can confirm that the above mixture copula leads to a non-convex concentration curve. The considered setup of the numerical study is as follows:

Line type	$\pi(\boldsymbol{X})$	$\mu(\boldsymbol{X})$	C	ABC
dotted	$\mathcal{G}am(1,1)$	$\mathcal{G}am(1,1)$	(5.1) with $\theta = 0.8$	10%

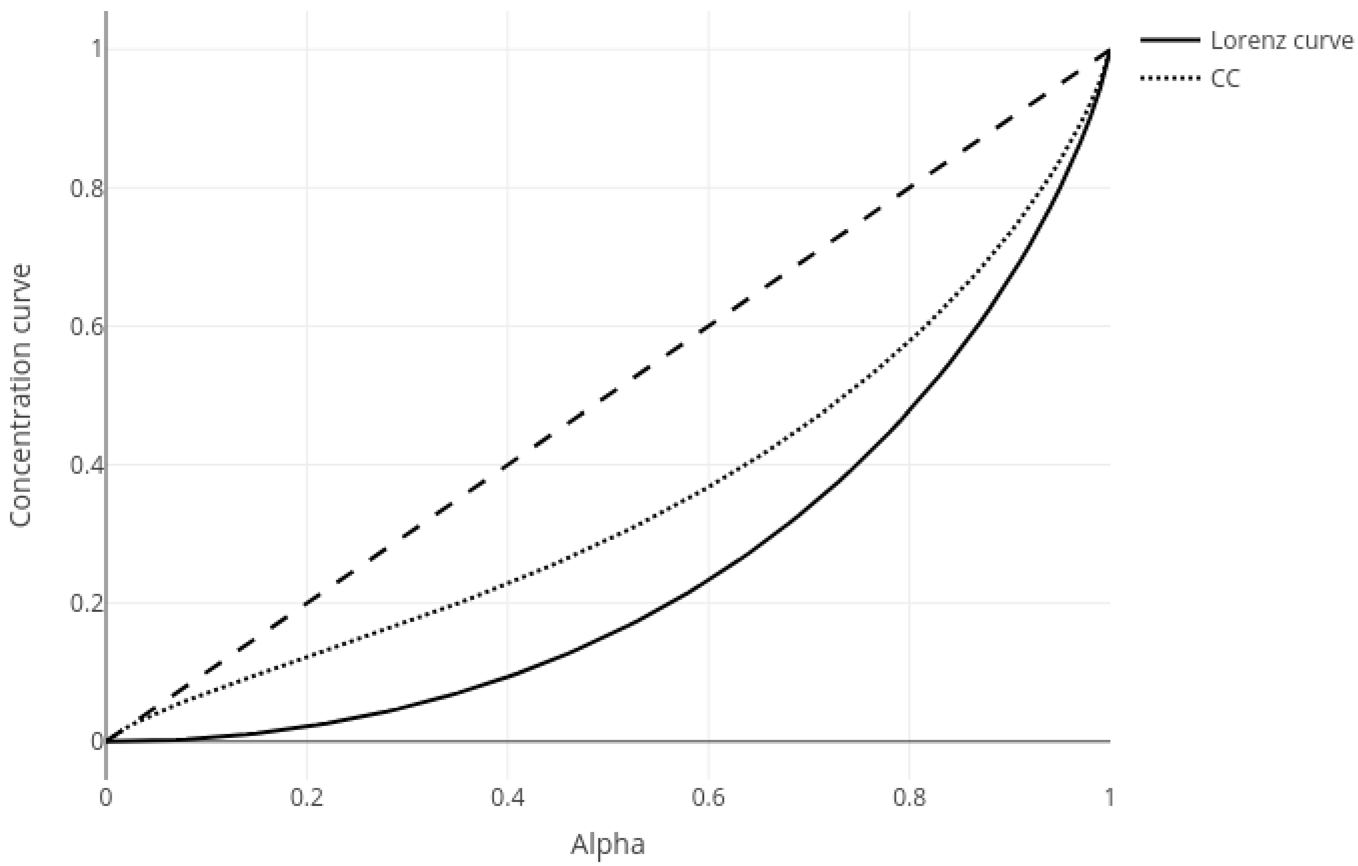

Figure 5.5: Lorenz curve and a concentration curves for a non-regression dependent copula.

6 Case study

6.1 Frequency

We consider here a data set corresponding to a French motor third-party liability insurance portfolio available in the `CASdatasets` package in `R`. Specifically, we look at the data set freMTPL2freq which contains 678,013 observations of the number of claims (response Y) together with nine explanatory variables ($\boldsymbol{X} = (X_1, \ldots, X_9)$). The explanatory variables correspond to several characteristics of

- the policy: exposure
- the policyholder: age, density of inhabitants in the home city, region, area, bonus-malus
- the car: power, age, brand, fuel type.

This data set has been recently explored with different statistical and machine learning techniques in Noll et al. (2018). We refer the reader to this case study for a broad description of the data and models' implementation details. Noll et al. (2018) compare different models based on in- and out-of-sample errors. The exact definitions for the in-sample and out-of-sample errors are given by formulas (2.2) and (2.3) in Noll et al. (2018). In this section we compare some of these models (also built upon a Poisson deviance loss function) using the goodness-of-lift metrics introduced in this paper.

More specifically, we consider in this section the following models of Noll et al. (2018) for the predictors $\pi_k(\boldsymbol{X}_k)$:

- glm1 – Poisson GLM with a log-link function and all explanatory variables
- glm3 – same as glm1 but without area and region variables
- pbm1 – boosted SBS (Standardized Binary Splits) tree (depth = 1, iterations = 30)
- pbm3 – boosted SBS tree (depth = 3, iterations = 50)
- pbm3.s2 – boosted SBS tree (depth = 3, iterations = 50, shrinkage = 0.5)
- glm1.pbm3 – boosted SBS tree starting from glm1 fit (depth = 3, iterations = 50)
- nn – shallow neural network (20 neurons with one hidden layer).

We partition the data set into a training sample of 610,000 observations and a testing sample comprising the remaining observations.

Figure 6.1 shows the in- and out-of-sample errors for the models under study together with bootstrapped 95% confidence intervals. The bounds are derived for in- and out-of-sample errors individually, so only vertical and horizontal distances are meaningful. In particular, the oval shape is due to spline smoothing through the points (in-sample error, out-of-sample error): {(lower, observed), (observed,higher), (higher,observed), (observed,lower)}. Overall, in-sample error and out-of-sample error classify the models in a similar way, except the boosted tree model (pbm3) and its shrunken version. For the latter models, introducing a shrinkage factor increases the in-sample error while it reduces the out-of-sample error. This is not surprising as the introduction of a shrinkage factor aims to avoid overfitting issues. We also note that the boosted GLM model (glm1.pbm3) improves substantially over the original GLM model (glm1). However, it does not outperform the boosted SBS tree (pbm3). The latter observation indicates that the fixed structural form imposed to the expected claim frequency by the GLM model does not provide any additional explanatory insights compared

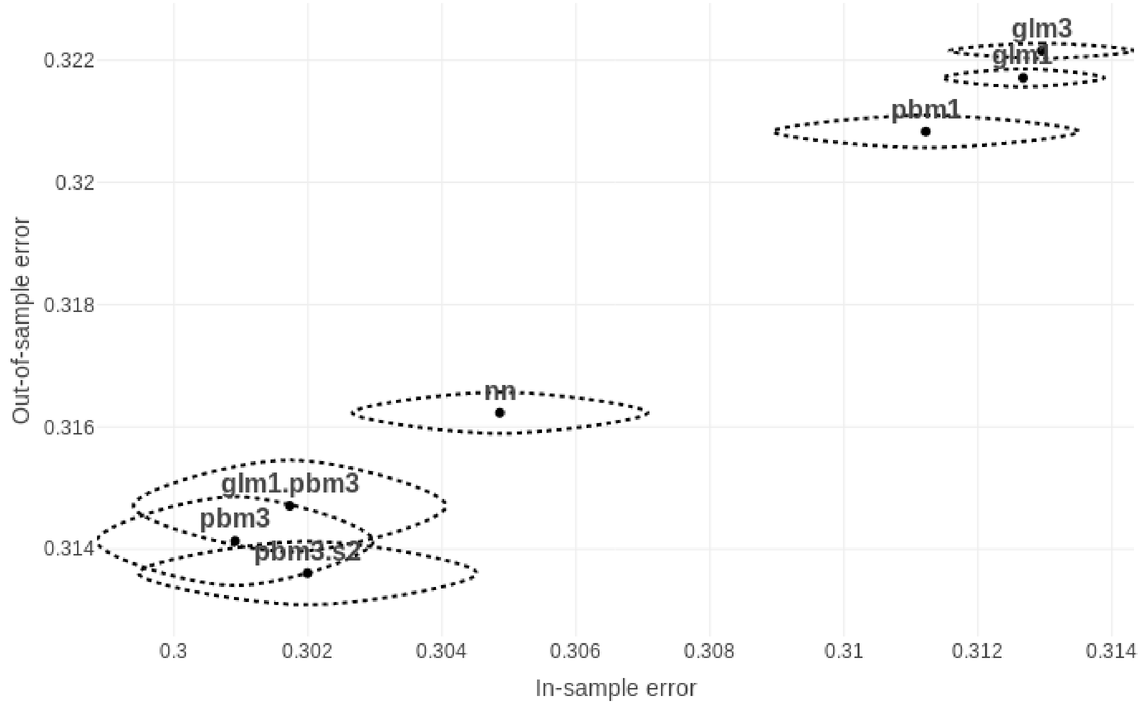

Figure 6.1: In- and out-of-sample estimation errors for models under consideration.

to the boosted SBS tree. Finally, the optimal model with respect to the out-of-sample error metric is the boosted tree model with a shrinkage factor (pbm3.s2).

Looking at the bootstrapped confidence intervals, all the models except the ones based on (pbm3) are nicely separated. It also seems that boosted methods yield more varying results than GLMs or the neural network model (for out-of-sample error).

Let us now turn to the goodness-of-lift metrics discussed in this paper. In the following, we use the empirical versions of the concentration curve $\widehat{\mathrm{CC}}$ and the Lorenz curve $\widehat{\mathrm{LC}}$ computed on the testing sample in order to get ABC and ICC values. In case the number of observations are insufficient, a smoothed version of the empirical concentration curve $\widehat{\mathrm{CC}}$ could be used instead. Here, the testing sample size is judged as sufficient to simply rely on $\widehat{\mathrm{CC}}$, depicted in Figure 6.2 for two of the considered models. The remaining models are close to the one of these tow presented clearly fomring two groups of curves. We notice that the curves form two groups for α larger than 0.15. The higher group of curves is related to models glm1, glm3 and pbm1, which are also the three worst models according to the out-of-sample errors.

ABC and ICC values are displayed in Figure 6.3 also with the same visualization of bootstrapped confidence intervals. We notice that ICC metric classifies the models as the out-of-sample error metric, except for models pbm3 and pbm3.s2. While both metrics agree that these two last models are the best ones, pbm3.s2 outperforms pbm3 according to the out-of-sample error while with ICC it is the other way around.

Regarding the ABC values, we observe that a model with a low ICC can either have low or large ABC. For instance, glm1.pbm3, which is one of the best model according to ICC,

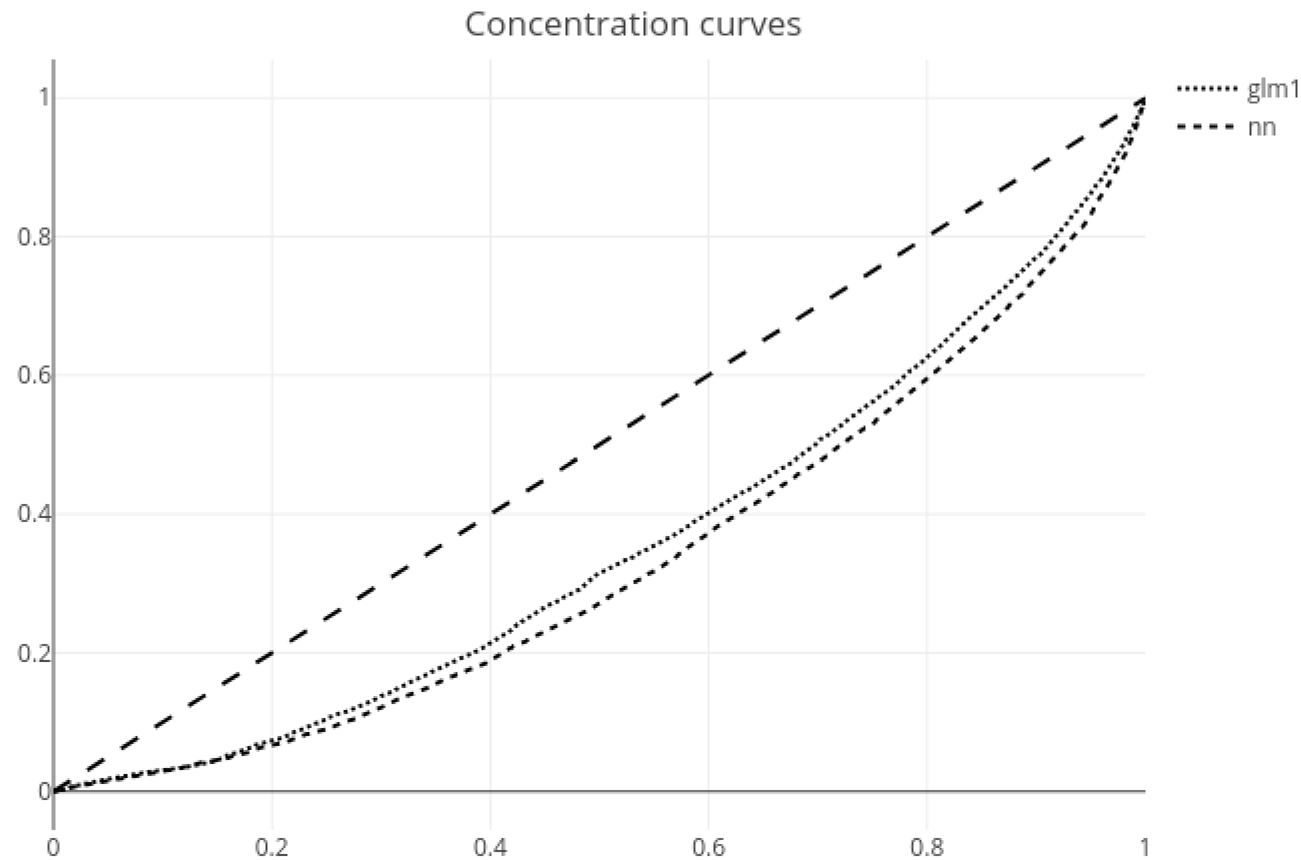

Figure 6.2: $\widehat{\mathrm{CC}}$ for models under consideration.

has the highest ABC, while pbm3.s2 has both low ICC and ABC. If we compare glm1.pbm3 and pbm3.s2 that have similar degrees of lift according to ICC, we notice that ABC metric favors pbm3.s2 that is less variable than glm1.pbm3, which is in line with Section 5.2. In the same way, while pbm3 and pbm3.s2 have similar ICC, pbm3.s2 outperforms pbm3 according to ABC, pbm3.s2 being less variable than pbm3 (since both models have the same number of trees while pbm3.s2 uses a shrinkage parameter). Finally, the optimal model with respect to ABC is pbm3.s2.

To end this example, we display in Figure 6.4 ICC and ABC as functions of α (i.e. integrating over the interval $[0, \alpha]$ instead of the whole interval $[0, 1]$). We present only curves for two models as the remaining curves look fairly similar and we do not want to blur the picture. In order to distinguish the curves more accurately we would have to zoom into specific regions (of alpha). However, even here, if we look at pbm1 and glm1.pbm3, we see that the latter has always lower ICC while the ABC values cross at around 91% quantile. Hence, from that threshold, pbm1 outperforms glm1.pbm3 according to ABC metric.

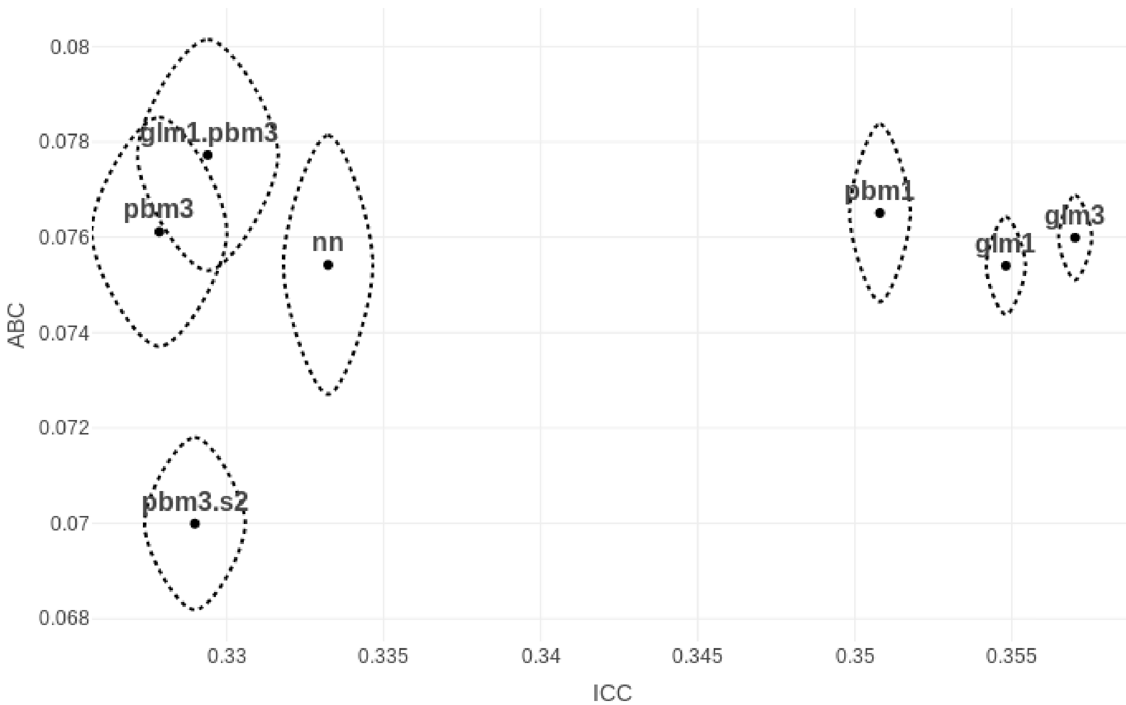

Figure 6.3: Estimated ABC and ICC values for models under consideration.

6.2 Frequency and severity

In the previous section we discussed claim frequencies. In this section we would like to touch upon severity analysis and use it as an example of how the introduced concepts work with mixed models. In particular, we use `freMTPL2sev` which contains around 26,639 aggregated claims corresponding to the previously described frequencies.

We take the model glm3 as the frequency model and consider the same set of covariates for the severity model using Gamma distribution with log link. This way we obtain a combined (frequency-severity) model for the pure premium. Alternatively, we consider a Tweedie GLM model for the pure premium directly. We use power of 1.6 (decided by a simple grid search) to account for the substantial mass at zero.

In Figure 6.5 we depict the concentration curves for both models. We observe that both curves differ over two-thirds of the interval. In particular, the Tweedie model assigns higher claims to the less risky policies than the combined model, whereas lower claims to the average risk profiles. This can help price setters to match pure premium models with segmentation objectives or marketing strategies.

The corresponding measures for both models are as follows:

Line type	model	ICC	ABC
solid	combined	0.3896	0.1336
dashed	Tweedie	0.3816	0.0446

Let us notice that we have removed the highest claim amount observation of the data set which happened to be assigned to the test set (which is actually used to derive the curves

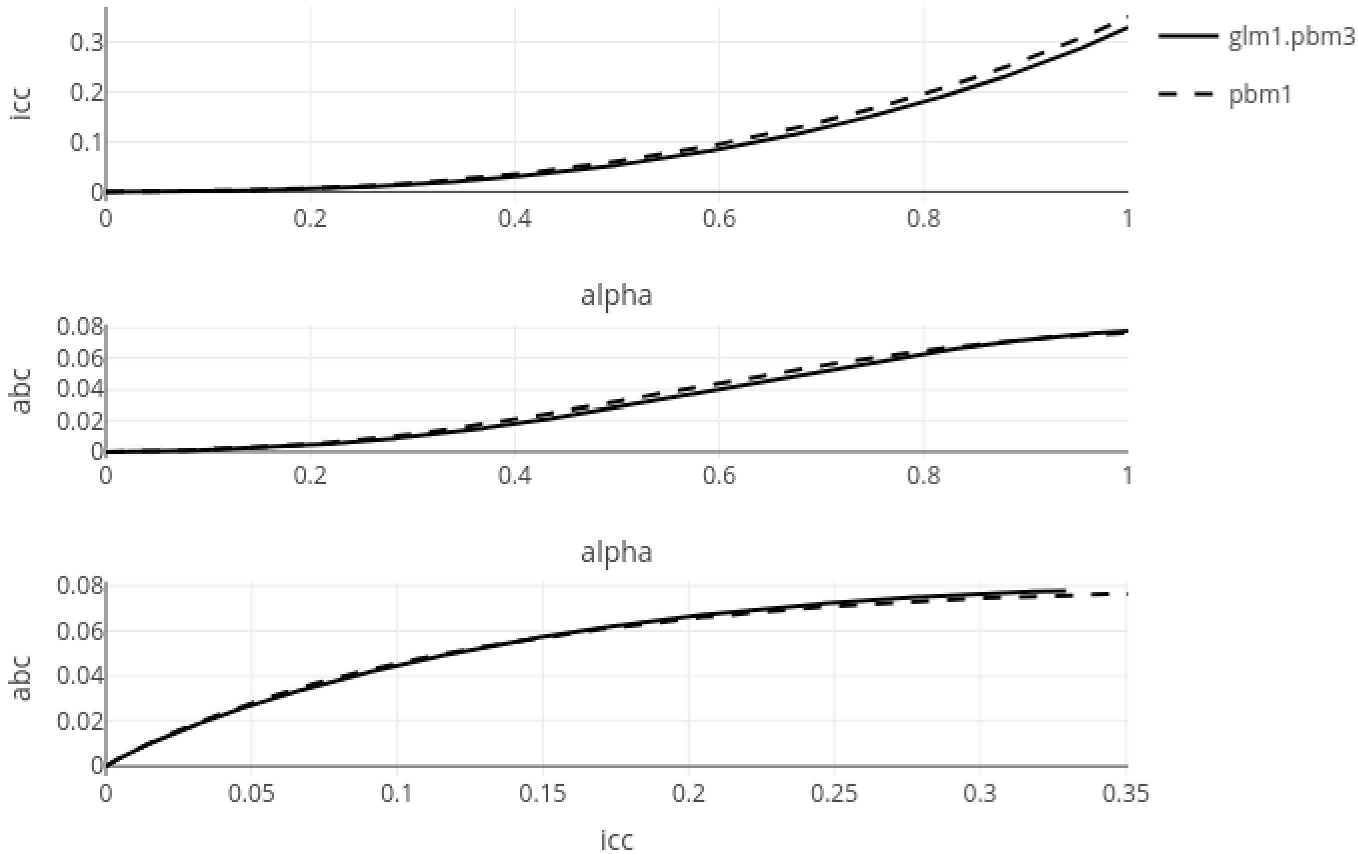

Figure 6.4: Estimated ABC and ICC values for models under consideration.

and measures). This highest observation yields a large (approximately 50%) jump in the concentration curves at points which label the corresponding estimated risk profiles.

7 Discussion

In this paper, we have proposed new metrics to assess the performances of a predictive model: ABC and ICC based on the graphs of the concentration and Lorenz curves, and their integrated versions. These indicators properly quantify the degree of lift achieved by the predictor under consideration, and possess an intuitive interpretation, useful for insurance applications. The interest of these metrics for insurance practice has been demonstrated based on simulated examples, as well as on a case study on the number of claims observed in a French motor insurance data set.

Measures of association for the pairs $(Y, \pi(\boldsymbol{X}))$ are also very useful to evaluate the performances of a given pricing model. Besides Gini coefficient, many criteria use concordance probabilities, such as Kendall's tau. We refer the interested reader to Denuit et al. (2018) for an illustration with binary responses. However, in this approach, we cannot assess the association for $(\mu(\boldsymbol{X}), \pi(\boldsymbol{X}))$ and Y is only a noisy version of $\mu(\boldsymbol{X})$.

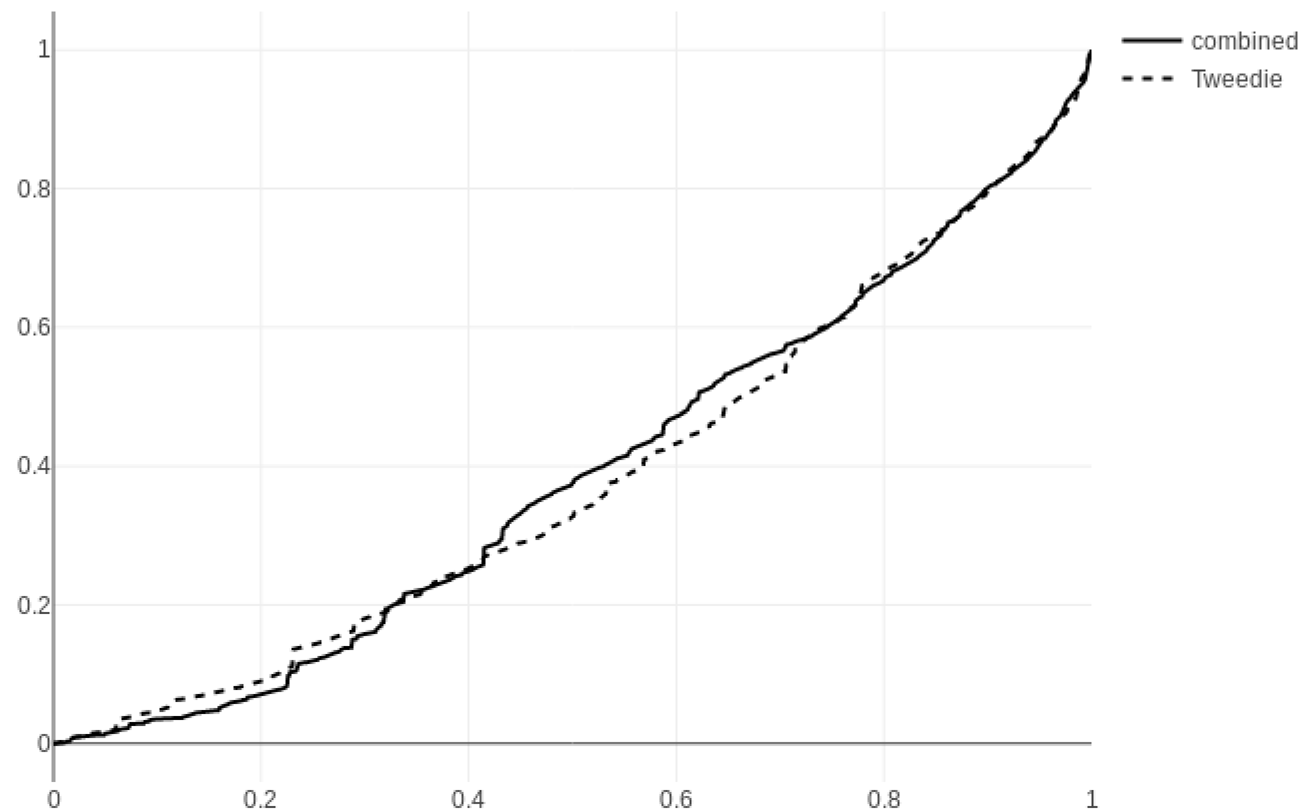

Figure 6.5: Concentration curves for the frequency-severity and Tweedie models.

Of course, no metric outperforms its competitors in every possible situation. Hence, ABC and ICC indicators enrich the actuarial toolkit, providing the analyst with new criteria to evaluate the performances of a given predictor.

References

[1] Denuit, M. (2010). Positive dependence of signals. Journal of Applied Probability 47, 893-897.

[2] Denuit, M., Dhaene, J., Goovaerts, M.J., Kaas, R. (2005). Actuarial Theory for Dependent Risks: Measures, Orders and Models. Wiley, New York.

[3] Denuit, M., Mesfioui, M. (2013). A sufficient condition of crossing-type for the bivariate orthant convex order. Statistics and Probability Letters 83, 157-162

[4] Denuit, M., Mesfioui, M. (2017). Preserving the Rothschild-Stiglitz type increase in risk with background risk: A characterization. Insurance: Mathematics and Economics 72, 1-5.

[5] Denuit, M., Mesfioui, M., Trufin, J. (2018). Bounds on concordance-based validation statistics in regression models for binary responses. Methodology and Computing in Applied Probability, in press.

[6] Egozcue, M., Garcia, L.-F., Wong, W.-K., Zitikis, R. (2011). Grüss-type bounds for covariances and the notion of quadrant dependence in expectation. Central European Journal of Mathematics 9, 1288-1297.

[7] Frees, E.W., Meyers, G., Cummings, A.D. (2011). Summarizing insurance scores using a Gini index. Journal of the American Statistical Association 106, 1085-1098.

[8] Frees, E.W., Meyers, G., Cummings, A.D. (2013). Insurance ratemaking and a Gini index. Journal of Risk and Insurance 81, 335-366.

[9] Gneiting, T., Raftery, A.E. (2007). Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102, 359-378.

[10] Gourieroux, C. (1992). Courbes de performance, de sélection et de discrimination. Annales d'Économie et de Statistique 28, 107-123.

[11] Gourieroux, C., Jasiak, J. (2011). The Econometrics of Individual Risk: Credit, Insurance, and Marketing. Princeton University Press.

[12] Henckaerts, R., Antonio, K., Clijsters, M., Verbelen, R. (2018). A data driven binning strategy for the construction of insurance tariff classes. Scandinavian Actuarial Journal, 2018(8), 681-705.

[13] Kaas, R., Hesselager, O. (1995). Ordering claim size distributions and mixed Poisson probabilities. Insurance: Mathematics and Economics 17, 193-201.

[14] Meyers, G., Cummings, A. D. (2009). "Goodness of Fit" vs. "Goodness of Lift". Actuarial Review 36, 16-17.

[15] Muliere, P., Petrone, S. (1992). Generalized Lorenz curve and monotone dependence orderings. Metron 50, 19-38.

[16] Noll, A., Salzmann, R., Wüthrich, M. (2018). Case study: French motor third-party liability claims. Available at SSRN: https://ssrn.com/abstract=3164764

[17] Shaked, M., Sordo, M. A., Suarez-Llorens, A. (2012). Global dependence stochastic orders. Methodology and Computing in Applied Probability 14, 617-648.

[18] Tevet, D. (2013). Exploring model lift: Is your model worth implementing. Actuarial Review 40, 10-13.

[19] Wright, R. (1987). Expectation dependence of random variables, with an application in portfolio theory. Theory and Decision 22, 111-124.

[20] Yitzhaki, S. (2003). Gini's mean difference: a superior measure of variability for non-normal distributions. Metron LXI(2), 285-316.

[21] Yitzhaki, S., Schechtman, E. (2013). The Gini Methodology: A Primer on Statistical Methodology. Springer.