Isotonic regression for variance estimation and its role in mean estimation and model validation

Łukasz Delong

University of Warsaw, Faculty of Economic Sciences

Joint work with Mario Wüthrich, ETH Zürich

Introduction

- Let us start with a general set-up of actuarial and statistical learning in the framework of GLMs,
- We consider a response variable Y with mean and variance:

$$\mathbb{E}[Y|\boldsymbol{x}] = \mu(\boldsymbol{x}, \boldsymbol{\theta}) \qquad \text{and} \qquad \mathrm{Var}[Y|\boldsymbol{x}] = \phi V(\mu(\boldsymbol{x}, \boldsymbol{\theta})),$$

  where $\boldsymbol{x} \in \mathbb{R}^d$ is a d-dimensional feature vector that characterizes the response and $\boldsymbol{\theta} \in \mathbb{R}^d$ is a d-dimensional parameter vector of the functional μ,
- We assume:

$$g(\mu(\boldsymbol{x}, \boldsymbol{\theta})) = \eta(\boldsymbol{x}, \boldsymbol{\theta}) = \boldsymbol{x}^\top \boldsymbol{\theta},$$

  for a given strictly monotone and smooth link function g,
- The goal is to estimate and validate the mean functional $\boldsymbol{x} \mapsto \mu(\boldsymbol{x})$, i.e. the vector of parameters $\boldsymbol{\theta}$, based on $(Y_i, \mathbf{x}_i)_{i=1}^n$ using minimal information about the distribution of $Y_i|\mathbf{x}_i$ and making minimal subjective decisions in estimation process.

Introduction

- The variance function still enters statistical estimation and inference at least at three problems:
  - Asymptotically efficient estimation of mean functionals by the quasi(pseudo)-likelihood method,
  - Back-transformation of maximum likelihood estimates on log scale to original scale for lognormally distributed responses,
  - Tests and validation plots, such as reliability plots (lift plots) for validation of the auto-calibration property of predictors.

Asymptotically efficient estimation of mean functionals by the quasi(pseudo)-likelihood method

Quasi(pseudo)-likelihood method

- In the framework of GLMs, if the distribution of the response is properly specified, and if the distribution is a member of the exponential dispersion family, the parameter $\boldsymbol{\theta} \in \mathbb{R}^d$ is estimated by minimizing the deviance loss function L of the true distribution, i.e. we solve:

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \mathbb{R}^d}{\arg\min} \sum_{i=1}^{n} L\Big(Y_i, \mu(\boldsymbol{x}_i, \boldsymbol{\theta}), V\left(\mu(\boldsymbol{x}_i, \boldsymbol{\theta})\right)\Big),$$

- In practice, it is very unlikely that the response exactly follows the chosen distribution,
- We can apply the quasi(pseudo)-likelihood method – we choose any deviance loss function derived from a particular distribution from the EDF and the parameter $\boldsymbol{\theta} \in \mathbb{R}^d$ is still estimated by minimizing the deviance loss function,
- Since deviance loss functions are strictly consistent scoring functions for mean estimation, the quasi(pseudo)-likelihood estimate is strongly consistent,
- We can gain asymptotic estimation efficiency in the quasi(pseudo)-likelihood method by properly specifying the variance function of the response.

The key problem is to estimate variance function and incorporate its estimate in estimation of mean functional

Quasi(pseudo)-likelihood method

- Estimation of the variance function and its impact on mean estimation is not very popular in an actuarial context where the focus is mainly on GLMs,
- At the same time, the problem is well-known in statistics,
- The results by Chiou, Muller, 1999, *Non-parametric quasi-likelihood estimation*, The Annals of Applied Statistics, show if the variance function is estimated with a local polynomial smoother then the asymptotic limiting distribution of the vector of parameters in the mean functional is the same as for the quasi-likelihood estimates obtained under correct specification of the variance function,
- However, without carefully chosen bandwidth and degree of polynomial good estimates cannot be obtained in the framework of smoothing methods.

Non-parametric quasi(pseudo)-likelihood method

- **Our contribution:** We explore the non-parametric quasi(pseudo)-likelihood method with isotonic regression as a non-parametric variance estimator – we validate isotonic regression as an alternative to local regression,
- We believe it is simple, requires minimal assumptions and no intervention in variance estimation (no bandwidth and no hyperparameters selection), and can be applied in many real-life actuarial problems.

- In order to apply isotonic regression we work under the assumption that $\mu \mapsto V(\mu)$ is increasing,
- In the actuarial science the Tweedie family with a strictly increasing variance function $V(\mu) = \mu^p, \ p \geq 1$, plays a dominant role and fits many real-life application.

Non-parametric quasi(pseudo)-likelihood method

- There are other possible approaches to estimate the variance function of the response, or even the whole distribution of the response, e.g double GLM, GAMLSS, mixture of experts, cyclic GBM:
  - Delong, Ł., Lindholm, M., Zariksson, H., 2024, *On cyclic gradient boosting machines*,
- <u>However</u>, we have to specify and fit additional regression functions, which might be too sophisticated for an application (image we want to improve Gamma GLM with double GLM for Inverse Gaussian claim amounts), or perform hyperparameters optimization.

Isotonic regression

- Isotonic regression is a rank based non-parametric and monotonic regression approach that preserves monotonicity in pre-specified ranks of original predictor $(\pi(\boldsymbol{x}_i))_{i=1}^n$,
- Assume we have data points $(Y_i, \pi(\boldsymbol{x}_i))_{i=1}^n$ and positive case weights $(w_i)_{i=1}^n$. Would like to fit a non-parametric regression model to the responses $(Y_i)_{i=1}^n$ which respects the ranks $(\pi(\boldsymbol{x}_i))_{i=1}^n$,
- We consider the following estimation problem:

$$\widehat{\boldsymbol{\mu}} = \underset{\boldsymbol{\mu}=(\mu_1,\ldots,\mu_n)^\top \in \mathbb{R}^n}{\arg\min} \sum_{i=1}^n w_i \left(Y_i - \mu_i\right)^2,$$

$$\text{subject to } \mu_j \leq \mu_k \iff \pi(\boldsymbol{x}_j) \leq \pi(\boldsymbol{x}_k) \text{ for all } 1 \leq j, k \leq n.$$

- Estimation problem can be solved using the pool adjacent violators (PAV) algorithm. The resulting solution can also be written as the following min-max formula:

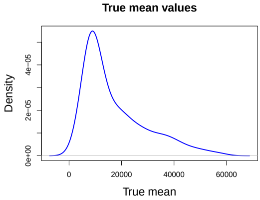
$$\widehat{\mu}_i = \min_{j=i,\ldots,n} \max_{k=1,\ldots,j} \frac{1}{\sum_{l=k}^j w_l} \sum_{l=k}^j w_l Y_l, \quad 1 \leq i \leq n.$$

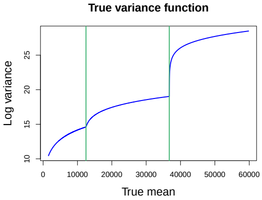
Synthetic data set

- In our experiments we consider claims sizes,
- We use a synthetic data set that was built on the data set `swmotorcycle` from the R package CASdatasets – we know the ground truth about the distributions of the claim sizes and their parameters,
- We estimate a regression function for the claim sizes on the original data and use this regression as the mean functional for our synthetic polices,
- We choose a special variance function which has different regimes for different ranges of the expectations:

$$V(\mu) = \mu^2 \mathbf{1}\{\mu < \mu_1^*\} + \big(c_1 + \mu^2 \log(\mu)\big)\mathbf{1}\{\mu_1^* \leq \mu < \mu_2^*\} + \big(c_2 + \mu^3\big)\mathbf{1}\{\mu \geq \mu_2^*\},$$

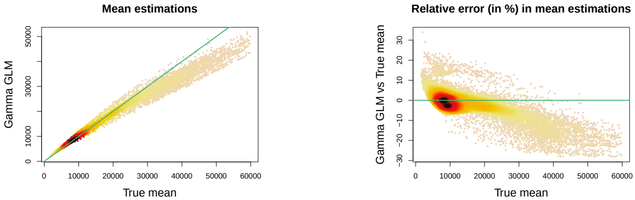
- We simulate $n = 20,000$ claim sizes $(Y_i)_{i=1}^n$ for our policies (medium-sized insurer) from lognormal distributions with the two moments $\big(\mu(\boldsymbol{x}_i), \phi V(\mu(\boldsymbol{x}_i))\big)_{i=1}^n$ specified as above.





Classical Gamma GLM

- We start with a classical actuarial approach,
- We fit a Gamma GLM with quadratic variance function $V(\mu) = \mu^2$ and log-link function $g(\mu) = \log(\mu)$ to our observations,
- Since we know the true means, the fit of the Gamma GLM is very poor in terms of mean estimates – this should give a warning to actuaries that variance function is important,



- We can fit GLMs with other parametric forms of the variance function in the Tweedie family with $V(\mu) = \mu^p$ and we always get poor estimates,
- We can improve our estimates by jointly modelling the mean and the variance function of the response by using isotonic regression for variance estimation.

Quasi GLM with variance estimation

Quasi-likelihood with Non-Parametric Variance Estimation

- **Step 1:** We estimate the expected value of the response with a classical GLM with link function g and variance function V^{GLM} implied by the selected GLM. This provides an initial parameter estimate $\widehat{\boldsymbol{\theta}}^0$ and the estimated mean values $(\widehat{\mu}^0(\boldsymbol{x}_i))_{i=1}^n$ for all instances $1 \le i \le n$.
- **Step 2:** We estimate the variance of the response using the crude estimator

$$\widehat{v}_i^0 = \frac{(Y_i - \widehat{\mu}^0(\boldsymbol{x}_i))^2}{1 - h_i^0},$$

  where $(h_i^0)_{i=1}^n$ denote the hat values of the GLM fitted at the initial step.
- **Step 3:** We estimate the variance function of the response based on the observations $(\widehat{v}_i^0, \widehat{\mu}^0(\boldsymbol{x}_i))_{i=1}^n$. We use an isotonic regression assuming that the true variance function $V(\mu(\boldsymbol{x}))$ is monotonically increasing in $\mu(\boldsymbol{x})$. This step provides us with an estimate of the variance function $\mu \mapsto \widehat{V}^0(\mu)$ and the variances $\widehat{V}^0(\widehat{\mu}^0(\boldsymbol{x}_i))$ for all instances $1 \le i \le n$.
- **Step 4:** We iterate for $k = 1$ to K:

  (i) We estimate the expected value of the response with a GLM with link function g using the quasi-likelihood method. In the deviance loss function we use the estimated variance function $\mu \mapsto \widehat{V}^{k-1}(\mu)$ as the variance function of the EDF family.

  (ii) We get new estimates of $\widehat{\boldsymbol{\theta}}^k$ and $(\widehat{\mu}^k(\boldsymbol{x}_i))_{i=1}^n$ for all instances $1 \le i \le n$.

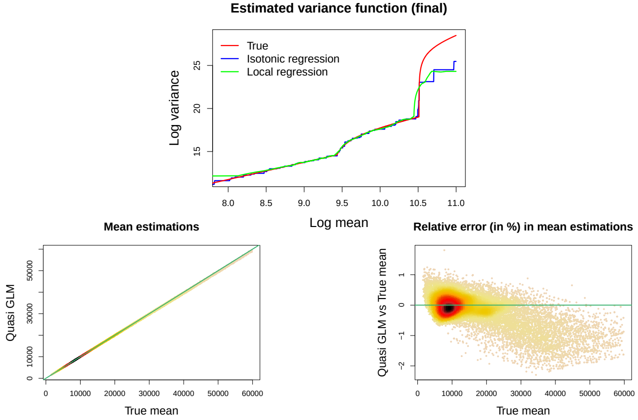
  (iii) We re-estimate the variance of the response using the crude estimator:

$$\widehat{v}_i^k = \frac{(Y_i - \widehat{\mu}^k(\boldsymbol{x}_i))^2}{1 - h_i^k}.$$

  (iv) We re-estimate the variance function of the response based on the observations $(\widehat{v}_i^k, \widehat{\mu}^k(\boldsymbol{x}_i))_{i=1}^n$ using an isotonic regression and assuming monotonicity in the mean estimates. We get new estimates of $\mu \mapsto \widehat{V}^k(\mu)$ and $\widehat{V}^k(\widehat{\mu}^k(\boldsymbol{x}_i))$ for all instances $1 \le i \le n$.

Quasi GLM with variance estimation

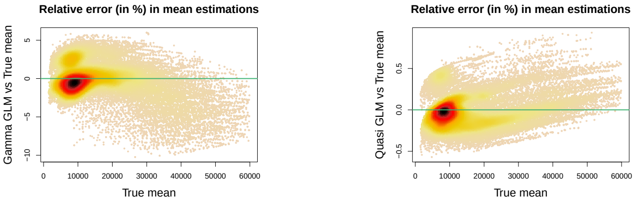
- In terms of variance, the isotonic regression gives accurate results,
- In terms of mean, we gain a major improvement in accuracy. We get small estimation errors of order $+1\%$ and -2%, compared to $+30\%$ and -30% for the Gamma GLM,



- We conclude that on a finite sample we should easily improve mean estimation with quasi-likelihood with non-parametric isotonic variance estimation.

Quasi GLM with variance estimation

- The theory of quasi-likelihood methods says that when the sample size increases, we may misspecify the true variance function, but asymptotically the mean estimates converge to the true means,
- We present the estimation errors for our synthetic data set with an increased sample size of $n = 100,000$,



- Even for a large insurer it should be beneficial to include variance function to achieve accurate mean estimates.

Back-transformation of maximum likelihood estimates on log scale to original scale for lognormally distributed responses

Maximum likelihood for lognormal response

- In our data set the responses follow lognormal distributions – it is natural to perform MLE under the true distribution to gain full efficiency of the mean estimator,
- MLE is performed on logged responses and an exponential transformation is applied to the mean and variance estimates on the log scale to get the mean estimate on the original scale,
- If we misspecify the variance function on the log scale, then the MLE for the mean value on the original scale is no longer strongly consistent – variance estimation is a must here.

The key problem is to estimate variance function on log scale and incorporate its estimate in estimation of mean functional on original scale

- **Our contribution:** We use the results from the first part of this presentation to advise to fit the variance function of logged observations and transform the mean and variance estimates from logged scale to original scale.

Maximum likelihood for lognormal response

- For simplicity we present the result for logarithmic link function,
- We consider a lognormal response variable Y with mean and variance:

$$\mathbb{E}[Y|\boldsymbol{x}] = \mu(\boldsymbol{x}, \boldsymbol{\theta}) = e^{\boldsymbol{x}^\top \boldsymbol{\theta}} \qquad \text{and} \qquad \mathrm{Var}[Y|\boldsymbol{x}] = \phi V(\mu(\boldsymbol{x}, \boldsymbol{\theta})),$$

- If $Y|\boldsymbol{x}$ has a lognormal distribution with mean $\mathbb{E}[Y|\boldsymbol{x}]$ and variance $\mathrm{Var}[Y|\boldsymbol{x}]$, then $Z = \log(Y)$, conditional on $\boldsymbol{x}$, is normally distributed,
- We introduce an auxiliary response variable on the log scale:

$$\widetilde{Z} = Z + \frac{1}{2}\mathrm{Var}[Z|\boldsymbol{x}],$$

with the first two moments:

$$\begin{aligned}
\mathbb{E}[\widetilde{Z}|\boldsymbol{x}] &= \mathbb{E}[Z|\boldsymbol{x}] + \frac{1}{2}\mathrm{Var}[Z|\boldsymbol{x}] = \log(\mu(\boldsymbol{x}, \boldsymbol{\theta})) = \boldsymbol{x}^\top \boldsymbol{\theta} = \eta(\boldsymbol{x}, \boldsymbol{\theta}), \\
\mathrm{Var}[\widetilde{Z}|\boldsymbol{x}] &= \log\left(1 + \frac{\mathrm{Var}[Y|\boldsymbol{x}]}{\left(\mathbb{E}[Y|\boldsymbol{x}]\right)^2}\right) \\
&= \log\left(1 + \frac{V(\mu(\boldsymbol{x}, \boldsymbol{\theta}))}{\left(\mu(\boldsymbol{x}, \boldsymbol{\theta})\right)^2}\right) =: \widetilde{V}\left(\eta(\boldsymbol{x}, \boldsymbol{\theta})\right) = \mathrm{Var}[Z|\boldsymbol{x}],
\end{aligned}$$

where $\widetilde{V}$ is a variance function of the response on the log scale seen as a function of the linear predictor η (a function of the logarithm of the mean value of the response on the original scale),
- We assume that $\eta \mapsto \tilde{V}(\eta)$ is increasing.

Maximum likelihood for lognormal response

Lognormal Linear Model with Non-Parametric Variance Estimation

- **Step 1:** We estimate the linear predictor with a Gaussian GLM with identity link and unit weights based on the observations $(Z_i, \boldsymbol{x}_i)_{i=1}^n$. We get an initial estimate of the parameter $\widehat{\boldsymbol{\theta}}^0$, the linear predictors $(\widehat{\eta}^0(\boldsymbol{x}_i))_{i=1}^n$ and the estimated mean values $(\widehat{\mu}^0(\boldsymbol{x}_i))_{i=1}^n$ for all instances $1 \leq i \leq n$.
- **Step 2:** We estimate the variance of the response on the log scale using the crude estimator:

$$\widehat{\widehat{v}}_i^0 = \frac{\left(Z_i - \widehat{\eta}_i^0\right)^2}{1 - h_i^0},$$

  where $(h_i^0)_{i=1}^n$ denote the hat values of the Gaussian GLM fitted in the initial step.
- **Step 3:** We estimate the variance function of the response on the log scale based on the observations $(\widehat{\widehat{v}}_i^0, \widehat{\eta}^0(\boldsymbol{x}_i))_{i=1}^n$. We use an isotonic regression assuming that the true variance function $\widetilde{V}(\eta(\boldsymbol{x}))$ is monotonically increasing in $\eta(\boldsymbol{x})$. This step gives us an estimate of the variance function $\eta \mapsto \widehat{\widehat{V}}^0(\eta)$ and the variances $\widehat{\widehat{V}}^0(\widehat{\eta}^0(\boldsymbol{x}_i))$ for all instances $1 \leq i \leq n$.

Maximum likelihood for lognormal response

Lognormal Linear Model with Non-Parametric Variance Estimation

- **Step 4:** We iterate for $k = 1$ to K:

  (i) We estimate the linear predictor with a Gaussian GLM with identity link, the weights $\left(\widehat{\widetilde{V}}^{k-1}(\widehat{\mu}^{k-1}(\boldsymbol{x}_i))\right)^{-1}$ and the offset $-\frac{1}{2}\widehat{\widetilde{V}}^{k-1}(\widehat{\eta}^{k-1}(\boldsymbol{x}_i))$ based on the observations $(Z_i, \boldsymbol{x}_i)_{i=1}^n$.

  (ii) We get new estimates of $\widehat{\boldsymbol{\theta}}^k$, $(\widehat{\eta}^k(\boldsymbol{x}_i))_{i=1}^n$ and $(\widehat{\mu}^k(\boldsymbol{x}_i))_{i=1}^n$ for all instances $1 \leq i \leq n$.

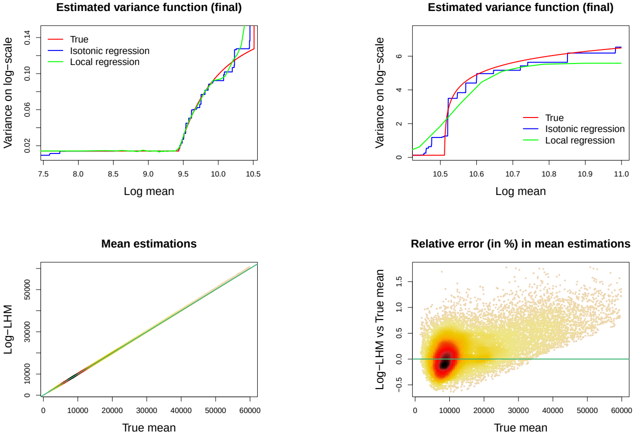
  (iii) We re-estimate the variance of the response on the log scale using the crude estimator:

$$\widehat{\widetilde{v}}_i^k = \frac{\left(Z_i + \frac{1}{2}\widehat{\widetilde{V}}^{k-1}(\widehat{\eta}^{k-1}(\boldsymbol{x}_i)) - \widehat{\eta}^k(\boldsymbol{x}_i)\right)^2}{1 - h_i^k}.$$

  (iv) We re-estimate the variance function of the response on the log scale based on the observations $(\widehat{\widetilde{v}}_i^k, \widehat{\eta}^k(\boldsymbol{x}_i))_{i=1}^n$ using an isotonic regression. We get new estimates of $\eta \mapsto \widehat{\widetilde{V}}^k(\eta)$ and $\widehat{\widetilde{V}}^k(\widehat{\eta}^k(\boldsymbol{x}_i))$ for all instances $1 \leq i \leq n$.

Maximum likelihood for lognormal response

- We arrive at precise estimates of the expected values and the variances of the responses,



- Our algorithm describes how to fit lognormal distributions in a regression setting with variance function depending on mean (an alternative approach would be to fit GAMLSS).

Reliability plots (lift plots) for validation of the auto-calibration property of predictors

Auto-calibration property of predictors

- A regression function $\boldsymbol{x} \mapsto \mu(\boldsymbol{x})$ is auto-calibrated for $(Y, \mathbf{X})$ if

$$\mu(\mathbf{X}) = \mathbb{E}\left[\, Y \,|\, \mu(\mathbf{X})\right],$$

- In actuarial pricing, auto-calibration means that every price cohort $\mu(\mathbf{X})$ is in average self-financing for its claim Y. This implies that there is no systematic cross-financing within the price system $\mu(\cdot)$,
- The goal is to build a mean estimator $\widehat{\mu}$, which accurately approximates the best-estimate price $\mu^* = \mathbb{E}[Y|\mathbf{X}]$, and which fulfills the auto-calibration property.

Auto-calibration property of predictors

- A diagnostic tool for checking auto-calibration of a regression function is a reliability diagram (lift plot),
- In modern statistical learning approach, we compare the estimated mean values $\left(\widehat{\mu}(\boldsymbol{x}_i)\right)_{i=1}^n$ against their recalibrated values $\left(\widehat{\mu}_{\mathrm{rc}}(\boldsymbol{x}_i)\right)_{i=1}^n$, where $\widehat{\mu}_{\mathrm{rc}}$ is constructed from $\widehat{\mu}$ with an isotonic regression,
- Such reliability diagrams are called CORP, since:
  - C: the reliability diagrams and associated numerical measures of miscalibration are consistent in the classical statistical sense of convergence to population characteristics (consistency);
  - O: the reliability diagrams are optimally binned (optimality);
  - R: the reliability diagrams do not require any tuning parameters nor implementation decision (reproducibility); and
  - P: the reliability diagrams are implemented via the PAV algorithm (pool adjacent violators algorithm).
- If the points $\left(\widehat{\mu}(\boldsymbol{x}_i), \widehat{\mu}_{\mathrm{rc}}(\boldsymbol{x}_i)\right)_{i=1}^n$ of a reliability diagram lie close to its diagonal, the mean estimator $\widehat{\mu}$ is auto-calibrated. We would like to formally evaluate the significance of the observed deviations from the diagonal.

Auto-calibration property of predictors

- Murphy's score decomposition of a strictly consistent scoring function gives us:

$$S(\mathbf{Y}, \widehat{\mu}) = \mathrm{UNC}_S(\mathbf{Y}, \widehat{\mu}) - \mathrm{DSC}_S(\mathbf{Y}, \widehat{\mu}) + \mathrm{MCB}_S(\mathbf{Y}, \widehat{\mu}),$$

  with the uncertainty, discrimination and miscalibration statistics:

$$\begin{aligned} \mathrm{UNC}_S(\mathbf{Y}, \widehat{\mu}) &= S(\mathbf{Y}, \bar{\mu}), \\ \mathrm{DSC}_S(\mathbf{Y}, \widehat{\mu}) &= S(\mathbf{Y}, \bar{\mu}) - S(\mathbf{Y}, \widehat{\mu}_{\mathrm{rc}}) \geq 0, \\ \mathrm{MCB}_S(\mathbf{Y}, \widehat{\mu}) &= S(\mathbf{Y}, \widehat{\mu}) - S(v, \widehat{\mu}_{\mathrm{rc}}) \geq 0, \end{aligned}$$

  with $\bar{\mu} = \sum_{i=1}^{n} Y_i/n$ being the empirical mean not considering any features,
- A small value of the miscalibration statistics $\mathrm{MCB}_S(\mathbf{Y}, \widehat{\mu})$ supports the hypothesis that the estimated mean values $\widehat{\mu}$ are auto-calibrated,
- We set the null hypothesis $H_0 : \mathrm{MCB}_S(\mathbf{Y}, \widehat{\mu}) = 0$ against the alternative $H_1 : \mathrm{MCB}_S(\mathbf{Y}, \widehat{\mu}) > 0$ and formally evaluate the significance of a positive miscalibration statistics.

The key problem is to construct a point-wise consistency band for a reliability diagram and derive a critical value for the miscalibration test

- **Our contribution:** We propose techniques based on bootstrap and isotonic regresion for variance estimation for validating the auto-calibration property of mean predictors.

Auto-calibration property of predictors

Semi-Parametric Bootstrap Approach

- We assume that the estimated means $\left(\widehat{\mu}(\boldsymbol{x}_i)\right)_{i=1}^n$ are given,
- **Step 1:** We estimate the variances of $Y_i|\boldsymbol{x}_i$ for all instances $1 \leq i \leq n$ under the estimated means $\left(\widehat{\mu}(\boldsymbol{x}_i)\right)_{i=1}^n$ by an isotonic regression assuming that these variances are monotonically increasing in the means. We fit an isotonic regression to $\left(\widehat{v}_i, \widehat{\mu}(\boldsymbol{x}_i)\right)_{i=1}^n$, where

$$\widehat{v}_i = \frac{(Y_i - \widehat{\mu}(\boldsymbol{x}_i))^2}{1 - h_i},$$

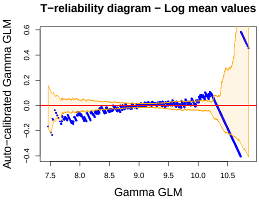
  and $(h_i)_{i=1}^n$ are the hat values from the fitted model. The isotonic regression provides us with variance estimates $\widehat{V}(\boldsymbol{x}_i)$ for all instances $1 \leq i \leq n$ under the estimated means.
- **Step 2:** We calculate the quantile residuals defined by $\widehat{\varepsilon}_i = F\left(Y_i, \widehat{\mu}(\boldsymbol{x}_i), \widehat{V}(\boldsymbol{x}_i)\right)$ for all instances $1 \leq i \leq n$, where we use a two-parametric cumulative distribution function F with the first two (estimated) moments $\left(\widehat{\mu}(\boldsymbol{x}_i), \widehat{V}(\boldsymbol{x}_i)\right)_{i=1}^n$.
- **Step 3:** Under an i.i.d. assumption for the quantile residuals, we bootstrap $(\varepsilon_i^*)_{i=1}^n$ from $(\widehat{\varepsilon}_i)_{i=1}^n$ to receive the bootstrap observations:

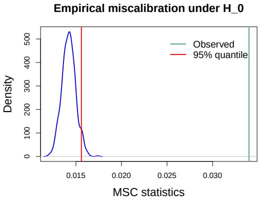
$$Y_i^* = F^{-1}\left(\varepsilon_i^*, \widehat{\mu}(\boldsymbol{x}_i), \widehat{V}(\boldsymbol{x}_i)\right).$$

- **Step 4:** Based on the new sample $\left(Y_i^*, \widehat{\mu}(\boldsymbol{x}_i)\right)_{i=1}^n$, we recalibrate the mean estimator $\widehat{\mu}$ with an isotonic regression, plot the T-reliability diagram and calculate the miscalibration statistics. From repreating this semi-parametric bootstrap, we can construct empirical confidence bands for the T-reliability diagram and find critical values for the miscalibration test, under the null hypothesis that $\left(\widehat{\mu}(\boldsymbol{x}_i)\right)_{i=1}^n$ are the best-estimate prices, hence, auto-calibrated.

Classical Gamma GLM

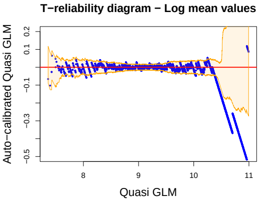
- Using our diagnostic tools, we can conclude that the mean estimates from the classical Gamma GLM are not auto-calibrated,
- The Gamma GLM seems to over-estimate the true means of policyholders with small expected claims, and under-estimate the true means of policyholders with large expected claims – in practice, such a misesimation leads to a "wrong" premium, which in turn may imply adverse selection.

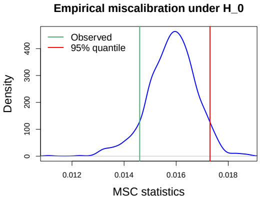




Quasi GLM with variance estimation

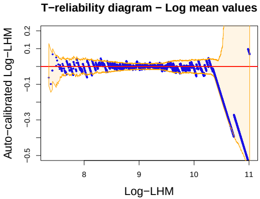
- If we switch to the non-parametric quasi-likelihood estimation with isotonic regression, the results show that the miscalibration observed for the Gamma GLM vanishes for most estimated mean values,
- The Quasi GLM does not have a bias (the mean estimates from the Quasi GLM are auto-calibrated and we cannot reject the null hypothesis of having an auto-calibrated model).

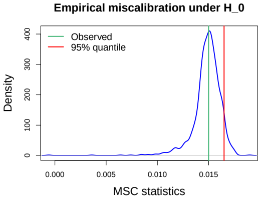




Maximum likelihood for lognormal response

- The mean estimates achieved with our algorithm are auto-calibrated and we cannot reject the null hypothesis of having an auto-calibrated model.





- We believe that our diagnostic tools should be useful in detecting bias and miscalibration of mean predictions.

Conclusion

- We have emphasized the significance and benefits of an accurate variance estimation for mean estimation in regression models,
- We have proposed an isotonic regression as a tool for non-parametric variance estimation as a monotonic function of the mean,
- We have recommended diagnostic tests for validating the auto-calibration property of predictor,
- Our approach with joint mean-variance estimation also applies to GAMs, regression trees, neural networks or any other more complex regression framework,
- For more details, please see the paper: Ł. Delong, M. Wüthrich, 2024, *Isotonic regression for variance estimation and its role in mean estimation and model validation*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4477677.

Thank you very much.

Łukasz Delong
University of Warsaw, Faculty of Economic Sciences
E-mail: l.delong@uw.edu.pl
Homepage: www.lukaszdelong.pl