Direct Analysis of Pre-Adjusted Loss Cost, Frequency or Severity in Tweedie Models

Sheng G. Shi, Ph.D.

Abstract

Response data (loss cost, claim frequency or claim severity) are often pre-adjusted with known factors and directly analyzed with generalized linear models (GLM). This paper shows that the exposure weights should also be adjusted if the Tweedie distribution with log link is used in such direct analysis. An advantage of the direct analysis over GLM offsetting is that the structure of the original dataset may be simplified significantly after removing the known factors. Direct analysis is a convenient tool for directly modeling loss ratio and for removing known territory factors from the dataset. Implementation in EMBLEM and SAS is discussed, and a computationally efficient SAS macro is provided for Tweedie models.

Keywords. Predictive Modeling; GLM offset; Ratemaking; Tweedie Model; EMBLEM; SAS.

1. INTRODUCTION

In insurance ratemaking, response data are often pre-adjusted with known factors before predictive modeling. However, the effect of adjustment on exposure weights is usually either ignored or not linked to the response distribution. This is particularly the case when the response variable is loss cost, which is assumed to follow a Tweedie distribution of power p ($1<p<2$).

Application of the GLM offset feature in property-casualty predictive modeling has been discussed recently by Yan et. al.[7]. They translated the analysis on loss ratio into an analysis on loss cost with premium offset. In this paper, we will show how loss ratio, viewed as loss cost pre-adjusted with premium rates, can be analyzed directly. An advantage of the direct analysis is that the structure of original dataset may be simplified significantly for subsequent analysis. We first show, in general, how the exposure weights should be modified in Tweedie models (including the special case of Poisson and Gamma) with pre-adjusted loss cost, claim frequency or claim severity as the response.

2. CONNECTION BETWEEN OFFSETS AND PRE-ADJUSTMENT

In this section, we give two propositions that connect GLM offsets with pre-adjustment. Proposition 1 builds a simple linkage between the offsets and pre-adjustment. Proposition 2 establishes a foundation for data simplification.

2.1 Assumptions and Notations

Suppose that there are two rating factors *U* and *V*, where *U* has *m* categories and *V* has *n* categories. Denote u_i as the relativity of the i^{th} category ($i = 1,2,...,m$) of *U* and v_j as the relativity of the j^{th} category ($j = 1,2,...,n$) of *V*. Let Y_{ij} be a random variable for the ratio of interest in the rating cell with the i^{th} level of *U* and the j^{th} level of *V* such that $Y_{ij} = X_{ij} / w_{ij}$. When the ratio of interest is loss cost, $X_{ij} = L_{ij}$ as loss amount and $w_{ij} = e_{ij}$ as earned exposure. When the ratio of interest is claim frequency, $X_{ij} = c_{ij}$ as claim count and $w_{ij} = e_{ij}$. When the ratio of interest is claim severity, $X_{ij} = L_{ij}$ and $w_{ij} = c_{ij}$. Assume that the Y_{ij}'s are mutually independent and Y_{ij} follows a Tweedie distribution with power parameter *p* such that

$$E(Y_{ij}) = u_i v_j, \qquad (2.1)$$

$$Var(Y_{ij}) = \phi(u_i v_j)^p / w_{ij} \qquad (2.2)$$

where ϕ is a constant dispersion parameter [3]. To include the dispersed Poisson and Gamma as special cases of the Tweedie distribution, we focus on the range of power parameter, $1 \le p \le 2$.

As in a typical analysis, we assume that the power parameter *p* and the constant dispersion parameter ϕ are known or have been pre-determined. We will use log link in all the models.

2.2 Propositions

2.2.1 Simple link between GLM offsets and pre-adjustment

With the Tweedie model, an offset problem can be translated into a pre-adjustment problem and vice versa as shown in the proposition below. This interchangeability also allows us to have a model with both pre-adjustment and offsets.

Proposition 1

Under the assumptions and notations above, if u_i's are known, then fitting the following Tweedie model (in Eq. (2.3)) of power *p* with weights w_{ij} and $\log(u_i)$ as an offset,

$$\log E(Y_{ij}) = \log(v_j) + \log(u_i), \qquad (2.3)$$

where $i = 1,2,...,m$ and $j = 1,2,...,n$, is equivalent to fitting the Tweedie model of power *p* below (in Eq. (2.4)) with pre-adjusted response variable $Z_{ij} = Y_{ij} / u_i$ and weights $w_{ij} u_i^{2-p}$,

$$\log E(Z_{ij}) = \log(v_j), \qquad (2.4)$$

where $i = 1,2,...,m$ and $j = 1,2,...,n$.

In other words, Z_{ij} can be viewed to follow Tweedie distribution of the same power *p* and dispersion parameter ϕ as Y_{ij}, but with different weights.

Proof

Note that the Tweedie distribution belongs to the exponential dispersion family, which is closed under a scale transformation (cf. [3] Formula 6 on p. 72). Thus, Z_{ij} follows a Tweedie distribution with power parameter *p*. Based on Eq. (2.1) and Eq. (2.2) above,

$$E(Z_{ij}) = E(Y_{ij})/u_i = u_i v_j / u_i = v_j \tag{2.5}$$

and

$$Var(Z_{ij}) = Var(Y_{ij})/u_i^{\,2} = \phi(u_i v_j)^p /(w_{ij} u_i^{\,2}) = \phi v_j^{\,p} /(w_{ij} u_i^{\,2-p}). \tag{2.6}$$

To show that two models are equivalent, let l_{ij} be the log-likelihood function for $Y_{ij} = y_{ij}$. Then, according to the property of the exponential dispersion family, we have

$$\frac{\partial l_{ij}}{\partial \mu_{ij}} = \frac{y_{ij} - \mu_{ij}}{\phi V(\mu_{ij})/w_{ij}} \tag{2.7}$$

where the mean $\mu_{ij} = E(Y_{ij}) = u_i v_j$ and the variance function $V(\mu_{ij}) = \mu_{ij}^{\,p}$. To obtain the maximum likelihood estimate $\hat{v}_j$ of v_j, we set for $j = 1,2,\ldots,n$,

$$\sum_i \frac{\partial l_{ij}}{\partial v_j} = \sum_i \frac{\partial l_{ij}}{\partial \mu_{ij}} \frac{\partial \mu_{ij}}{\partial v_j} = \sum_i \frac{y_{ij} - u_i v_j}{\phi(u_i v_j)^p / w_{ij}} u_i = 0 \tag{2.8}$$

which leads to the estimate for the model specified in Eq. (2.3),

$$\hat{v}_j = \frac{\sum_i w_{ij} u_i^{\,2-p} y_{ij} / u_i}{\sum_i w_{ij} u_i^{\,2-p}}. \tag{2.9}$$

Now, let l_{ij}^* be the log-likelihood function for $Z_{ij} = z_{ij}$. Then,

$$\frac{\partial l_{ij}^*}{\partial v_j} = \frac{z_{ij} - v_j}{\phi V(v_j)/(w_{ij} u_i^{\,2-p})} \tag{2.10}$$

where the mean $v_j = E(Z_{ij})$ and the variance function $V(v_j) = v_j^{\,p}$. To obtain the maximum likelihood estimate $\hat{v}_j^*$ of v_j, we set for $j = 1,2,\ldots,n$,

$$\sum_i \frac{\partial l_{ij}^*}{\partial v_j} = \sum_i \frac{z_{ij} - v_j}{\phi v_j^{\,p} /(w_{ij} u_i^{\,2-p})} = 0 \tag{2.11}$$

which leads to the estimate for the model specified in Eq. (2.4),

$$\hat{v}_j^* = \frac{\sum_i w_{ij} u_i^{2-p} z_{ij}}{\sum_i w_{ij} u_i^{2-p}} = \hat{v}_j \, . \tag{2.12}$$

It is easy to verify that $\sum_i \frac{\partial^2 l_{ij}(\hat{v}_j)}{\partial v_j^2} < 0$ and $\frac{\partial^2 l_{ij}^*(\hat{v}_j^*)}{\partial v_j^2} < 0$ for the maxima. Q.E.D.

Note that the right side of Eq. (2.4) is not related to the index *i*. Thus, it may be simplified by collapsing over the rating factor *U* as discussed in Section 2.2.2.

Example 1

Loss ratio can be viewed as loss cost L_{ij} / e_{ij} pre-adjusted with the premium rates u_i in a rating plan:

Loss Ratio = Losses/Earned Premiums

= Losses/(Exposures*Rates) = (Losses/Exposures)/Rates

= (Loss Cost)/Rates.

Assume that the loss cost L_{ij} / e_{ij} follows Tweedie of power *p*. Then, the loss ratio $L_{ij} /(e_{ij} u_i)$ can be analyzed with the Tweedie model of power *p*, but the model weights need to be adjusted to Exposures*Rates^(2-*p*) = $e_{ij} u_i^{2-p}$.

2.2.2 Pre-adjustment for data simplification

Aggregating data reduces the number of records in a dataset and simplifies the data structure. This can be especially beneficial when aggregating across high-dimensional variables, such as territory. From a modeling perspective, this is achieved by collapsing on the GLM offset variable, but subsequent analyses will then need to be done with pre-adjusted data as shown in the proposition below.

Proposition 2

Under the assumptions and notations above, if u_i's are known, then fitting the following Tweedie model (in Eq. (2.13)) of power *p* with weights w_{ij} and $\log(u_i)$ as an offset

$$\log E(Y_{ij}) = \log(v_j) + \log(u_i), \tag{2.13}$$

where $i = 1,2,...,m$ and $j = 1,2,...,n$, is equivalent to fitting the simplified Tweedie model of power *p* below (in Eq. (2.14)) with weights $\sum_i w_{ij} u_i^{2-p}$,

$$\log E(Z_j) = \log(v_j);\ j = 1,2,...,n, \tag{2.14}$$

where
$$Z_j = \frac{\sum_i w_{ij} u_i^{2-p} (Y_{ij} / u_i)}{\sum_i w_{ij} u_i^{2-p}};\ j = 1,2,...,n. \tag{2.15}$$

In other words, Z_j can be viewed to follow the Tweedie distribution of the same power *p* and dispersion parameter ϕ as Y_{ij}, but with different weights (cf. [4]).

Proof

Note that the Tweedie distribution belongs to the exponential dispersion family, which is closed under a scale transformation and follows the convolution formula (cf. [3] Formula 10 on p. 74). Write $Z_{ij} = Y_{ij} / u_i$. We know from Proposition 1 that Z_{ij} follows the Tweedie distribution of the power *p* with mean v_j, dispersion parameter ϕ and prior weights $w_{ij} u_i^{2-p}$. Therefore, for $j = 1,2,...,n$,

$$Z_j = \frac{\sum_i w_{ij} u_i^{2-p} Z_{ij}}{\sum_i w_{ij} u_i^{2-p}} \tag{2.16}$$

is still Tweedie distributed with the power parameter *p* and

$$E(Z_j) = \frac{\sum_i w_{ij} u_i^{2-p} E(Y_{ij} / u_i)}{\sum_i w_{ij} u_i^{2-p}} = \frac{\sum_i w_{ij} u_i^{2-p} (u_i v_j / u_i)}{\sum_i w_{ij} u_i^{2-p}} = v_j, \tag{2.17}$$

$$
\begin{aligned}
Var(Z_j) &= \frac{\sum_i (w_{ij}u_i{}^{2-p})^2 * Var(Y_{ij}/u_i)}{(\sum_i w_{ij}u_i{}^{2-p})^2} \\
&= \frac{\sum_i (w_{ij}u_i{}^{2-p})^2 * (\phi u_i{}^p v_j{}^p / w_{ij}u_i{}^2)}{(\sum_i w_{ij}u_i{}^{2-p})^2} \\
&= \phi v_j{}^p * \frac{\sum_i (w_{ij}u_i{}^{2-p})^2 / (w_{ij}u_i{}^{2-p})}{(\sum_i w_{ij}u_i{}^{2-p})^2} \\
&= \phi v_j{}^p / \sum_i w_{ij}u_i{}^{2-p}.
\end{aligned}
\tag{2.18}
$$

To show the two models are equivalent for estimating v_j, we note from the proof of Proposition 1 that the maximum likelihood estimate for the model specified by Eq. (2.13) is given in Eq. (2.9). From Eq. (2.17), it is rather trivial that the maximum likelihood estimate for the model specified by Eq. (2.14) is the same as that in Eq. (2.9), because only a single Z_j is involved for estimating v_j. Q.E.D.

Example 2

In a loss ratio analysis, a dataset with numerous premium rate levels may be simplified by collapsing over the premium variable. Note that a unique premium rate level is defined by a unique combination of all rating variables in a rating plan. The data size can be reduced drastically in many cases by collapsing over the premium variable. Before collapsing, loss ratios $L_{ij}/(e_{ij}u_i)$ are recorded for each exposure, where u_i's are premium rates. We are interested in fitting a Tweedie model of power *p* with other covariates that are combined into v_j. After collapsing, we can equivalently model "weighted loss ratios" $(\sum_i L_{ij}u_i^{1-p})/(\sum_i e_{ij}u_i^{2-p})$ with adjusted exposure weights $\sum_i e_{ij}u_i^{2-p}$. Note that the weighted loss ratios are not of the form $(\sum_i L_{ij})/(\sum_i e_{ij}u_i)$.

Example 3

In a loss cost analysis, a dataset with numerous territories may be simplified by collapsing over the territory variable. Both loss cost and exposure weights need to be adjusted by known territory relativities for Tweedie models.

3. IMPLEMENTATION

Suppose that the EMBLEM (cf. [1] and [2]) data source is in a summarized table such that each record has an observed level (indexed by *i*) of a rating factor (for example, territory) to be collapsed, an observed level (indexed by *j*) of a populated combination of other rating factors, along with the number of claims (c_{ij}), the loss amount (L_{ij}) and the exposure (e_{ij}) at the level (*i, j*). Assume that the original predictive models are as in Table 1 and log link is used for all models. With the log link, u_i is specified as an offset in accordance with the EMBLEM logic. The dispersion parameter is either specified or estimated wherever appropriate.

With pre-adjustment, the response variable and the weight variable before collapsing are given in Table 2 in accordance with Proposition 1. After collapsing, the response variable and the weight variable in Table 3 are ready for simplified analysis in accordance with Proposition 2.

Table 1. Description of original predictive models

Model	Distribution	Response Variable	Weight Variable	Offset
Frequency	Poisson	Claim frequency, c_{ij}/e_{ij}	Number of exposures, e_{ij}	u_i
Severity	Gamma	Claim severity, L_{ij}/c_{ij}	Number of claims, c_{ij}	u_i
Loss cost	Tweedie(*p*)	Loss cost, L_{ij}/e_{ij}	Number of exposures, e_{ij}	u_i

Table 2. Description of pre-adjustment models before collapsing

Model	Distribution	Response Variable	Weight Variable
Frequency	Poisson	Adjusted claim frequency, $c_{ij}/(e_{ij}u_i)$	Number of adjusted exposures, $e_{ij}u_i$
Severity	Gamma	Adjusted claim severity, $L_{ij}/(c_{ij}u_i)$	Number of claims, c_{ij}
Loss cost	Tweedie(*p*)	Adjusted loss cost, $L_{ij}/(e_{ij}u_i)$	Number of adjusted exposures, $e_{ij}u_i^{2-p}$

Table 3. Description of pre-adjustment models after collapsing

Model	Distribution	Response Variable	Weight Variable
Frequency	Poisson	Weighted sum of adjusted claim frequency, $(\sum_i c_{ij})/(\sum_i e_{ij}u_i)$	Total number of adjusted exposures, $\sum_i e_{ij}u_i$
Severity	Gamma	Weighted sum of adjusted claim severity, $(\sum_i L_{ij}/u_i)/(\sum_i c_{ij})$	Total number of claims, $\sum_i c_{ij}$
Loss cost	Tweedie(*p*)	Weighted sum of adjusted loss amount $(\sum_i L_{ij}u_i^{1-p})/(\sum_i e_{ij}u_i^{2-p})$	Total number of adjusted exposures, $\sum_i e_{ij}u_i^{2-p}$

Implementation in SAS can be done similarly. With known Tweedie power and dispersion parameters, the GENMOD procedure can be adopted with user defined distribution.[5]

4. REMARKS

Throughout this paper, we assumed that both the Tweedie power and dispersion parameters are known. In practice, the power parameter *p* is often taken from prior modeling experience, while the dispersion parameter is estimated using the Pearson, Deviance or the likelihood approach [3]. Compared to the likelihood approach, an estimated dispersion parameter using either the Pearson or Deviance can be significantly different for *p* in the mid-range of the interval (1, 2). In the SAS environment, PROC NLMIXED may be used for simultaneous estimation of all Tweedie parameters using the code written by Flynn [7], but convergence may become a problem with a large dataset and numerous class variables. As an alternative, the code in Appendix A may be applied.

We assumed that the dispersion parameter ϕ is a constant. However, it is often more appropriate to allow ϕ to vary with different rating cells such that $\phi = \phi_{ij}$, especially in a loss cost model [6]. In such a case, if we insist on fitting a model with fixed ϕ, then a different set of weights may be necessary for an accurate solution. On the other hand, if ϕ is allowed to vary, we may put any adjustment on weights into ϕ, leaving the original weights untouched.

The choice of weights in Eq. (2.4) and Eq. (2.14) affects both the accuracy of the model estimates and the validity of hypothesis tests even if an estimate of v_j is unbiased.

Both propositions 1 and 2 can be generalized to the case with more than two rating factors. Note that multiple adjustments can be combined and sequenced with index *i* and other covariates may be combined and sequenced with index *j*.

Acknowledgment

The author thanks Trevor Handley, FCAS, MAAA for reviewing this paper.

Appendix A

The SAS macro provided in this appendix may be used experimentally for Tweedie models. The macro is based on the orthogonal property between the mean parameter μ and the power/dispersion parameter (p,ϕ) [3], which allows their separate optimizations. It iterates until convergence between the μ-step with PROC GENMOD and the (p,ϕ)-step with PROC NLMIXED. The GENMOD procedure is easy to converge and has a handy CLASS statement, which is suitable for Tweedie models with known (p,ϕ) and high dimension of μ. This approach reduces the burden on PROC NLMIXED so that it is used only to estimate (p,ϕ) with μ assumed known.

```
****************************************************************
*                MACRO FOR TWEEDIE MODEL                       *
*                                                              *
*       Author: Sheng G. Shi                                   *
*       Paramters:                                             *
*         dn -- dataset name                                   *
*         vformat -- list of formats                           *
*         vclass -- class variables                            *
*         wght -- weight variable                              *
*         resp -- response variable (must be non-negative)     *
*         pred -- predictors                                   *
*         clmcnt -- claim counts                               *
*         offset -- offset variable                            *
*      Warning:                                                *
*        Check output for convergence of GENMOD and NLMIXED;   *
*        Check log for results;                                *
*        Title3 will be over-written.                          *
****************************************************************;
%macro tweedie(dn=,vformat=,vclass=,wght=,resp=,pred=,clmcnt=,offset=);
/* Initialization */
title3;
data Est_save_;
  format p_ phi_ sigma_ p_lower p_upper phi_lower phi_upper
         sigma_lower sigma_upper 15.4 p_change sigma_change 15.4;
  p_ = 1.5;
  p_lower = .;
```

```
  p_upper = .;
  phi_ = 1;
  phi_lower = .;
  phi_upper = .;
  sigma_ = 1;
  sigma_lower = .;
  sigma_upper = .;
  p_change = .;
  sigma_change = .;
  call symput('p',trim(left(put(p_,15.4))));
  call symput('phi',trim(left(put(phi_,15.4))));
  call symput('sigma',trim(left(put(sigma_,15.4))));
  call symput('p_lower',trim(left(put(p_,15.4))));
  call symput('p_upper',trim(left(put(p_,15.4))));
  call symput('phi_lower',trim(left(put(phi_,15.4))));
  call symput('phi_upper',trim(left(put(phi_,15.4))));
  call symput('sigma_lower',trim(left(put(sigma_,15.4))));
  call symput('sigma_upper',trim(left(put(sigma_,15.4))));
run;
/* Maximum likelihood estimation */
%let converge = 0;
%let i=1;
%do %until ((&converge eq 1) or (&i gt 10));
  title3 "Optimization Step &i";
  %optimize(&dn,&vformat,&vclass,&wght,&resp,&pred,&clmcnt,&offset,0);
  %let i = %eval(&i+1);
    data Est_save_(drop=p_old sigma_old);
        set Est_save_ end=last;
        p_old = p_;
        sigma_old = sigma_;
        retain p_old sigma_old;
        output;
        if last then do;
          p_ = &p;
          p_lower = &p_lower;
          p_upper = &p_upper;
          phi_ = &phi;
          phi_lower = &phi_lower;
          phi_upper = &phi_upper;
          sigma_ = &sigma;
          sigma_lower = &sigma_lower;
          sigma_upper = &sigma_upper;
          p_change = abs(p_-p_old);
          sigma_change = abs(sigma_-sigma_old);
          p_old = p_;
          sigma_old = sigma_;
          if (p_change le 1e-5) and (p_change ne .)
                and (sigma_change le 1e-5) and (sigma_change ne .) then
              call symput('converge','1');
          output;
        end;
    run;
%end;
```

```
/* Results */
%if (&converge eq 1) %then %do;
  title3 'Tweedie Model with Converged Parameter Estimates';
  %optimize(&dn,&vformat,&vclass,&wght,&resp,&pred,&clmcnt,&offset,1);
  %put Converged;
  %put Power parameter = &p with 95% C.I. (&p_lower, &p_upper);
  %put Dispersion parameter = &phi with 95% C.I. (&phi_lower, &phi_upper);
  %put SAS scale parameter = &sigma with 95% C.I. (&sigma_lower,
&sigma_upper);
%end; %else %do;
  %put Not converged after 10 iterations: ;
  %put Power parameter = &p;
  %put Dispersion parameter = &phi;
  %put SAS scale parameter = &sigma;
  %put at the end of 10th iteration.;
%end;
title3;
%mend tweedie;
%macro optimize(dn,vformat,vclass,wght,resp,pred,clmcnt,offset,flag);
proc genmod data=&dn;
  format &vformat;
  class  &vclass
       /param=glm;
  p_ = &p;
  mu_ = _MEAN_;
  y_ = _RESP_;
  v_ = mu_**p_;
  if y_ gt 0 then
    d_ = 2*(y_*(y_**(1-p_)-mu_**(1-p_))/(1-p_)-(y_**(2-p_)-mu_**(2-p_))/(2-
p_));
  else
    d_ = 2*(mu_**(2-p_))/(2-p_);
  variance var = v_;
  deviance dev = d_;
  weight &wght;
  model &resp = &pred
      /link=log noscale scale=&sigma
            %if %length(&offset) eq 0 %then ;
            %else offset=&offset;;
  output out=Out_mu_ pred=yhat_;
run;
%if &flag ne 1 %then %do;
ods trace on;
ods output ParameterEstimates=Est_;
proc nlmixed data=Out_mu_;
  format p_ 15.4 phi_ 15.4;
  parms  p_=&p phi_=&phi;
  bounds 1<p_<2, phi_>0;
  n_ = &clmcnt;
  w_ = &wght;
  y_ = &resp;
  mu_ = yhat_;
  t_ = y_*mu_**(1-p_)/(1-p_)-mu_**(2-p_)/(2-p_);
```

```
  a_ = (2-p_)/(p_-1);
  if (n_ eq 0) then
    rll_ = (w_/phi_)*t_;
  else
    rll_ = n_*((a_+1)*log(w_/phi_)+a_*log(y_)-a_*log(p_-1)-log(2-p_))
        -lgamma(n_+1)-lgamma(n_*a_)-log(y_)+(w_/phi_)*t_;
  /* log likelihood of (p_,phi_) with mu_ known */
  model y_ ~ general(rll_);
run;
ods trace off;
data _null_;
  set Est_;
  if Parameter eq 'p_' then do;
    call symput('p',trim(left(put(Estimate,15.4))));
    call symput('p_lower',trim(left(put(Lower,15.4))));
    call symput('p_upper',trim(left(put(Upper,15.4))));
  end; else if Parameter eq 'phi_' then do;
    call symput('phi',trim(left(put(Estimate,15.4))));
    call symput('phi_lower',trim(left(put(Lower,15.4))));
    call symput('phi_upper',trim(left(put(Upper,15.4))));
    call symput('sigma',trim(left(put(sqrt(Estimate),15.4))));
    call symput('sigma_lower',trim(left(put(sqrt(Lower),15.4))));
    call symput('sigma_upper',trim(left(put(sqrt(Upper),15.4))));
  end;
run;
%end;
%mend optimize;
**************************************************************;
```

Here is an example that calls the **%*tweedie*** macro:

```
%tweedie(dn=CarData,
         vformat=ModelYr 4.
                 SYM $symfmt.,
         vclass=ModelYr SYM,
         wght=EExp,
         resp=LossCost,
         pred=ModelYr SYM,
         clmcnt=ClaimCnt,
         offset=LogEP

);
```

5. REFERENCES

[1] EMBLEM Getting Started Guide, EMB Software Limited, 1999-2008.
[2] EMBLEM User's Guide, EMB Software Limited, 1999-2008.
[3] Jørgensen, B., and M. C. P. de Souza, "Fitting Tweedie's compound Poisson model to insurance claims data", *Scandinavian Actuarial Journal*, 1994, pp. 69-93.
[4] Ohlsson, E. and B. Johansson, "Credibility theory and GLM revised", *Research Report* 2003:15, Mathematical Statistics Stockholm University, 2003.
[5] SAS/STAT User's Guide, Cary, NC: SAS Institute Inc., 2002-2004.
[6] Smyth, G. K. and B. Jørgensen, "Fitting Tweedie's compound Poisson model to insurance claims data: dispersion modelling", *ASTIN Bulletin*, 2002, 32 (1), pp. 143-157.
[7] Yan, J., and J. Guszcza, M. Flynn, C. P. Wu, "Applications of the Offset in Property-Casualty Predictive Modeling", Casualty Actuarial Society E-Forum, Winter 2009, pp. 366-385.

Biography of the Author

Sheng G. Shi has fifteen years of experience in analyzing insurance data. He has a Ph.D. in statistics from the University of Texas at Austin. He is currently with Safeco Insurance, a member of Liberty Mutual Group.