Measuring the Value of Risk Cost Models

D. Semenovich, C. Dolman

Abstract: In this paper we demonstrate connections between several popular diagnostics used to assess the predictive power of risk cost models and introduce a new family of metrics based on economic principles. We also remind readers how any machine learning regression algorithm that supports per observation weights can be used to model risk cost without modification.

1 Introduction

Following innovations in machine learning and computational statistics, a large variety of new modeling techniques are being applied to premium rating. In order to carry out model comparison and selection in this regime it is particularly valuable to develop metrics that allow us to evaluate predictive power of candidate models with respect to the insurance outcome without relying on the knowledge of their internal structure.

Common diagnostics used today include calibration plots, quantile charts, double lift or loss ratio plots, Lorenz curves and the Gini index (Berry et al., 2009; Goldburd et al., 2016). The relationships between these tools and the potential economic value of the models are not necessarily well understood (Meyers, 2008; Meyers and Cummings, 2009). In this paper we establish a precise connection between the traditional diagnostics and the economic value and take advantage of the resulting intuition to motivate a new family of model-agnostic evaluation metrics. Finally, we remind readers of the transformations required to effectively apply standard machine learning algorithm implementations to problems with non-uniform per observation exposures.

2 Single period optimal pricing problem

To illustrate the economic rationale for the Gini index and related diagnostics, we need to first consider a simple model of single period optimal pricing (Talluri and van Ryzin, 2004).

In this model we seek to maximise the total profit objective for a cohort of n policies subject to a constraint on the minimum retention level D, where for the i-th policy with risk characteristics $\mathbf{x}_i$ the proposed premium is denoted p_i, the expected demand[1] $d_i(p_i, \mathbf{x}_i)$ is a function of premium and $c(\mathbf{x}_i)$ corresponds to the expected cost of claims:

$$\begin{aligned} \underset{p_1,\cdots,p_n}{\text{maximise}} \quad & \sum_{i=1}^{n} \big(p_i - c(\mathbf{x}_i)\big) d(p_i, \mathbf{x}_i) \\ \text{subject to} \quad & \sum_{i=1}^{n} d_i(p_i, \mathbf{x}_i) \geq D. \end{aligned}$$

[1] As we are dealing with demand levels for individual policies, d_i can also be interpreted as a probability.

Here the decision variables are premiums $p_i \geq 0$. We can rewrite the same problem using policy *demand* as the decision variable, assuming one-to-one correspondence between premium and demand $p(d_i, \mathbf{x}_i) = d^{-1}(d_i, \mathbf{x}_i)$:

$$\begin{aligned} \underset{d_1,\dots,d_n}{\text{maximise}} \quad & \sum_{i=1}^{n} \big(p(d_i, \mathbf{x}_i) - c(\mathbf{x}_i)\big) d_i = R(d_1, \dots, d_n) \\ \text{subject to} \quad & \sum_{i=1}^{n} d_i \geq D, \end{aligned}$$

where $R(d_1, \dots, d_n)$ denotes the total profit over the single period. We can then formulate the Lagrangian:

$$L(d_1, \dots, d_n, \lambda) = \sum_{i=1}^{n} \big(p(d_i, \mathbf{x}_i) - c(\mathbf{x}_i)\big) d_i + \lambda \Big(\sum_{i=1}^{n} d_i - D \Big)$$

and write the optimality conditions[2] as:

$$\begin{aligned} \frac{\partial L}{\partial d_i} &= 0, \quad 1 \leq i \leq n, \\ \frac{\partial L}{\partial \lambda} &= 0, \\ \lambda &\geq 0. \end{aligned}$$

Observe that:

$$\frac{\partial L}{\partial d_i} = \frac{\partial R}{\partial d_i} + \lambda$$

and that therefore if the portfolio is priced optimally, marginal profit with respect to demand for each policy is constant:

$$\frac{\partial R}{\partial d_i} = -\lambda. \tag{1}$$

This condition is intuitive — should $\frac{\partial R}{\partial d_i} \neq \frac{\partial R}{\partial d_j}$ for some i and j, we can reallocate demand between contracts i and j in such a way as to increase total profit.

Finally, let us express $\frac{\partial R}{\partial d_i}$ as a function of premium p_i and $\epsilon_i = \frac{\partial d_i}{\partial p_i}\frac{p_i}{d_i}$, the price elasticity of demand:

$$\begin{aligned} \frac{\partial R}{\partial d_i} &= \frac{\partial p_i}{\partial d_i} d_i + p_i - c(\mathbf{x}_i) \\ &= \left(\frac{\partial d_i}{\partial p_i} \frac{p_i}{d_i} \right)^{-1} p_i + p_i - c(\mathbf{x}_i) \\ &= p_i \left(1 + \frac{1}{\epsilon_i} \right) - c(\mathbf{x}_i). \end{aligned} \tag{2}$$

Substituting into (1) and assuming $\lambda = 0$, i.e. that the constraint on demand is not binding, we get:

$$\frac{c(\mathbf{x}_i)}{p_i} = 1 + \frac{1}{\epsilon_i},$$

[2] The solution does not need to be unique in general, however, for monotone demand functions from certain parametric families e.g. logistic and probit, the optimisation problem is convex which would mean that the solution is unique or solutions form a convex set.

or that if we price to maximise profit the loss ratio should be equal to $1 + \frac{1}{\epsilon_i}$ for every contract.

3 Calibration, Lorenz curve and Gini index

We now formally introduce four common diagnostics that are used to evaluate risk cost models — the Lorenz curve, the Gini index, the calibration plot and the quantile chart. Following Frees et al. (2011), we define the Lorenz curve with respect to a threshold parameter t as follows:

$$\begin{aligned} x_{\text{Lorenz}}(t) &= \mathbb{E}\big[\mathbb{I}\big(c(X) \geq t\big)\big] \\ &= \Pr\big(c(X) \geq t\big) = 1 - F_{c(X)}(t) = S(t), \\ y_{\text{Lorenz}}(t) &= \frac{\mathbb{E}\big[Y\mathbb{I}\big(c(X) \geq t\big)\big]}{\mathbb{E}[Y]}. \end{aligned}$$

In the above, Y corresponds to the cost of claims, X to the risk characteristics and $c(X)$ to the predictions of the risk cost model being evaluated, with $F_{c(X)}(t)$ denoting the cumulative distribution function of $c(X)$. Expectations are with respect to the joint distribution of (Y, X). Note that the sort order of the x axis is reversed relative to the traditional definition.

In practice, the curve is approximated using data held out from model estimation $\big\{(y_i, e_i, \mathbf{x}_i)\big\}_{i=1}^{n}$ with e_i denoting per observation exposure:

$$\begin{aligned} \hat{x}_{\text{Lorenz}}(t) &= \frac{\sum_{i=1}^{n} e_i \mathbb{I}\big(c(\mathbf{x}_i) \geq t\big)}{\sum_{i=1}^{n} e_i}, \\ \hat{y}_{\text{Lorenz}}(t) &= \frac{\sum_{i=1}^{n} y_i \mathbb{I}\big(c(\mathbf{x}_i) \geq t\big)}{\sum_{i=1}^{n} y_i}. \end{aligned} \tag{3}$$

The Gini index is then commonly defined with reference to the Lorenz curve:

$$\begin{aligned} G &= 2\int_0^1 y_{\text{Lorenz}}(t)\, dS(t) - 1 \\ &= 2\int_0^1 y_{\text{Lorenz}}(t) - x_{\text{Lorenz}}(t)\, dS(t), \end{aligned} \tag{4}$$

represented by twice the shaded area in Figure 3.

The calibration plot compares model predictions with the claims outcome Y:

$$\begin{aligned} x_{\text{Calibration}}(t) &= t, \\ y_{\text{Calibration}}(t) &= \mathbb{E}\big[Y \mid c(X) = t\big]. \end{aligned}$$

At last, the quantile chart is a rescaling of the calibration plot along the x axis to the same units as the Lorenz curve (the proportion of policies with $c(X) \geq t$):

$$\begin{aligned} x_{\text{Quantile}}(t) &= S(t), \\ y_{\text{Quantile}}(t) &= \mathbb{E}\big[Y \mid c(X) = t\big]. \end{aligned}$$

Finite sample approximations of the quantile chart similar to (3) can be developed, e.g. the so called "decile plot" (Goldburd et al., 2016).

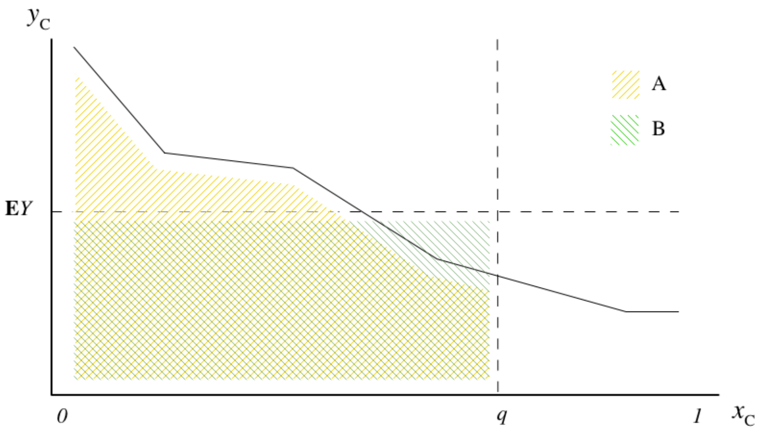

Figure 1: A quantile plot. Under assumption of $\epsilon = -1$, the area $A - B$ represents the economic gain from a demand neutral price change where we raise prices for q of total policies with highest absolute marginal profit with respect to demand as to forgo q units of demand and then offset that loss of demand through a price reduction for the entire portfolio, gaining q units of demand. Here $\mathbb{E}(Y)$ corresponds to the average cost of claims per policy.

We can immediately see the connection between the quantile chart and the Lorenz curve:

$$\begin{aligned} \int_0^{S(u)} y_{\text{Quantile}}\, dx_{\text{Quantile}} &= \int_0^{S(u)} \mathbb{E}\big[Y \,|\, c(X) = t\big]\, dS(t) \\ &= \int_\infty^{u} \mathbb{E}\big[Y \,|\, c(X) = t\big]\, \frac{dS}{dt}\, dt \\ &= \int_u^{\infty} \mathbb{E}\big[Y \,|\, c(X) = t\big]\, \Pr\big(c(X) = t\big)\, dt \\ &= \mathbb{E}\big[Y \mathbb{I}\big(c(X) \geq u\big)\big] \\ &= y_{\text{Lorenz}}(u)\, \mathbb{E}[Y]. \end{aligned} \tag{5}$$

This demonstrates that the Lorenz curve is the integral of the corresponding quantile plot, scaled by the average cost of claims per policy $\mathbb{E}[Y]$. A more detailed discussion of the correspondences between different model evaluation metrics, albeit limited to the binary classification case, can be found in Vuk and Curk (2006).

4 Economic interpretation of model diagnostics

We now combine the results from Sections 2 and 3 in order to gain an economic intuition for the diagnostic metrics.

We first observe that if we take $\epsilon_i = -1$ (a strong assumption yet not entirely unreasonable in many cases), then the equation (2) reduces to:

$$\frac{\partial R}{\partial d_i} = -c(\mathbf{x}_i),$$

allowing the interpretation of the quantile plot in terms of marginal profit — namely we can take the values on the y axis to represent *actual* (measured using realised claims experience Y) negative marginal profit with respect to demand for that segment. Unless the graph is perfectly flat, the optimality condition (1) of constant marginal profit is not satisfied and we can improve portfolio performance by rebalancing demand through price adjustments.

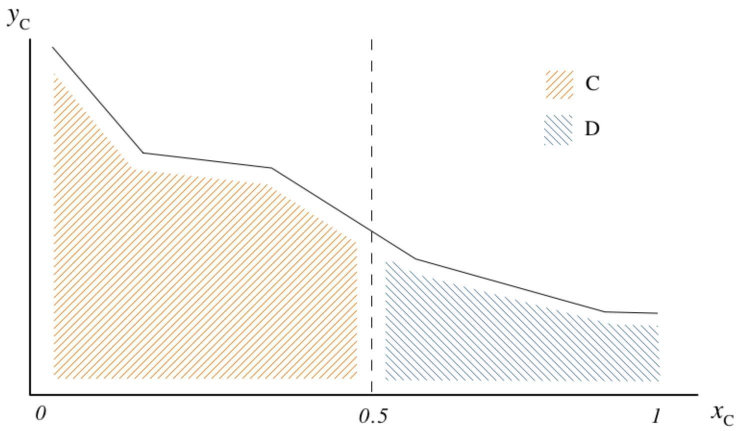

Figure 2: A quantile plot. Area $C - D$ corresponds to the economic value of a price adjustment where we increase premiums for policies in C and simultaneously reduce premiums for policies in D so as to effect offsetting demand changes of 0.5 and -0.5 units respectively.

Consider the quantile plot in Figure 1 and areas identified as A and B. In this case, the area A corresponds to the change in profit[3] for a q unit total demand change (achieved through a requisite change in premiums) applied to the fraction q of the policies with the highest expected absolute value of marginal profit with respect to demand according to the model $c(X)$:

$$\mathbb{E}\left[\frac{\partial R}{\partial d}\mathbb{I}\big(c(X) \geq S^{-1}(q)\big)\right].$$

Area B corresponds to the change in profit if the prices are changed for the entire portfolio as to effect a change of q units of demand, $q\mathbb{E}\left(\frac{\partial R}{\partial d}\right)$. The value $A - B$ can then be written as:

$$A - B = \mathbb{E}\left[\frac{\partial R}{\partial d}\mathbb{I}\big(c(X) \geq S^{-1}(q)\big)\right] - q\mathbb{E}\left[\frac{\partial R}{\partial d}\right], \tag{6}$$

representing the economic gain (per policy) from a targeted, demand neutral price change whereby we raise prices for q fraction of total policies as to forgo q units of demand and then offset the loss of demand through a price reduction for the entire portfolio, gaining q units of demand.

Next consider the quantile plot in Figure 2 and the areas identified as C and D. Using similar reasoning to above, we can write:

$$C - D = \mathbb{E}\left[\frac{\partial R}{\partial d}\mathbb{I}\big(c(X) \geq S^{-1}(0.5)\big)\right] - \mathbb{E}\left[\frac{\partial R}{\partial d}\mathbb{I}\big(c(X) \leq S^{-1}(0.5)\big)\right].$$

This area corresponds to the economic value of a price adjustment where we increase premiums for policies in C and simultaneously reduce premiums for policies in D so as to effect offsetting demand changes of 0.5 and -0.5 units respectively.

Owing to the relationship demonstrated in Section 3, we can also interpret quantities $A - B$ and $C - D$ with reference to the Lorenz curve (see Figure 3). We observe that $A - B$ corresponds to the distance between the Lorenz curve and the line $y = x$ divided by the constant $\mathbb{E}[Y]$. The Gini index, therefore, being twice the area between the Lorenz curve and the same line, as seen from (4), represents the *average economic value* of price change decisions of type (6) as we vary the threshold q, scaled by the constant $\frac{\mathbb{E}(\mathbb{Y})}{2}$.

[3] Note that since we are dealing with derivatives, effects are assumed symmetrical for price increases and decreases.

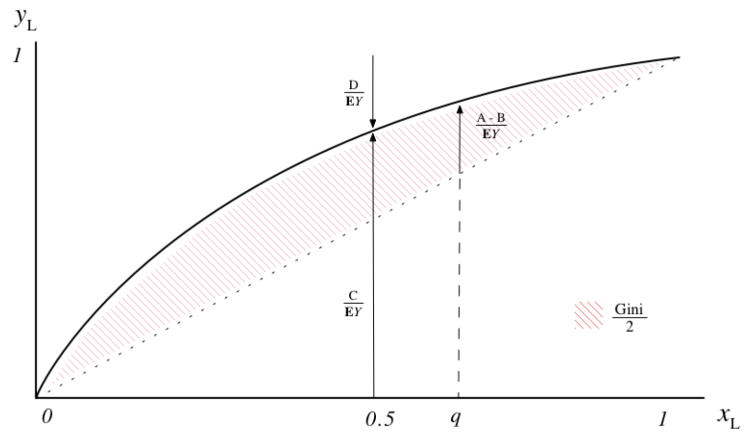

Figure 3: Lorenz Curve and the Gini index. We observe that the area $A - B$ from Figure 1 corresponds to the distance between the Lorenz curve and the line $y = x$ up to constant $\mathbb{E}[Y]$. The Gini coefficient is equal to twice the area between the Lorenz curve and the same line.

5 Marginal profit plots

We are now in the position to define a new family of models diagnostics, parametrised through the choice of elasticity assumption ϵ (still constant across the cohort of risks under consideration). The marginal profit plot (a generalisation of the quantile plot) is given by:

$$\begin{aligned} x^*_{\text{Quantile}}(t,\epsilon) &= \mathbb{E}\Big[\mathbb{I}\Big(c(X) - p(X)\Big(1+\frac{1}{\epsilon}\Big) \geq t\Big)\Big] \\ &= \Pr\Big(c(X) - p(X)\Big(1+\frac{1}{\epsilon}\Big) \geq t\Big) = S_\epsilon(t), \\ y^*_{\text{Quantile}}(t,\epsilon) &= \mathbb{E}\Big[Y - p(X)\Big(1+\frac{1}{\epsilon}\Big) \,\Big|\, c(X) - p(X)\Big(1+\frac{1}{\epsilon}\Big) = t\Big]. \end{aligned}$$

Note that we have introduced a new quantity $p(X)$, corresponding to the current premiums.

We can define the associated marginal profit Lorenz curve and the marginal profit Gini index using the relationship (5):

$$\begin{aligned} x^*_{\text{Lorenz}}(t,\epsilon) &= S_\epsilon(t), \\ y^*_{\text{Lorenz}}(t,\epsilon) &= \int_0^{S_\epsilon(t)} y^*_{\text{Quantile}}\, dx^*_{\text{Quantile}} \\ &= \int_0^{S_\epsilon(t)} \mathbb{E}\Big[Y - p(X)\Big(1+\frac{1}{\epsilon}\Big) \,\Big|\, c(X) - p(X)\Big(1+\frac{1}{\epsilon}\Big) = u\Big]\, dS_\epsilon(u) \\ &= \mathbb{E}\Big[\mathbb{I}\Big(c(X) - p(X)\Big(1+\frac{1}{\epsilon}\Big) \geq t\Big)\Big(Y - p(X)\Big(1+\frac{1}{\epsilon}\Big)\Big)\Big], \\ G^*(\epsilon) &= 2\int_0^1 y^*_{\text{Lorenz}}(t) - \mathbb{E}[Y]\, x^*_{\text{Lorenz}}(t)\;\, dS_\epsilon(t). \end{aligned}$$

We chose not to rescale y^*_{Lorenz} by $\mathbb{E}[Y]$. This is due to the well known difficulties with the definition of the Lorenz curve and associated quantities in situations where negative measurements are allowed (e.g. consider the case when $\mathbb{E}[Y] = 0$). This issue does not arise if we instead adopt unscaled "generalised" Lorenz curve following Shorrocks (1983).

With the choice of $\epsilon = -1$ we recover the standard definitions in Section 3 up to the constant $\mathbb{E}[Y]$. As we let $\epsilon \to -\infty$, we have an analogue of the so called "loss ratio chart", a plot comparing expected vs. actual loss ratios (Goldburd et al., 2016), but defined for the dollar margin $p(X) - c(X)$, rather than the ratio[4] $\frac{c(X)}{p(X)}$.

All of the interpretations developed in section 4 also apply for the marginal profit and related plots. It can be particularly informative to compare $G^*(\epsilon)$ values for the candidate models across a realistic range of elasticities.

Finally we observe that it is possible to use per observation elasticity values $\hat{\epsilon}_i$, however some care needs to be taken in this situation as the resulting statistics can be quite sensitive to the predictive uncertainty in the estimates $\hat{\epsilon}_i$. We intend to address this setting in more depth in a separate publication.

6 Machine learning for risk cost estimation

Many popular implementations of high performance machine learning algorithms, e.g. Ke et al. (2017) and Chen and Guestrin (2016), do not directly support per observation exposures e_i, which can appear to limit their applicability to risk cost estimation.

This can, however, be addressed via a simple transformation that converts data with non-uniform exposures $(y_i, \mathbf{x}_i, e_i)$, into the standard form of instance-label-weight triples $(l_i, \mathbf{x}_i, w_i)$:

- $\mathbf{x}_i$ is the feature vector for risk i
- label $l_i = \frac{y_i}{e_i}$ is the total cost of claims attaching to that policy y_i, divided by the exposure for the instance e_i,
- and $w_i = e_i$, the instance weight, is set to the exposure for that instance.

In the case of Poisson loss this is directly equivalent to adding exposure as an offset. Suppose we have an instance $\mathbf{x}_i$ and a corresponding linear predictor $\mathbf{x}_i^T \boldsymbol{\beta} = a_i$ where $\boldsymbol{\beta}$ is the parameter vector. For claim cost y_i, the Poisson negative log likelihood (loss) is:

$$\exp(a_i) - y_i a_i.$$

Normally for a proportional hazards model a_i would be replaced with $a_i + \log(e_i)$. It is easy to see that we obtain the same result if we set $l_i = \frac{y_i}{e_i}$ and $w_i = e_i$:

$$\begin{aligned} w_i\big(\exp(a_i) - l_i a_i\big) &= e_i\big(\exp(a_i) - \frac{y_i}{e_i} a_i\big) \\ &= \exp\big(a_i + \log(e_i)\big) - y_i a_i, \end{aligned}$$

as the term $y_i \log(e_i)$ can be omitted since it does not depend on a_i.

This reduction, namely training a model on the transformed data $(l_i, \mathbf{x}_i, w_i)$, is effective regardless of the choice of loss and link functions and can be applied e.g. when using squared or Tweedie losses.

[4] We can motivate the loss ratio chart in a similar manner if we consider marginal profit with respect to premium $\frac{\partial R}{\partial p_i}$ instead.

7 Conclusion

In this paper we have illustrated the connection between various risk cost model evaluation metrics and the associated economic value, a problem first posited by Meyers (2008), and have introduced a new family of diagnostics based on economic principles. Model evaluation involving contract level elasticity estimates will be addressed in future work.

References

Berry, J, Hemming, G, Matov, G, Morris, O., *Report of the Model Validation and Monitoring in Personal Lines Pricing Working Party*, GIRO Conference, 2009

Chen, T., Guestrin, C., *XGBoost: A Scalable Tree Boosting System*, KDD, 2016

Frees, E., Meyers, G., Cummings, D., *Summarizing insurance scores using a Gini index*, Journal of the American Statistical Association, 106 (495), pp. 1085–1098, 2011

Goldburd, M, Khare, A, Tevet, D. *Generalised Linear Models for Insurance Rating*, Casualty Actuarial Society Monograph Series Number 5, 2016.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T., *LightGBM*: A Highly Efficient Gradient Boosting Decision Tree, NIPS, 2017

Meyers, G., *The Value of Lift*. The Actuarial Review March 2008, Casualty Actuarial Society, 2008.

Meyers, G., Cummings, D., *"Goodness of Fit" vs. "Goodness of Lift"*, The Actuarial Review August 2009, Casualty Actuarial Society, 2008.

Shorrocks, A., *Ranking Income Distributions*, Economica, 50 (197), pp. 3-17, 1983

Talluri, K, van Ryzin, G. *The Theory and Practice of Revenue Management*, Springer, 2004

Vuk, M, Curk, T. *ROC Curve, Lift Chart and Calibration Plot*, Metodoloski zvezki 3, pp. 89–108, 2006