

# FAQ CTU ARY

FAQCTUARY 2022-2

## DOES AUTOCALIBRATION IMPROVE GOODNESS OF LIFT?

---

By Nicolas Ciatto, Harrison Verelst,  
Michel Denuit, and Julien Trufin

July 2022

## DISCLAIMER

The content of the FAQtuary's is for a pedagogical use only. Each business case is so specific that a careful analysis of the situation is needed before implementing a possible solution. Therefore, Detralytics does not accept any liability for any commercial use of the present document. Of course, the entire team remain available if the techniques presented in this FAQtuary required your attention.

Detralytics  
Avenue du Boulevard, 21 Box 5  
1210 Brussels  
[www.detralytics.com](http://www.detralytics.com)  
[info@detralytics.eu](mailto:info@detralytics.eu)

## ABSTRACT

Autocalibration is a desirable property, intimately related to the method of marginal totals that predates modern risk classification methods. Autocalibrated mean predictors  $\hat{\pi}$  for a response  $Y$  and feature variables  $\mathbf{X}$  are such that  $\hat{\pi}(\mathbf{X}) = E[Y | \hat{\pi}(\mathbf{X})]$ . Autocalibration is beneficial since it ensures that the information contained in  $\hat{\pi}$  is used without any systematic errors. Or stated differently, an autocalibrated predictor is optimal with respect to the information contained in it. Autocalibration can easily be implemented using the practical method proposed by Denuit, Charpentier and Trufin (2021), consisting in an extra local regression step. Professional practice favors Lorenz curves and Gini coefficients to assess the performances of a candidate premium. Under autocalibration, Denuit and Trufin (2021) established that the advanced diagnostic tools proposed by Denuit, Sznajder and Trufin (2019) reduce to Lorenz curve and Gini coefficients, respectively. The present note aims to assess the impact of autocalibration on the goodness of lift. Lift graphs compare average predicted and average actual loss costs when policies have been ranked according to the candidate premium and grouped within buckets. It is shown on a case study that autocalibration does not only restore global and local balances but also improve lift.

**Keywords:** Risk classification, Ratemaking, Lift curve, Goodness of lift, Machine learning.

# Contents

<b>1</b>	<b>Introduction and motivation</b>	<b>1</b>
<b>2</b>	<b>Ratemaking problem</b>	<b>2</b>
<b>3</b>	<b>Autocalibration</b>	<b>3</b>
3.1	Concept . . . . .	3
3.2	Autocalibrating a given predictor . . . . .	4
3.3	Local polynomial regression . . . . .	5
<b>4</b>	<b>Application to Poisson regression</b>	<b>6</b>
4.1	Poisson deviance . . . . .	6
4.2	Database . . . . .	7
4.3	Simple lift charts . . . . .	8
4.4	Double lift charts . . . . .	10
4.5	Out-of-sample Poisson deviance . . . . .	11
<b>5</b>	<b>Discussion</b>	<b>12</b>

# 1 Introduction and motivation

Boosting techniques and neural networks are particularly effective learning methods. However, the sum of fitted values can depart from the observed totals to a large extent and this often confuses actuarial analysts. To solve this problem, Denuit, Charpentier and Trufin (2021) proposed a new strategy based on the concept of autocalibration (see, e.g., Kruger and Ziegel, 2021) to restore global balance as well as local equilibrium in the spirit of the original method of marginal totals, or MMT in short. Dating back to the 1960s, when North-American actuaries pioneered risk classification with the help of minimum bias methods, the central idea behind MMT is that an acceptable set of premiums should reproduce the experience within sub-portfolios corresponding to each level of meaningful risk factors (like gender or age, for instance) and also the overall experience, i.e. be balanced for each level and in total. In insurance applications, autocalibration induces local balance and imposes financial equilibrium not only at portfolio level but also in any sufficiently large sub-portfolio. This concept thus appears to be particularly appealing in a ratemaking context. Denuit, Charpentier and Trufin (2021) show how to perform autocalibration by adding an extra step implementing MMT with the help of local regression.

The starting point of the present analysis is that global balance is often desirable in insurance studies. In ratemaking, it is important that the sum of estimates does not deviate too much from the sum of actual observations at both the entire portfolio level and also more locally, in meaningful classes of policyholders. The reason is obvious: the sum of the pure premiums must match the claim total as accurately as possible so that the insurance company is able to indemnify all third-parties and beneficiaries in execution of the contracts, without excess nor deficit, by the very definition of pure premium (expense loadings and cost-of-capital charges are added into the calculation at a later stage, when moving to commercial premiums). This naturally translates into a global balance constraint: considering that the total claim figures are representative of next-year's experience, it is important that the sum of fitted premiums matches the sum of insurance losses (taken as proxy for the total premium income), as closely as possible. But local equilibrium is also essential to guarantee a competitive pricing.

Flexible pricing tools like boosting and neural networks are able to produce premiums that correlate a lot with the response compared to models imposing a rigid form for the predictor (like Generalized Linear Models or GLMs, for instance). But more flexible models often do not impose any constraint on the replication of the observed totals. Thus, breaking the overall balance is the price to pay for this higher correlation. Autocalibration can then be used to restore global balance as well as local equilibrium in the spirit of the original MMT. This simple and effective solution to the problem is implemented by adding an extra step imple-

menting MMT within a local GLM analysis. Specifically, after the analysis has been performed with a method that does not necessarily respect marginal totals, a local constant GLM fit is achieved in order to restore the connection with MMT. The extra step should not be performed on the same data in order to avoid a potential source of overfitting (called nested-model bias). By using a local constant, or intercept-only GLM, these predictors define optimal neighborhoods to perform local averaging of observed losses.

The present note aims to explore the impact of autocalibration on lift charts that are often used in practice to assess the performances of a candidate premium. To draw such graphs, the data set is sorted based on the values of fitted premiums. The data are then bucketed into equally populated classes based on quantiles. The average predicted and average actual loss costs are then graphed for each bucket. This graph allows the analyst to check 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 in Denuit, Sznajder and Trufin (2019) identifies the condition required to observe an increasing trend in such a lift chart (that is, positive regression dependence). We know that autocalibration purposes to restore global and local balance but its impact on goodness of lift has not been assessed so far. Intuition suggests that autocalibration should also improve on lift and the present study aims to confirm this guess on the basis on an analysis of a classical motor insurance data set. Precisely, the case study performed in this note suggests that autocalibration is also beneficial in terms of lift.

The remainder of the text is structured as follows. Section 2 sets up the scene by recalling the ratemaking context. Section 3 presents the concept of autocalibration and the way it can be implemented. The case study in Section 4 assesses the effect of autocalibration on goodness of lift using a reference motor insurance database. The final Section 5 briefly concludes.

## 2 Ratemaking problem

Consider a claim-related response  $Y$  (frequency, severity or total losses) and a set of features  $X_1, \dots, X_p$  gathered in the vector  $\mathbf{X}$ . The dependence structure inside the random vector  $(Y, X_1, \dots, X_p)$  is exploited to extract the information contained in  $\mathbf{X}$  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  $\mu(\mathbf{X}) = \mathbb{E}[Y|\mathbf{X}]$  of the response  $Y$  given the available information  $\mathbf{X}$ . Henceforth,  $\mu(\mathbf{X})$  is referred to as the true (pure) premium.

Notice that the function  $\mathbf{x} \mapsto \mu(\mathbf{x}) = \mathbb{E}[Y|\mathbf{X} = \mathbf{x}]$  is generally unknown to the actuary, and may exhibit a complex behavior in  $\mathbf{x}$ . This is why this function is approximated by a (working, or actual) premium  $\mathbf{x} \mapsto \pi(\mathbf{x})$ . Sometimes, the work-

ing premium has a relatively simple structure compared to the unknown regression function  $\boldsymbol{x} \mapsto \mu(\boldsymbol{x})$ . This is the case when actuaries resort to Generalized Linear Models (GLMs) or Generalized Additive Models (GAMs) which are very transparent but highly constrained. Boosting for instance is more flexible and better captures the structure of the true pure premium, but at the cost of transparency. It is therefore important to be able to assess the quality of a candidate premium. This can be achieved using the pair  $(\mu(\mathbf{X}), \pi(\mathbf{X}))$  so that we are back to the bivariate case even if there were thousands of features comprised in  $\mathbf{X}$ .

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  $\mu(\mathbf{X})$  given the available features  $\mathbf{X}$ . Notice that the response  $Y$  itself does not play a direct role in the determination of the premium (beyond the definition of  $\mu(\mathbf{X})$  and the calibration of the supervised regression model delivering the prices  $\pi(\mathbf{X})$ ), as departures  $Y - \mu(\mathbf{X})$  are supposed cancel out when averaged over a sufficiently large portfolio (this is the very essence of insurance). The premium  $\pi(\mathbf{X})$  has to be as close as possible to the true premium  $\mu(\mathbf{X})$ . We refer the reader to Meyers 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  $\hat{\pi}(\mathbf{X})$  of unobserved  $\mu(\mathbf{X})$ .

For many models considered in insurance studies, the features  $\mathbf{X}$  are combined to form a score  $S$  (i.e. a real-valued function of the features  $X_1, \dots, X_p$ ) and the premium is a monotonic, say increasing, function of  $S$ . Formally,  $\pi(\mathbf{X}) = h(S)$  for some increasing function  $h$ . This is the case with GLMs, GAMs and boosting. 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. Also, double GLMs and GAMs for location, scale and shape parameters (or GAMLSS) include several scores. We refer the interested readers to Denuit, Hainaut and Trufin (2019) for a general presentation of these models. As these scores are all functions of the available features, we keep here the notation  $\pi(\mathbf{X})$  in the remainder of this text, to remain as general as possible.

## 3 Autocalibration

### 3.1 Concept

Assume that  $\hat{\pi}$  has been built from some training set (all formulas in the text are meant given this training set). The respective performances of competing models can then be assessed on the basis of a validation set  $\{(Y_i, \mathbf{X}_i), i = 1, 2, \dots, n\}$ ,

that has not been used to obtain  $\hat{\pi}$ . Provided the validation set comprises a large number of independent and identically distributed pairs  $(Y_i, \mathbf{X}_i)$ , this allows us to perform calculations with a generic random vector  $(Y, \mathbf{X})$ , independent of, and distributed as the observations contained in the training set. This approach is meaningful in insurance ratemaking where the actuary is typically in a data-rich situation.

The merits of a given pricing tool can be assessed using the pair  $(\mu(\mathbf{X}), \hat{\pi}(\mathbf{X}))$  so that we are back to the bivariate case even if there were thousands of features comprised in  $\mathbf{X}$ . To ease the exposition, we assume that predictor  $\hat{\pi}(\mathbf{X})$  under consideration, as well as the conditional expectation  $\mu(\mathbf{X})$  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  $\mathbf{X}$  and the function  $\hat{\pi}$  is a continuously increasing function of a real score built from  $\mathbf{X}$ .

What really matters is the correlation between  $\hat{\pi}(\mathbf{X})$  and  $\mu(\mathbf{X})$  but as  $\mu(\mathbf{X})$  is unobserved the actuary can only use its noisy version  $Y$  to reveal the agreement of the true premium  $\mu(\mathbf{X})$  with its working counterpart  $\hat{\pi}(\mathbf{X})$ . In insurance applications,  $\hat{\pi}(\mathbf{X})$  is supposed to be used as a premium so that correlation is important, but it is also essential that the sum of predictions  $\hat{\pi}(\mathbf{X})$  matches the sum of actual losses as closely as possible. This naturally leads to the concept of autocalibration. Recall that a predictor  $\hat{\pi}$  is said to be autocalibrated if  $\hat{\pi}(\mathbf{X}) = \mathbb{E}[Y|\hat{\pi}(\mathbf{X})]$ . We refer the reader to Kruger and Ziegel (2021) for a general presentation of this concept. Autocalibration ensures that the average response in a neighborhood defined with the help of the autocalibrated predictor under consideration matches premium income. This exactly corresponds to the MMT and is thus desirable for any candidate premium.

### 3.2 Autocalibrating a given predictor

There is no guarantee that the candidate premium  $\hat{\pi}$  satisfies autocalibration. Denuit, Charpentier and Trufin (2021) proposed an effective method to implement autocalibration in ratemaking studies. If  $s \mapsto \mathbb{E}[Y|\hat{\pi}(\mathbf{X}) = s]$  is continuously increasing, a simple way to restore global balance consists in switching from  $\hat{\pi}$  to its balance-corrected version  $\hat{\pi}_{\text{BC}}$  defined as

$$\hat{\pi}_{\text{BC}}(\mathbf{X}) = \mathbb{E}[Y|\hat{\pi}(\mathbf{X})]$$

that averages to  $\mathbb{E}[Y]$ , as shown in Property 5.1 in Denuit, Charpentier and Trufin (2021). This shows that  $\hat{\pi}$  can easily be autocalibrated by performing an extra step implementing MMT within a local GLM analysis. Specifically, after the analysis has been performed with a method that does not necessarily respect marginal

totals, a local constant GLM fit is achieved on the pairs  $(Y, \hat{\pi}(\mathbf{X}))$  in order to restore the connection with MMT. This produces the autocalibrated version  $\hat{\pi}_{\text{BC}}$  of  $\hat{\pi}$ . In the intercept-only GLM fitted locally, a rectangular weight function is used and premiums are corrected on the basis of a set of observations close to the one of interest (with proximity assessed with the help of the predictor to be corrected). The canonical link function is adopted so that maximum-likelihood estimates replicate observed totals.

Restoring balance through autocalibration reconciles the predicted and observed total on the entire training set as well as locally, in meaningful neighborhoods. In this way, we recover the balance properties imposed after the seminal works introducing MMT in the 1960s.

### 3.3 Local polynomial regression

A local polynomial regression approach allows the actuary to recover balance properties imposed in MMT. Recall that global balance (that is, equality of predicted and observed totals on the whole training) is one of the GLM likelihood equations under canonical link, corresponding to the intercept. The other likelihood equations in the GLM impose balanced at local levels. In order to implement these constraints to general machine learning procedures, we need to mimic the way local GLM proceeds for fitting, by defining meaningful neighborhoods for statistical learning.

To this end, the only feature entering the analysis is the premium  $\hat{\pi}$  needing autocalibration, so that the local GLM is fitted to  $(Y_i, e_i, \hat{\pi}(\mathbf{x}_i))$  comprised in the validation data set, for some relevant exposure  $e_i$ . An intuitively acceptable solution would consist in imposing marginal constraints on local neighborhoods defined by mean of  $\hat{\pi}$ . This allows for some local transfers of claims and premiums from neighboring policyholders and so implements local balance conditions in sub-portfolios corresponding to these neighborhoods. This is in essence the local GLM approach (see Loader, 1999, for a detailed account) that allows the actuary to maintain the relationship with MMT.

In order to obtain an autocalibrated version  $\hat{\pi}_{\text{BC}}$  of  $\hat{\pi}$ , let us consider a specific risk profile  $\mathbf{x}$ . Uniform weights are assigned to each  $(Y_i, e_i, \hat{\pi}(\mathbf{x}_i))$  in the neighborhood  $\mathcal{V}(\mathbf{x})$  of  $\mathbf{x}$ . Thus, a rectangular (or uniform) weight function is specified. The neighborhood  $\mathcal{V}(\mathbf{x})$  gathers  $\alpha\%$  of the data with the closest premiums  $\hat{\pi}(\mathbf{x}_i)$  to  $\hat{\pi}(\mathbf{x})$ . Here,  $\alpha$  acts as a smoothing parameter and controls the size of sub-portfolios where local balance is imposed. The optimal value for the parameter  $\alpha$  is selected by likelihood cross-validation.

The local GLM likelihood equation

$$\sum_{i \in \mathcal{V}(\mathbf{x})} y_i = \sum_{i \in \mathcal{V}(\mathbf{x})} e_i \hat{\pi}_{\text{BC}}(\mathbf{x})$$

thus matches MMT constraints: smoothing is ensured by transferring part of the experience at neighboring  $\hat{\pi}$  values to obtain  $\hat{\pi}_{\text{BC}}$ . A local constant, or intercept-only GLM thus implements local balance, or MMT in sub-portfolios gathering policyholders with about the same predicted value. The rectangular weight function involved in the statistical procedure optimally transfers part of claim experience between neighboring policyholders. This approach implements smoothness from a statistical point of view while remaining fully transparent and understandable. Indeed, local averaging can just be seen as an application of the mutuality principle at the heart of insurance.

## 4 Application to Poisson regression

### 4.1 Poisson deviance

Poisson deviance is by far the most widely used one in insurance applications. It applies for instance to claim counts in property and casualty insurance, death counts in life insurance, and numbers of transitions in health insurance. We assume in this section that we deal with a response  $Y$  obeying the Poisson distribution. Models can be compared on the basis of predictive deviance (or out-of-sample deviance). Denuit, Charpentier and Trufin (2021) established that minimizing predictive deviance can be achieved by

- (i) Increasing the overall bias measured on the response scale in the case of Poisson regression. As a consequence, adopting Poisson deviance as loss function outside GLMs may create a total premium income gap. This gap may become quite large with highly flexible models such as neural networks or boosting.
- (ii) Whereas increasing bias acts against the conservation of observed totals, the deviance can also be improved by increasing the dependence between  $\hat{\pi}$  and the response, so to decrease lower partial moments. In an insurance ratemaking context, the lower partial moment can be interpreted as best-profile premium income, that is, premium income for the sub-portfolio formed by gathering all policyholders with predicted premium at most equal to some threshold. These policyholders exhibit the best risk profiles according to the candidate premium  $\hat{\pi}$ .

Outside GLMs, Poisson regression minimizing deviance is likely to produce candidate premiums needing autocalibration. This is demonstrated in the remainder of this section by a case study.

## 4.2 Database

Let us consider a data set corresponding to a French motor third-party liability insurance portfolio available in the `CASdatasets` package in R. In particular, we consider the data set called `freMTPL2freq` which contains 678 013 observations of the number of claims (response  $Y$ ) together with nine explanatory variables ( $\mathbf{X} = (X_1, \dots, X_9)$ ). The explanatory variables correspond to several characteristics of

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

Moreover, the exposure-to-risk is also available for each policy, here the duration of observation expressed in year. It is important to mention that our continuity assumption was only for  $\mu(\mathbf{X})$  and  $\pi(\mathbf{X})$ . The response  $Y$  may well be discrete (integer-valued in the case of the number of claims considered here). This data set has been used with different statistical and machine learning techniques by several authors, including Noll et al. (2018). We refer the reader to this case study for an accurate description of the data.

By setting expected claim severities all equal to 1 (thus, adopting the mean severity as monetary unit), we can interpret the expected claim frequencies as pure premiums. We consider three candidate premiums corresponding to the following models in our case study:

- $\pi_{\text{glm}}$  obtained from a Poisson GLM with a log-link function and all explanatory variables.
- $\pi_{\text{gam}}$  obtained from a Poisson GAM with a log-link function and all explanatory variables. Functions optimally transforming continuous features for inclusion in the score are estimated using P-splines.
- $\pi_{\text{boost}}$  obtained from a boosted Poisson model. The `h2o` package is used for boosting, precisely the `h2o.gbm` function.

These models are fitted to the database under consideration as described on Arthur Charpentier’s github page [github.com/freakonometrics/autocalibration](https://github.com/freakonometrics/autocalibration).

To perform our analysis, the database has been divided into 3 sets :

1. The training set : 60% of the data are used to train the “base” model (GLM, GAM or Boosting).
2. The validation (or correcting) set : 20% of the data are used to calibrate the local GLM applied in the auto-calibration procedure. This represents 135 603 policies.

3. The test (or final) set : 20% of the data are kept to compare the different models on data that have not been used to train or calibrate the models.

All the graphs discussed hereafter are based on the data contained in the test set. The optimal value for  $\alpha$  is selected with the help of the LCV plot. This is performed with the help of the `lcvplot` function of the `locfit` package implementing local polynomial regression.

### 4.3 Simple lift charts

These lift charts are described in Tevet (2013). See also the CAS Study Note “Model Validation and Handout Data”. To draw such graphs, the data set is sorted based on the values of  $\hat{\pi}(\mathbf{X})$ . 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  $\hat{\pi}$  as well as the actual loss cost  $Y$ . The average predicted and average actual loss costs are then graphed for each class. Precisely, the procedure is as follows:

1. For each observation in the test set, compute the premiums for the model before and after auto-calibration.
2. The observations are ranked based on their premiums from the lowest to the highest. We will then have 2 data sets: one ranked according to the premiums before auto-calibration and one according to the premiums after auto-calibration.
3. The ranked observations are segmented into a number of bands of equal exposure. In our case, we use 10 bands.
4. For each band, compute the average of the premiums and compute the average of the observed values.
5. Plot the results with on the x-axis the bands and on the y-axis the average premium amounts and the average observed values. To make it more readable and interpretable, we normalized the values by dividing them by the average predicted value of the whole test set.

The results for Poisson GLM are displayed in Figure 1. It is important to realize that bands on each graph depend on the premiums produced by each model and so do not contain the same data. For each chart, as expected, the average predicted value increase from left to right as we have ranked the data from lowest to highest. As the average observed values follows the same pattern as the average predicted value and both curves nearly coincide, this suggests that both models

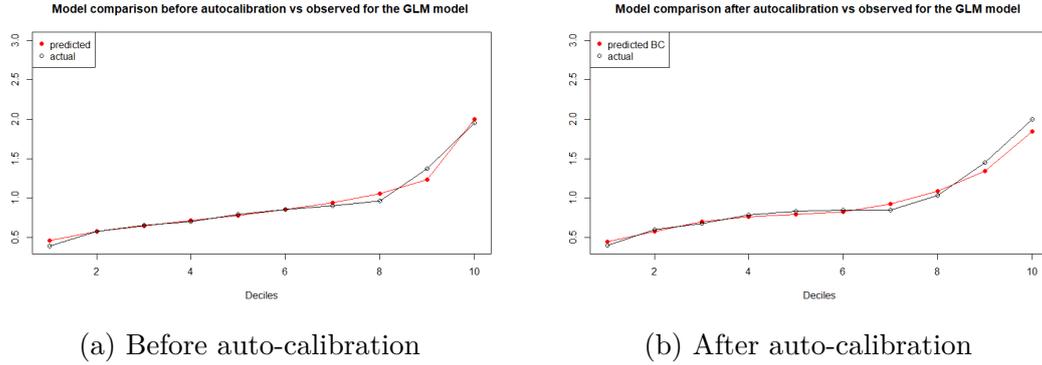


Figure 1: Lift chart for the Poisson GLM, before and after autocalibration.

(before and after auto-calibration) predict on average accurately the observed values. The lifts (value for last band - value for first band) for the models before and after auto-calibration are respectively equal to 1.53 and 1.4. Figure 1 shows that autocalibration has a very limited impact on  $\hat{\pi}_{\text{glm}}$ . This was expected since Poisson GLMs with log link impose local and global balance constraints.

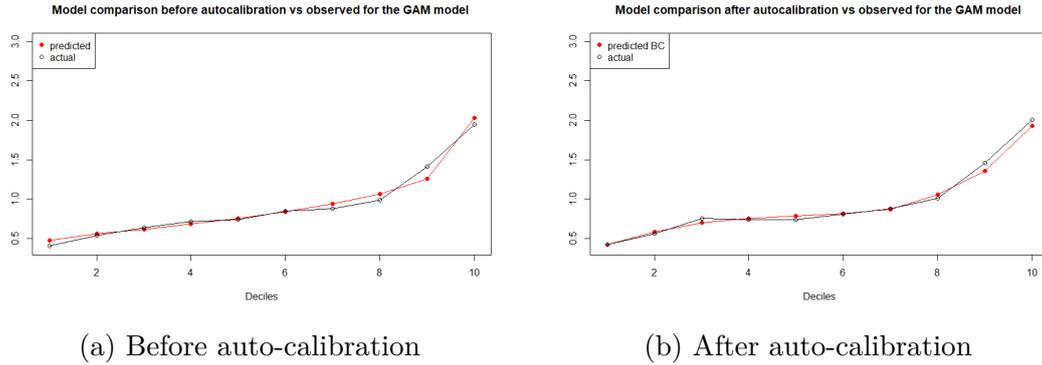


Figure 2: Lift chart for the Poisson GAM, before and after autocalibration.

The results for the Poisson GAM are displayed in Figure 2. The results are very similar to those obtained with the Poisson GLM and the same comments apply. Again, autocalibration has a very limited impact on  $\hat{\pi}_{\text{gam}}$ . The lifts for the models before and after auto-calibration are respectively equal to 1.55 and 1.51. The results for boosted Poisson regression are displayed in Figure 3. Marked differences before and after autocalibration are clearly visible there so that autocalibration greatly impacts on  $\hat{\pi}_{\text{boost}}$ . Before autocalibration, the boosting model under-estimates on average the observed values. This can be seen as the red curve is below the black curve for all bands on Figure 3a. This difference increases when

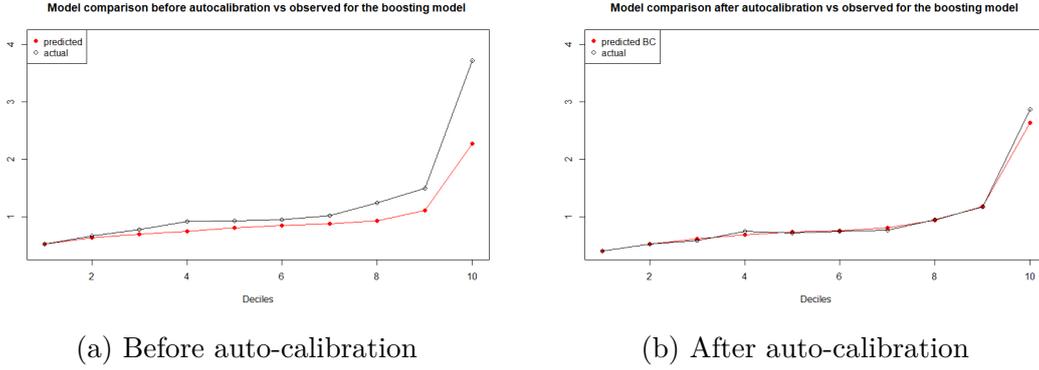


Figure 3: Lift chart for the boosted Poisson regression model, before and after autocalibration.

the premiums become higher. Autocalibration corrects for this bias. On Figure 3b, both the model curve and the observed values curve coincide. Only for the last bucket we do see a small difference. The lifts for the model before and after autocalibration are equal to 1.75 and 2.23, respectively. Once autocalibration has been implemented,  $\hat{\pi}_{\text{boost}}$  clearly outperforms  $\hat{\pi}_{\text{glm}}$  and  $\hat{\pi}_{\text{gam}}$  in terms of goodness of lift.

#### 4.4 Double lift charts

Simple lift charts compared each model to the observed values but it would be interesting to have both models (before and after calibration) on a same graph. That is where double lift charts are helpful. The procedure is the following :

1. For each observation in the test set, compute the premiums for the model before and after auto-calibration.
2. Compute the ratio of the premiums between the two models.
3. The observations are ranked based on their ratio from the lowest to the highest.
4. The ranked observations are segmented into a number of bands of equal exposure. In our case, we use 10 bands.
5. For each band, compute the average of the premiums for each model and compute the average of the observed values.
6. Plot the results with on the x-axis the bands and on the y-axis the average predicted and observed values. To make it more readable and interpretable,

we normalized the values by dividing the average predicted values for each bands by the average predicted value on the whole test set and the observed values by the average observed value on the whole test set.

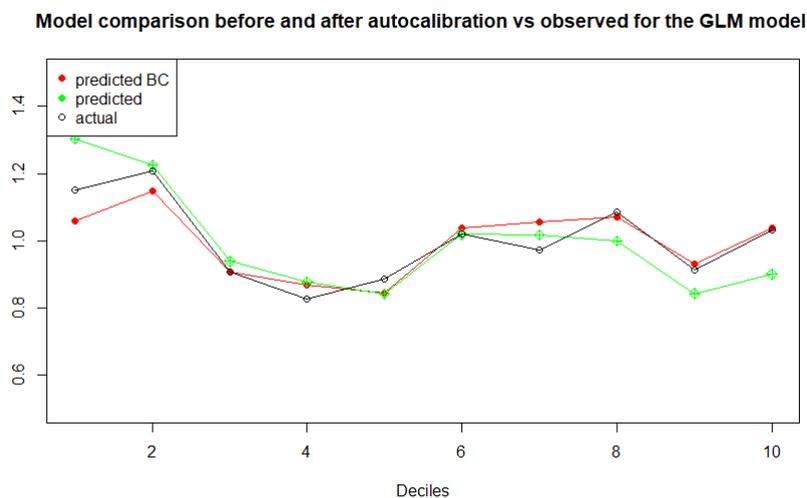


Figure 4: Double lift chart for the Poisson GLM.

Figure 4 displays the double lift chart for the Poisson GLM. For the last 3 bands (high value ratios), corresponding to observations where autocalibrated model predicts much larger value compared to non-autocalibrated model, the autocalibrated GLM model looks to outperforms the non-autocalibrated GLM model. The autocalibrated model curve and the average observed curve coincide but the non-autocalibrated GLM model under-estimates. Apart from that, both models looks quite similar in terms of average predictions. The same comments apply to the double lift chart for the Poisson GAM displayed in Figure 5.

Figure 6 displays the double lift chart for the boosted Poisson regression model. We can see the effect of the autocalibration. In the first and the last 2 bands where the ratio of the premium after and before autocalibration are the smallest and the largest, the autocalibration corrects the premium to bring it closer to the observed premium.

## 4.5 Out-of-sample Poisson deviance

In order to compare the different models, one can also use the out-of-sample, or predictive deviance. The smaller this deviance, the “better” the model. Figure 7 displays predictive deviance for the Poisson GLM, the Poisson GAM, and the

Model comparison before and after autocalibration vs observed for the GAM model

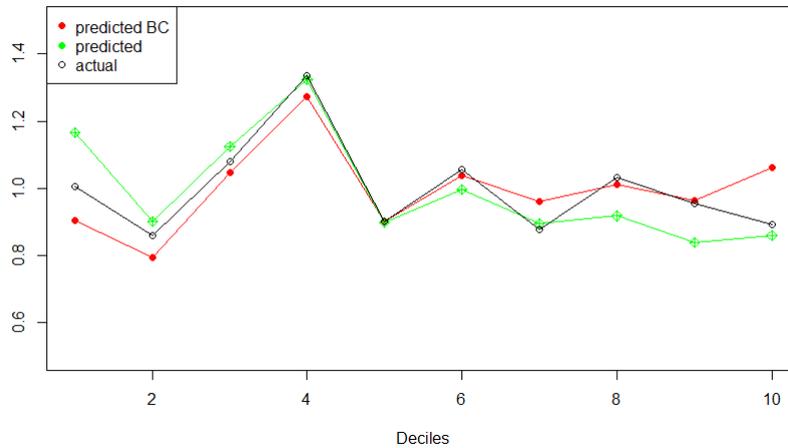


Figure 5: Double lift chart for the Poisson GAM.

boosted Poisson regression model, before and after autocalibration. We can see there that GAM does not outperform GLM on the database under consideration, whereas boosting achieves better performances. For the GLM and GAM models, there is no difference in deviance before or after the autocalibration procedure. For the boosted Poisson regression model, deviance is greatly reduced by applying the autocalibration method. This may be due to the fact that autocalibration has a much greater impact on boosting and that observed totals are more stable across subportfolios. Autocalibration can therefore be expected to deliver better fits on test data sets, as empirically established in the present study.

## 5 Discussion

Advanced learning models are able to produce scores that better correlate with the response, as well as with the true premium compared to classical GLMs. This comes from the additional freedom obtained by letting scores depend in a flexible way of available features, not only linearly. But breaking the overall balance is the price to pay for this higher correlation. Because no constraint on the replication of the observed total, or global balance is imposed, machine learning tools are also able to substantially increase overall bias.

To prevent this to occur, the balance-corrected version of any predictor can be obtained by local GLM, recognizing the nature of the response  $Y$ : local Poisson GLM for claim counts, local Gamma GLM for average claim severities and compound Poisson sums with Gamma-distributed terms for claim totals. The canonical link

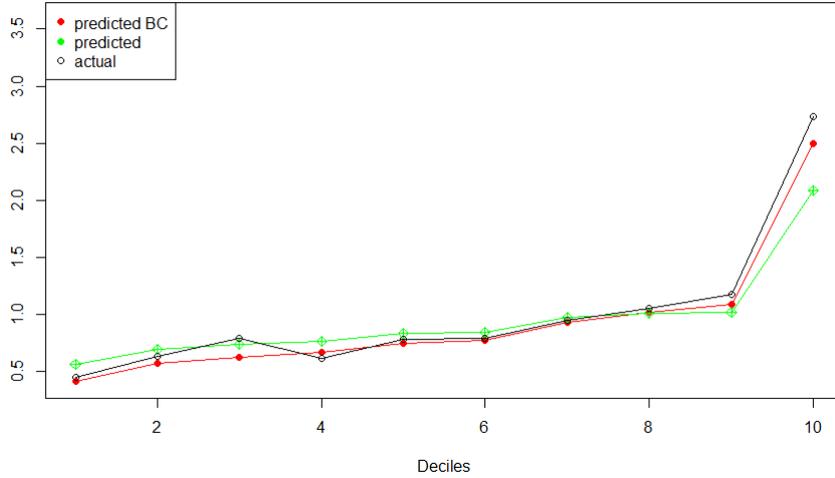


Figure 6: Double lift chart for boosted Poisson regression model.

function is adopted so that maximum-likelihood estimates replicate observed totals. An intercept-only GLM is fitted locally, with rectangular weight function, on a set of observations close to the one of interest with proximity assessed with the help of the predictor to be corrected.

In this approach, the supervised learning model is used to produce a real-valued signal  $\hat{\pi}(\mathbf{X})$ , reducing the high-dimensional feature space to the real line. In fact, the score of the model is enough to compute  $\hat{\pi}_{BC}$  so that the learning model is essentially used to assess proximity among individuals, before performing local averaging. In that respect, the proposed approach shares some similarities with  $k$ -NN, or  $k$  nearest-neighbors except that here, proximity is assessed with the help of the real-valued  $\hat{\pi}$  produced by the learning model.

The case study performed in this note shows that autocalibration not only restores global and local balance, but also improve performances measured in terms of lift induced by the candidate premium. This suggests that autocalibration is a key ingredient of the ratemaking process with advanced machine learning tools, reconciling modern tools with old actuarial MMT recipes.

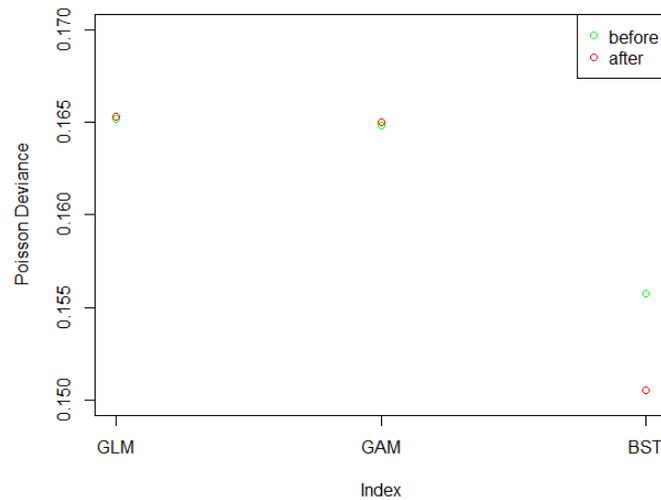


Figure 7: Predictive Poisson deviance before and after autocalibration for the GLM, GAM and boosted regression model.

## References

- [1] M. Denuit and J. Trufin. Lorenz curve, gini coefficient, and tweedie dominance for autocalibrated predictors. *LIDAM Discussion Paper ISBA 2021 / 36*, available from [dial.uclouvain.be](http://dial.uclouvain.be), 2021.
- [2] M. Denuit, D. Hainaut, and J. Trufin. *Effective Statistical Learning Methods for Actuaries. Volume 1: GLM and Extensions*. Springer Actuarial Lecture Notes Series, 2019.
- [3] M. Denuit, D. Sznajder, and J. Trufin. Model selection based on lorenz and concentration curves, gini indices and convex order. *Insurance: Mathematics and Economics*, 89:128–139, 2019.
- [4] M. Denuit, A. Charpentier, and J. Trufin. Autocalibration and tweedie-dominance for insurance pricing with machine learning. *Insurance: Mathematics and Economics*, 101:485–497, 2021.
- [5] F. Kruger and J. F. Ziegel. Generic conditions for forecast dominance. *Journal of Business & Economic Statistics*, 39:972–983, 2021.
- [6] C. Loader. *Local Regression and Likelihood*. Springer, New York, 1999.

- [7] G. Meyers and A. D. Cummings. “goodness of fit” vs. “goodness of lift”. *Actuarial Review*, 36:16–17, 2009.
- [8] A. Noll, R. Salzmann, and M. Wüthrich. Case study: French motor third-party liability claims. *Available at SSRN: <https://ssrn.com/abstract=3164764>*, 2018.
- [9] D. Tevet. Exploring model lift: Is your model worth implementing. *Actuarial Review*, 40:10–13, 2013.

## About the serie and the authors

### The FAQctuary's

The FAQctuary's are a series of educational papers dedicated to the insurance sector. Each issue addresses a specific actuarial topic, expressed as a question asked by market players. FAQctuary's are published by members of the Detralytics team and written in a clear and accessible language. The team combines academic expertise and business knowledge. Detralytics was founded to support companies in the advancement of actuarial science and the solving of the profession's future challenges. It is within the scope of this mission that we make our work available through our Detra Notes and FAQctuary's series.

All our publication are available at: <https://detralytics.com/our-publications/>

### Authors' biographies

#### Harrison Verelst

Harrison is part of the Talent Consolidation Program (TCP). Prior to joining Detralytics, Harrison worked at Belfius (BE) as Quantitative Risk Analyst and at QBE Re as Analytics analyst. His tasks ranged from the review of model risk provisions to technical support to Non-Life underwriting department and development of pricing tools. Harrison holds two Master's degrees in Mechanical Engineering and Quantitative Finance and he is currently in the second year of the Actuarial Science Master at ULB.

#### Nicolas Ciatto

Nicolas is part of the Talent Accelerator Program (TAP). Prior to joining Detralytics, Nicolas did an actuarial internship at Axa Belgium, in corporate healthcare. He worked on the implementation of Non-Life methods in the sight of a priori and a posteriori pricing. Nicolas holds a Master's degree in Actuarial Science from ULB (BE). His thesis focused on credibility methods.

#### Michel Denuit

Michel is Scientific Director at Detralytics, as well as a Professor in Actuarial Science at the Université Catholique de Louvain. Michel has established an international career for some two decades and has promoted many technical projects in collaboration with different actuarial market participants. He has written and co-written various books and publications. A full list of his publications is available at : <https://uclouvain.be/en/directories/michel.denuit>.

## **Julien Trufin**

Julien is Scientific Director at Detralytics, as well as a Professor in Actuarial Science at the department of mathematics of the Université Libre de Bruxelles. Julien is a qualified actuary of the Institute of Actuaries in Belgium (IA|BE) and has experience as a consultant, as well as a compelling academic background developed in prominent universities such as Université Laval (Canada), UCL and ULB (Belgium). He has written and co-written various books and publications. A full list of his publications is available at : <https://julien.trufin.web.ulb.be/>.



People drive actuarial innovation