Tweedie Compound Poisson Linear Models

Ratemaking and Product Management Seminar
Philadelphia, 03/21/2011

Yanwei (Wayne) Zhang
Director
Strategic Research & Economic Modeling
CNA Insurance Company
Yanwei.Zhang@cna.com

CNA

Disclaimer

The views expressed in this presentation are those of the author and do not necessarily reflect the views of CNA Financial Corporation or any of its subsidiaries. This presentation is for general informational purposes only.

Agenda

- Introduction to the Tweedie compound Poisson distribution
  - Construction and simulation of compound Poisson variables
  - Overview of the challenges on statistical inference
  - Investigation of the impact of the index parameter on inferences
  - Description of the data under study
- Compound Poisson linear models
  - Generalized linear models [GLM]
  - Generalized linear mixed models [GLMM]
    - Shrinkage estimates
    - Accounting for within-cohort correlations
  - Generalized additive models [GAM] / penalized splines
    - Specifying smoothing effects vs global linear trends
  - Zero-inflated compound Poisson models [ZICP]
    - Accounting for “bonus hunger”
    - Modeling patterns in the observed frequency of zeros
- Summary and conclusion

The Tweedie compound Poisson distribution

- The goal is to model the aggregate claim amount for a policy term.
- The well-known collective risk model:
  - The sum of an unknown number of individual claims

$$\boxed{Y = \sum_{i}^{T} X_i} \tag{1}$$

  - T is the number of claims, X_i is the loss amount for the i_{th} claim.
- A special case: the Tweedie compound Poisson distribution [CPois]

$$\boxed{T \sim Pois(\lambda),\ X_i \overset{\text{iid}}{\sim} Gamma(\alpha, \gamma),\ T \perp X_i.} \tag{2}$$

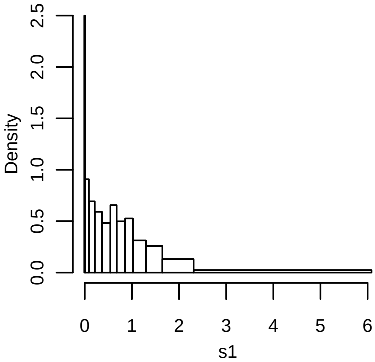
Motivations for employing the *CPois* distribution

- Reasonable assumptions: Poisson frequency and Gamma severity
- Capability to accommodate the aggregate loss distribution: it has a probability mass at zero accompanied by a continuous distribution on the positive values
- Belongs to the exponential dispersion family: $Var(Y) = \phi \cdot \mu^p$
  - $\phi > 0$: dispersion parameter, $p \in (1, 2)$: the index parameter
  - $V(\mu) = \mu^p$: the variance function
  - Various linear model forms can be readily handled for a given p
- The density is intractable, but can be approximated accurately and fast.
  - In general, compound distributions must be evaluated using the less efficient and much slower recursive algorithm.

Simulation of a *CPois* variable (1)

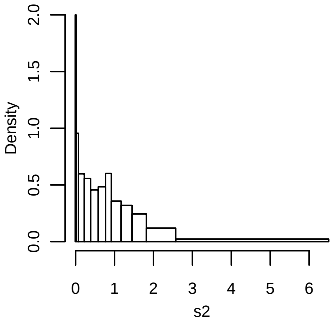
- It is straightforward to simulate from the *CPois* distribution.

```
library(tweedie)
n <- 300
mu <- 1;
phi <- 1;
p <- 1.7
s1 <- rtweedie(n, mu = mu,
        phi = phi, power = p)
```



Simulation of a *CPois* variable (2)

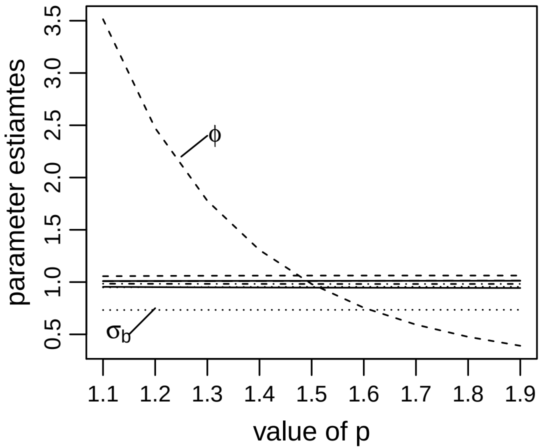
```
lambda <- mu^(2 - p) / (phi * (2 - p))
alpha <- (2 - p) / (p - 1)
gamma <- phi * (p - 1) * mu^(p - 1)
s2 <- sapply(rpois(n, lambda), function(x)
    ifelse(x > 0, sum(rgamma(x, alpha, scale = gamma)), 0))
```



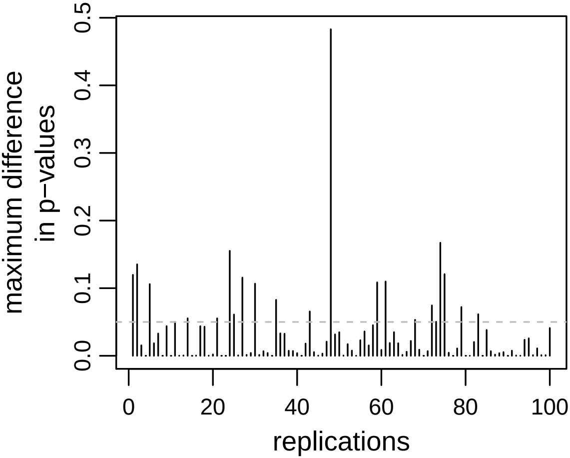
Existing challenges

- Available fitting methods require the index p to be known.
  - Pre-specify it with an "expert" selection.
    - What's the impact of the index p on inference?
    - Little impact on regression parameters
    - Significant impact on ϕ, thus on estimated standard errors and hypothesis tests
  - Inference on p, i.e., estimation of the variance function:
    - Full maximum likelihood estimation with density approximation
- Extensions of the *CPois* distribution:
  - The zero-inflated Poisson [ZIP] model has better performances than a regular Poisson model in modeling claim counts.
  - Excess zeros: "Hunger for bonus"
  - Patterns in observed frequencies of zeros
  - If $T \sim ZIP$, this yields a zero-inflated compound Poisson model [ZICP].
  - Extension to the severity part is more difficult!

Impact of p on parameter estimates



Impact of p on P-values



Data description

- Examples are illustrated using a data set:
  - A sample composed of 27,246 policies issued during 2006-2009.
  - 93.2% of the policies reported no claims.

Generalized linear models

$$\boxed{\eta(\boldsymbol{\mu}) = \mathbf{X}\beta} \tag{3}$$

- Denote $\boldsymbol{\sigma} = (\phi, p)'$ as the vector of nuisance parameters.
- For a given p (or $\boldsymbol{\sigma}$), we can estimate the model using the widely available Fisher's scoring algorithm: $\hat{\beta}(\boldsymbol{\sigma})$.
- We can profile out β from the likelihood and maximize the profile likelihood to estimate $\boldsymbol{\sigma}$ as

$$\boxed{\hat{\boldsymbol{\sigma}} = \arg\max_{\boldsymbol{\sigma}} \ell(\boldsymbol{\sigma}|\mathbf{y}, \hat{\beta}(\boldsymbol{\sigma})).} \tag{4}$$

- The likelihood is approximated using numerical methods, and then optimized subject to $\phi > 0$ and $p \in (1, 2)$.
- The estimate for β is $\hat{\beta}(\hat{\boldsymbol{\sigma}})$.

Fitting the model

- We specify a pure premium model:
  - Log link function
  - `LOSS` as the response variable
  - The log of the exposure as an offset
  - 12 predictors - their names are masked here

Inference results

```
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    -5.48427    0.32700 -16.771  < 2e-16 ***
var1           -0.53909    0.02715 -19.855  < 2e-16 ***
factor(var2)1 -0.17072    0.11328  -1.507  0.13181
factor(var3)1 -0.23210    0.08705  -2.666  0.00768 **
factor(var4)1 -0.04758    0.10541  -0.451  0.65172
var5           -0.10532    0.04399  -2.394  0.01667 *
var6           -0.19469    0.03690  -5.276 1.33e-07 ***
var7           -0.06089    0.04002  -1.521  0.12817
var8           -0.06276    0.04042  -1.553  0.12049
var9            0.16668    0.04248   3.924 8.74e-05 ***
var10           0.25248    0.03955   6.384 1.76e-10 ***
var11           0.05539    0.04428   1.251  0.21092
var12           0.07475    0.03581   2.088  0.03685 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(MLE estimate for the dispersion parameter is  22.829 ;
 MLE estimate for the index parameter is  1.4749 )

Residual deviance: 138337  on 27233  degrees of freedom
AIC:  26148
```

Generalized linear mixed models

- Extend the GLMs by including random effects:

$$\eta(\boldsymbol{\mu}) = \mathbf{X}\beta + \mathbf{Z}\mathbf{b}$$
$$\mathbf{b} \sim (\mathbf{0}, \boldsymbol{\Sigma})$$

- The distribution on **b** shrinks its estimate toward zero.
- The Bülmann credibility formula is a special case of the (Normal) mixed model with only the intercept.
- Existing inference method: Penalized Quasi-likelihood
  - Not suited to estimating *p*- the objective function maximized is not truly an approximation of the likelihood
  - Likelihood ratio tests to compare nested models?

Estimation in GLMM

- We consider full maximum likelihood estimation methods that maximize the marginal likelihood

$$\boxed{p(\mathbf{y}|\boldsymbol{\beta},\phi,p,\boldsymbol{\Sigma}) = \int p(\mathbf{y}|\boldsymbol{\beta},\phi,p,\mathbf{b})\cdot p(\mathbf{b}|\boldsymbol{\Sigma})d\mathbf{b}.} \tag{5}$$

- This integral is intractable and must be evaluated numerically.
  1. Laplace approximations
     - Integrate out **b** using the second-order Taylor approximation to the joint likelihood at the conditional mode of **b**.
     - Conditional mode of **b** is found using Penalized Iteratively Re-weighted Least Squares.
  2. Adaptive Gauss-Hermite quadrature
     - Higher-order integral approximation
     - Collapse to the Laplace method when only one knot is specified
     - More accurate at the cost of slower speed
     - Limited to a single grouping factor

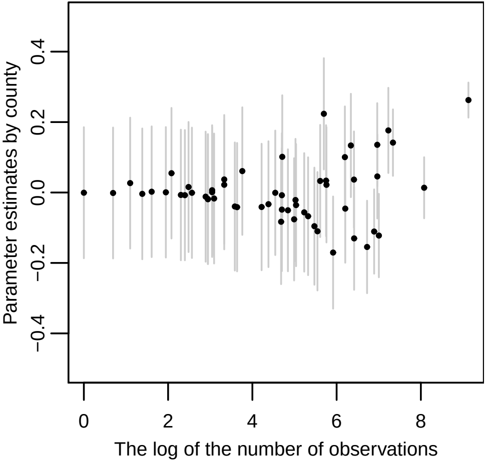
Fitting the model

- We allow intercepts to vary by `COUNTY`
- This will account for the within county correlation: closer risks are more alike
- This will also shrink parameter estimates:
  - Estimates for small counties are pulled toward the overall mean for lack of credibility

Inference results

```
Random effects:
 Groups   Name        Variance  Std.Dev.
 COUNTY   (Intercept)  0.034618 0.18606
 Residual             22.686004 4.76298
Number of obs: 27246, groups: COUNTY, 56

Fixed effects:
              Estimate Std. Error t value
(Intercept)   -5.54023    0.28477 -19.455
var1          -0.54251    0.02333 -23.258
factor(var2)1 -0.18056    0.09762  -1.850
factor(var3)1 -0.22919    0.07530  -3.044
factor(var4)1 -0.07363    0.09514  -0.774
var5          -0.10870    0.03794  -2.865
var6          -0.19327    0.03176  -6.086
var7          -0.05482    0.03452  -1.588
var8          -0.05690    0.03484  -1.633
var9           0.21623    0.05443   3.973
var10          0.23819    0.05598   4.255
var11          0.10114    0.04767   2.122
var12          0.07608    0.03080   2.470

Estimated scale parameter: 22.686
Estimated index parameter: 1.4757
```

County estimates



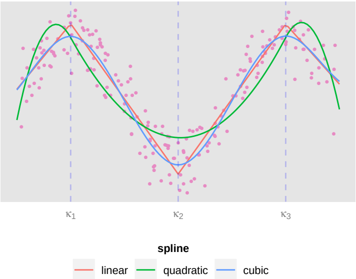
Introduction to splines

- Splines offer a flexible means of modeling nonlinear pattern:
  - It is hard to find an appropriate parametric nonlinear model.
- Model the pattern using piece-wise polynomials (basis functions):
  - Number of cut-off points (knots)
  - Positioning of the knots

Form	X	Z
Linear	x	$(x-\kappa_1)_+, (x-\kappa_2)_+$
Quadratic	x, x^2	$(x-\kappa_1)_+^2, (x-\kappa_2)_+^2$
Cubic	x, x^2, x^3	$(x-\kappa_1)_+^3, (x-\kappa_2)_+^3$
Radial	x	$\|x-\kappa_1\|, \|x-\kappa_2\|$

Table: Basis functions.

$(x-\kappa)_+ = (x-\kappa)\cdot(x-\kappa>0)$



Spline bases in GLM

- These basis functions can be used in a linear model as (e.g., with linear basis functions)

$$\boxed{\eta(\mu_i) = \beta_0 + \beta_1 x_i + \sum_{k=1}^{K} b_k (x_i - \kappa_k)_+.} \tag{6}$$

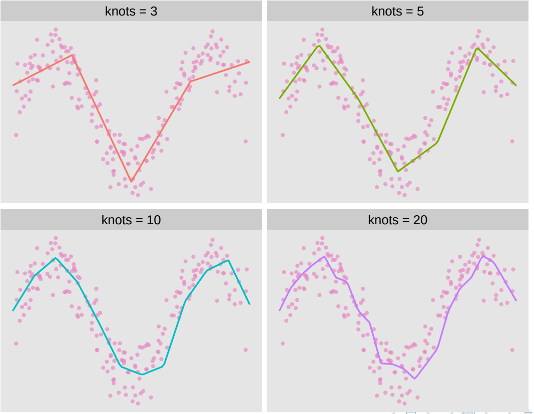
- Using matrix notation,

$$\boxed{\eta(\boldsymbol{\mu}) = \mathbf{X}\beta + \mathbf{Zb}.} \tag{7}$$

  - $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ is the coefficients for intercept and x;
  - $\mathbf{b} = (b_1, \cdots, b_K)'$ is the coefficients for the basis functions having knots;
  - $\mathbf{X}_i = (1, x_i)$ and $\mathbf{Z}_i = [(x_i - \kappa_1)_+, \cdots, (x_i - \kappa_K)_+]$ design matrix.

Problem with choices of spline knots

- Too few - not enough to describe the pattern.
- Too many - wiggly fit, including too much noise.



Additive models: penalized splines

- To avoid wiggly fit, we impose the constraints $\mathbf{b}^T\mathbf{b} < C$.
- This "penalty" is equivalent to assuming

$$\boxed{b_k \sim N(0, \sigma_b^2).} \tag{8}$$

- This provides a convenient way to estimate additive models using the mixed model software.

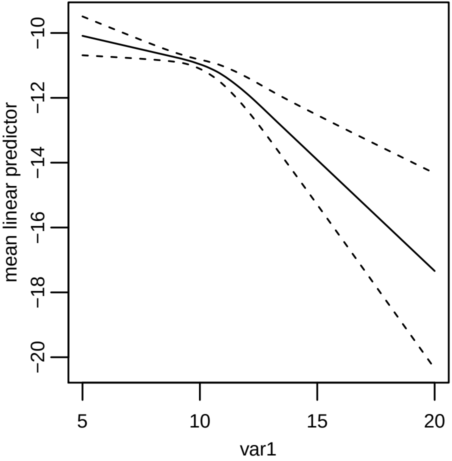
Fitting the model

- We specify a smoothing effect for `var1` using a linear spline.
- We use 15 knots, determined by empirical quantiles.
- Fit the model using the mixed-model estimation method.

Inference results

```
Random effects:
 Groups    Name Variance  Std.Dev.
 f.var1    tp     0.015549 0.12469
 Residual       22.727942 4.76738
Number of obs: 27246, groups: f.var1, 14

Fixed effects:
               Estimate Std. Error t value
(Intercept)   -11.12784    0.24438  -45.54
var1.fx1       -0.22747    0.17502   -1.30
factor(var2)1  -0.15661    0.09742   -1.61
factor(var3)1  -0.21359    0.07490   -2.85
factor(var4)1  -0.05137    0.09054   -0.57
var5           -0.11730    0.03803   -3.08
var6           -0.19423    0.03168   -6.13
var7           -0.05469    0.03439   -1.59
var8           -0.06505    0.03477   -1.87
var9            0.16463    0.03646    4.51
var10           0.24712    0.03398    7.27
var11           0.05807    0.03798    1.53
var12           0.07783    0.03080    2.53

Estimated scale parameter: 22.7279
Estimated index parameter: 1.4763
```

Smoothing effect on var1



The Zero-inflated compound Poisson distribution

- Zero-inflated Poisson model to account for excess zeros in count data:

$$T_i \sim \begin{cases} 0 & \text{with probability } q_i, \\ Pois(\lambda_i) & \text{with probability } 1 - q_i. \end{cases} \tag{9}$$

- Replacing the latent Poisson variable by the above zero-inflated Poisson, we have a zero-inflated compound Poisson:

$$Y_i \sim \begin{cases} 0 & \text{with probability } q_i, \\ CPois(\mu_i, \phi, p) & \text{with probability } 1 - q_i. \end{cases} \tag{10}$$

  - The zero-inflation part generates the excess zeros with probability q_i.
  - The compound Poisson part generates the random claim amount from the compound Poisson process.

The ZICP model

- Under this assumption, the probability of observing a zero is

$$\boxed{Pr(Y_i = 0) = q_i + (1 - q_i) \cdot \exp\left(-\frac{\mu_i^{2-p}}{\phi(2-p)}\right).} \tag{11}$$

- We allow covariates to be incorporated in both parts such that

$$\boxed{\varphi(\mathbf{q}) = \mathbf{G}\gamma, \quad \eta(\boldsymbol{\mu}) = \mathbf{B}\boldsymbol{\beta}.} \tag{12}$$

- The zero-inflation part enables one to
  - Investigate the claim underreporting behavior due to bonus hunger.
  - More adequately model the patterns in the observed frequency of zeros .

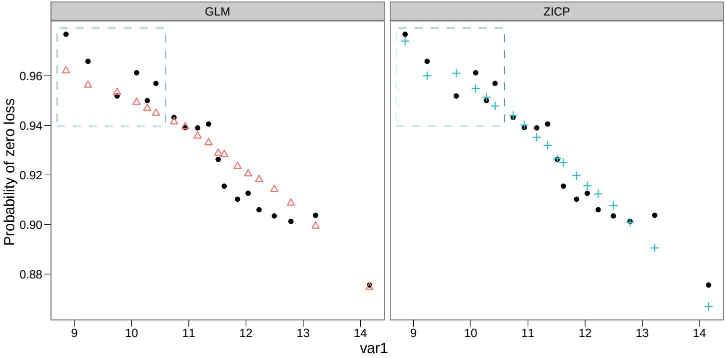
Fitting the model

- We specify four relevant covariates in the zero-inflation part.
- The offset term is only used for the compound Poisson part.

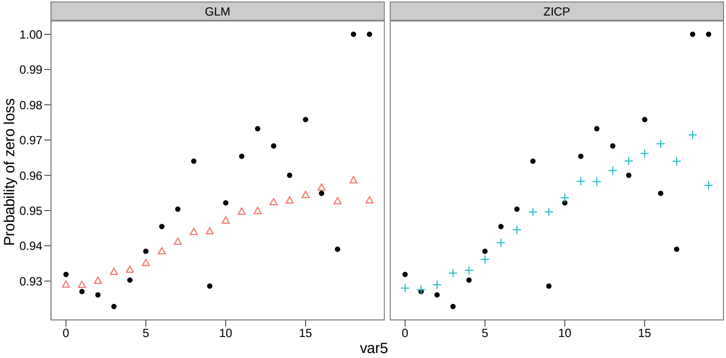
Inference results

```
Zero-inflation model coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  5.76568    0.87536   6.587 4.50e-11 ***
var1        -0.57326    0.08098  -7.079 1.45e-12 ***
var5         0.26870    0.08934   3.008 0.002633 **
var12       -0.29465    0.07935  -3.713 0.000205 ***
var6         0.39966    0.11359   3.519 0.000434 ***

Compound Poisson model coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -3.08997    0.38046  -8.122 4.60e-16 ***
var1          -0.70988    0.03144 -22.580  < 2e-16 ***
factor(var2)1 -0.17217    0.09748  -1.766  0.07735 .
factor(var3)1 -0.21038    0.07560  -2.783  0.00539 **
factor(var4)1 -0.03911    0.09126  -0.429  0.66820
var5          -0.01280    0.05290  -0.242  0.80879
var6          -0.08766    0.04214  -2.080  0.03753 *
var7          -0.05532    0.03574  -1.548  0.12167
var8          -0.06335    0.03617  -1.751  0.07988 .
var9           0.15679    0.03732   4.202 2.65e-05 ***
var10          0.24797    0.03419   7.254 4.06e-13 ***
var11          0.05167    0.03990   1.295  0.19532
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(MLE estimate for the dispersion parameter is  19.079 ;
 MLE estimate for the index parameter is  1.486 )
```

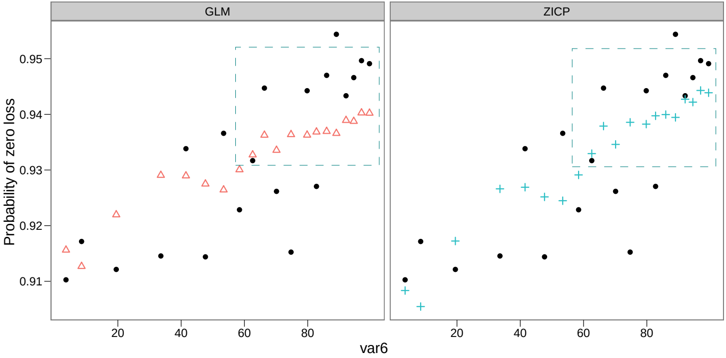
Predicted probability of zeros (1)



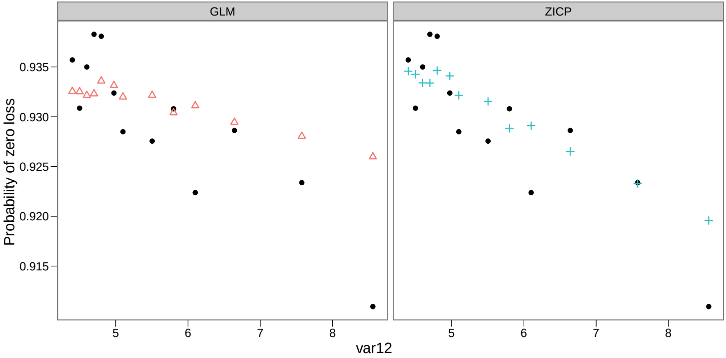
Predicted probability of zeros (2)



Predicted probability of zeros (3)



Predicted probability of zeros (3)



Model comparisons

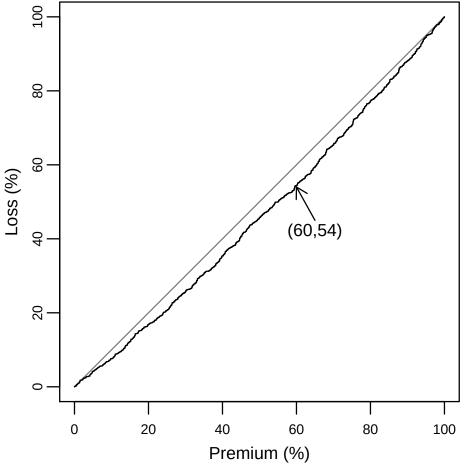
- The information criteria
- The 10-fold cross validation mean squared error (not quite informative)
- The Gini index
  - Let y_i be the loss, P_i be the baseline premium, S_i be the insurance score (predictions from the model) and $R_i = S_i/P_i$ be the relativity.
  - Sort the observations by the relativity in an increasing order.
  - Compute the empirical cumulative premium and loss distributions as

$$\hat{F}_P(s) = \frac{\sum_{i=1}^n P_i \cdot \mathbb{1}(R_i \leq s)}{\sum_{i=1}^n P_i}, \ \hat{F}_L(s) = \frac{\sum_{i=1}^n y_i \cdot \mathbb{1}(R_i \leq s)}{\sum_{i=1}^n y_i}. \tag{13}$$

  - The graph $\left(\hat{F}_P(s), \hat{F}_L(s)\right)$ is an ordered Lorenz curve.

	Loglikelihood	AIC	BIC	MSE	Gini
GLM	-13067.43	26147.85	26267.61	24.98	-1.62(2.13)
ZICP	-13022.18	26078.36	26217.98	24.95	6.92(2.10)

The ordered Lorenz curve



Summary

- Reviewed the compound Poisson distribution.
- Discussed the challenges on statistical inference.
- Presented MLE methods for estimating various linear models.
- Illustrated these techniques through an example.