Modeling the Severity of Random Events with the SAS/ETS® SEVERITY Procedure

Mahesh V. Joshi, SAS Institute Inc., Cary NC

ABSTRACT

The new SEVERITY procedure fits probability distributions for the severity (magnitude) of random events. Important examples include events with negative impact (such as the distribution of losses claimed under insurance policies, the magnitude of damages caused by natural disasters, and the severity of outbreaks of a disease) and events with positive impact (such as order sizes for products characterized by intermittent demand).

The SEVERITY procedure can use any of eight different parametric families of probability distributions. If these eight built-in distributions are not sufficient for your problem, then you can extend the SEVERITY procedure to support any parametric family by specifying its CDF and PDF functions with SAS® programming statements and the FCMP procedure.

The SEVERITY procedure can model the effects of exogenous variables on the severity of an event. It can also model censored or truncated data, which is an important feature, especially for insurance problems. Data on payments under insurance policies are subject to deductibles (which cause left truncation of the data) and are also subject to coverage limits (which produce right censoring).

INTRODUCTION

Insurance and banking industries, among others, need to understand the nature of losses that were incurred in the past in order to manage future risk. This is important both for maintaining a healthy and profitable business and also for complying with various industry regulations. A stochastic modeling approach is often beneficial not only because it provides quantitative estimation of the expected loss and the risk associated with it but also because it enables what-if analysis of various business scenarios. In a typical application of this approach, past data is used to find the best-fitting probability distributions for the frequency of loss occurrences and the magnitude (severity) of the losses.

In addition to allowing inference of the characteristics that are unique to the frequency and magnitude, independent analysis of frequency and severity distributions allows more modeling flexibility, especially when factors that affect the frequency are different from the factors that affect the magnitude. The best frequency and severity distributions are then combined to compute the aggregate distribution of the losses. Various quantities of business interest can be inferred from the aggregate distribution. Examples of such quantities include the estimate of the aggregate loss that is incurred in a given time period and the value at risk (VaR), which is computed as a specific percentile of the aggregate distribution. These quantities enable companies to comply with risk management regulations by allocating sufficient capital and honoring the liability commitments.

A stochastic modeling approach is also useful for modeling any random event of interest whose frequency and magnitude are expected to follow atypical probability distributions. Examples of such events are damages caused by natural disasters, disease outbreaks, and product orders characterized by intermittent demand.

The SEVERITY procedure (new in SAS/ETS 9.22) enables you to model severity distributions, which are continuous parametric probability distributions. The existing SAS/ETS COUNTREG procedure enables you to fit frequency distributions, which are discrete parametric probability distributions. The focus of this paper is to introduce and illustrate the features of the SEVERITY procedure.

The SEVERITY procedure estimates parameters of any arbitrary continuous probability distribution that is used to model magnitude (severity) of a continuous-valued event of interest. PROC SEVERITY is especially useful when the severity of an event does not follow typical distributions (such as the normal distribution) that are often assumed by standard statistical methods. It provides a default set of probability distribution models for various distributions that are used for severity modeling. In the simplest form, you can estimate the parameters of any of these distributions by using a list of severity values that are recorded in a SAS data set. The values can optionally be grouped by a set of BY variables. PROC SEVERITY computes the estimates of the model parameters, their standard errors, and their covariance structure by using the maximum likelihood method for each of the BY groups. When multiple distributions are fitted, PROC SEVERITY also chooses the best distribution according to a specified selection criterion.

You can request that the procedure output the status of the estimation process, the parameter estimates and their standard errors, the estimated covariance structure of the parameters, the statistics of fit, estimated cumulative distribution function (CDF) for each of the specified distributions, and the empirical distribution function (EDF) estimate.

The following key features of PROC SEVERITY make it different from other SAS procedures that can estimate continuous probability distributions:

- PROC SEVERITY enables you to fit a distribution model when the severity values are left-truncated or right-censored or both. This is especially useful in applications with an insurance-type model where a severity (loss) is reported and recorded only if it is greater than the deductible amount (left-truncation) and a severity value greater than or equal to the policy limit is recorded at the limit (right-censoring). The procedure also enables you to specify a *probability of observability* for the left-truncated data, which is a probability of observing values that are greater than the left-truncation threshold. In certain applications, this additional information can be useful to more correctly model the distribution of the severity of events.

  When left-truncation or right-censoring is specified, PROC SEVERITY can compute the empirical distribution function (EDF) estimate by using Kaplan-Meier's product-limit estimator or one of its robust variants.

- PROC SEVERITY can model the effect of exogenous or regressor variables on a probability distribution, providing that the distribution has a scale parameter. A linear combination of the regressor variables is assumed to affect the scale parameter via an exponential link function.

  If a distribution does not have a scale parameter, then either it needs to have another parameter that can be derived from a scale parameter by using a supported transformation or it needs to be reparameterized to have a scale parameter. If neither of these is possible, then regression effects cannot be modeled.

- PROC SEVERITY enables you to define any arbitrary continuous parametric distribution model and to estimate its parameters. You just need to define the key components of the distribution, such as its probability density function (PDF) and cumulative distribution function (CDF), as a set of functions and subroutines that are written with the FCMP procedure, which is part of Base SAS software. As long as the functions and subroutines follow certain rules, PROC SEVERITY can fit the distribution model defined by them.

The following sections illustrate the features of PROC SEVERITY with some examples inspired by the real-world applications.

GETTING STARTED WITH A SIMPLE MODEL

Let Y denote the random variable that represents the severity of the event of interest. The SEVERITY procedure fits an error model of the form $Y \sim \mathcal{F}(\Theta)$, where $\mathcal{F}$ is a probability distribution parameterized by a set of parameters Θ. In other words, PROC SEVERITY enables you to confirm or reject the hypothesis that the distribution $\mathcal{F}$ adequately describes the probability distribution of Y.

PROC SEVERITY provides a default set of probability distribution models that includes the Burr, exponential, gamma, generalized Pareto, inverse Gaussian (Wald), lognormal, Pareto, and Weibull distributions. A summary of these models is provided in Table 1.

The simplest way to use the SEVERITY procedure is to fit all of the predefined distributions to a set of values that are recorded in a variable of a SAS data set and let the procedure identify the best-fitting distribution according to a model selection criterion. For example, the following statements fit all the predefined distribution models to the values of the variable Y in the data set Mylib.MyData and identify the best model as the one with the smallest value for the Kolmogorov-Smirnov (KS) statistic:

```
proc severity data=mylib.mydata;
    model y / crit=ks;
run;
```

Internally, PROC SEVERITY forms a candidate error model for each of the predefined distributions, estimates the parameters of each candidate model, and computes the statistics of fit for each estimated model. The summary of the estimation and selection process is as follows. More details are available in the *SAS/ETS 9.22 User's Guide*.

1. The parameters of each distribution model are initialized either with the values you specified or with the default parameter initialization method provided for the distribution. The parameters are initialized by using the method of moments for all the predefined distributions except for the gamma and the Weibull distributions. For the gamma distribution, approximate maximum likelihood estimates are used. For the Weibull distribution, the method of percentile matching is used.

2. The parameters of each model are optimized using the maximum likelihood method. Given data that consist of N observations, the likelihood of the data is computed as $L = \prod_{i=1}^{N} f(y_i; \Theta)$, where $f(y_i; \Theta)$ denotes the probability density function (PDF) of the distribution. The log likelihood of the data is then maximized using a nonlinear optimization method. PROC SEVERITY enables you to choose various aspects of the optimization algorithm such as the optimization technique and the stopping criterion.

3. For each model the goodness of fit is computed using the specified selection criterion, and the model with the best value is selected. Seven different statistics of fit can be used as selection criteria: log likelihood, Akaike's information criterion (AIC), corrected Akaike's information criterion (AICC), Schwarz Bayesian information criterion (BIC), Kolmogorov-Smirnov statistic (KS), Anderson-Darling statistic (AD), and Cramér-von-Mises statistic (CvM). The first four are likelihood-based statistics, and the last three are computed by comparing the empirical distribution function (EDF) of the data with the estimated cumulative distribution function (CDF).

4. Various diagnostic output and ODS graphics for each of the candidate models are prepared to enable better comparison of the characteristics of the fitted models.

Table 1 Predefined SEVERITY Distributions

Name	Distribution	Parameters	PDF (f) and CDF (F)	
BURR	Burr	$\theta > 0, \alpha > 0,$ $\gamma > 0$	$f(x) =$	$\frac{\alpha \gamma z^{\gamma}}{x(1+z^{\gamma})^{(\alpha+1)}}$
			$F(x) =$	$1 - \left(\frac{1}{1+z^{\gamma}}\right)^{\alpha}$
EXP	Exponential	$\theta > 0$	$f(x) =$	$\frac{1}{\theta} e^{-z}$
			$F(x) =$	$1 - e^{-z}$
GAMMA	Gamma	$\theta > 0, \alpha > 0$	$f(x) =$	$\frac{z^{\alpha} e^{-z}}{x \Gamma(\alpha)}$
			$F(x) =$	$\frac{\gamma(\alpha, z)}{\Gamma(\alpha)}$
GPD	Generalized Pareto	$\theta > 0, \xi > 0$	$f(x) =$	$\frac{1}{\theta} (1 + \xi z)^{-1-1/\xi}$
			$F(x) =$	$1 - (1 + \xi z)^{-1/\xi}$
IGAUSS	Inverse Gaussian (Wald)	$\theta > 0, \alpha > 0$	$f(x) =$	$\frac{1}{\theta} \sqrt{\frac{\alpha}{2\pi z^3}} \, e^{\frac{-\alpha(z-1)^2}{2z}}$
			$F(x) =$	$\Phi\left((z-1)\sqrt{\frac{\alpha}{z}}\right) +$ $\Phi\left(-(z+1)\sqrt{\frac{\alpha}{z}}\right) e^{2\alpha}$
LOGN	Lognormal	μ (no bounds), $\sigma > 0$	$f(x) =$	$\frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log(x)-\mu}{\sigma}\right)^2}$
			$F(x) =$	$\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)$
PARETO	Pareto	$\theta > 0, \alpha > 0$	$f(x) =$	$\frac{\alpha\theta^{\alpha}}{(x+\theta)^{\alpha+1}}$
			$F(x) =$	$1 - \left(\frac{\theta}{x+\theta}\right)^{\alpha}$
WEIBULL	Weibull	$\theta > 0, \tau > 0$	$f(x) =$	$\frac{1}{x} \tau z^{\tau} e^{-z^{\tau}}$
			$F(x) =$	$1 - e^{-z^{\tau}}$

Notes:
1. $z = x/\theta$, wherever z is used.
2. θ denotes the scale parameter for all the distributions. For LOGN, $\log(\theta) = \mu$.
3. Parameters are listed in the order in which they are defined in the distribution model.
4. $\gamma(a, b) = \int_0^b t^{a-1} e^{-t} dt$ is the lower incomplete gamma function.
5. $\Phi(y) = \frac{1}{2}\left(1 + \text{erf}\left(\frac{y}{\sqrt{2}}\right)\right)$ is the standard normal CDF.

EXAMPLE: Fitting Predefined Distributions

The following DATA step statements simulate a sample from a lognormal distribution with population parameters $\mu = 1.5$ and $\sigma = 0.25$, and store the sample in the variable Y of a data set Work.Test_Sev1.

```
/*------------- Simple Lognormal Example -------------*/
data test_sev1(keep=y label='Simple Lognormal Sample');
    call streaminit(45678);
    label y='Response Variable';
    Mu = 1.5;
    Sigma = 0.25;
    do n = 1 to 100;
        y = exp(Mu) * rand('LOGNORMAL')**Sigma;
        output;
    end;
run;
```

The following statements enable ODS Graphics, fit all the predefined distribution models to the values of Y, and identify the best distribution according to the corrected Akaike's information criterion (AICC):

```
ods graphics on;

proc severity data=test_sev1;
    model y / crit=aicc;
run;
```

The ODS GRAPHICS ON statement enables PROC SEVERITY to generate the default graphics, the PROC SEVERITY statement specifies the input data set, and the MODEL statement specifies the variable to be modeled along with the model selection criterion.

Some of the default output displayed by this step is shown in Figure 1 through Figure 3. First, information about the input data set and the model selection table are displayed, as shown in Figure 1. The "Model Selection Table" displays the convergence status, the value of the selection criterion, and the selection status for each of the candidate models. The Converged column indicates whether the estimation process for a given distribution model has converged, might have converged, or failed. The Selected column indicates whether a given distribution has the best fit for the data according to the selection criterion. For this example, the lognormal distribution model is selected, because it has the lowest value for the selection criterion.

Figure 1 Data Set Information and Model Selection Table

```
                      The SEVERITY Procedure

                          Input Data Set

              Name              WORK.TEST_SEV1
              Label     Simple Lognormal Sample


                     Model Selection Table

                                      Corrected
                                       Akaike's
                                     Information
         Distribution    Converged     Criterion     Selected

         Burr            Yes           322.50845     No
         Exp             Yes           508.12287     No
         Gamma           Yes           320.50264     No
         Igauss          Yes           319.61652     No
         Logn            Yes           319.56579     Yes
         Pareto          Yes           510.28172     No
         Gpd             Yes           510.20576     No
         Weibull         Yes           334.82373     No
```

Next, comparative plots are prepared. These plots enable you to visually verify how the models differ from each other and from the nonparametric estimates. The first plot in Figure 2 compares the cumulative distribution function (CDF) estimates of all the models and the empirical distribution function (EDF) estimates. The second plot compares the probability density function (PDF) estimates of all the models and the nonparametric kernel and histogram estimates. Both plots indicate that the Exp (exponential), Pareto, and Gpd (generalized Pareto) distributions are a poor fit. The Weibull distribution is also a poor fit, although not as poor as Exp, Pareto, and Gpd. The Burr and Gamma differ significantly from the Igauss and Logn distributions in the central portion of the range of Y values, while the latter two fit the data almost identically.

Figure 2 Comparison of CDF and PDF Estimates of the Fitted Models
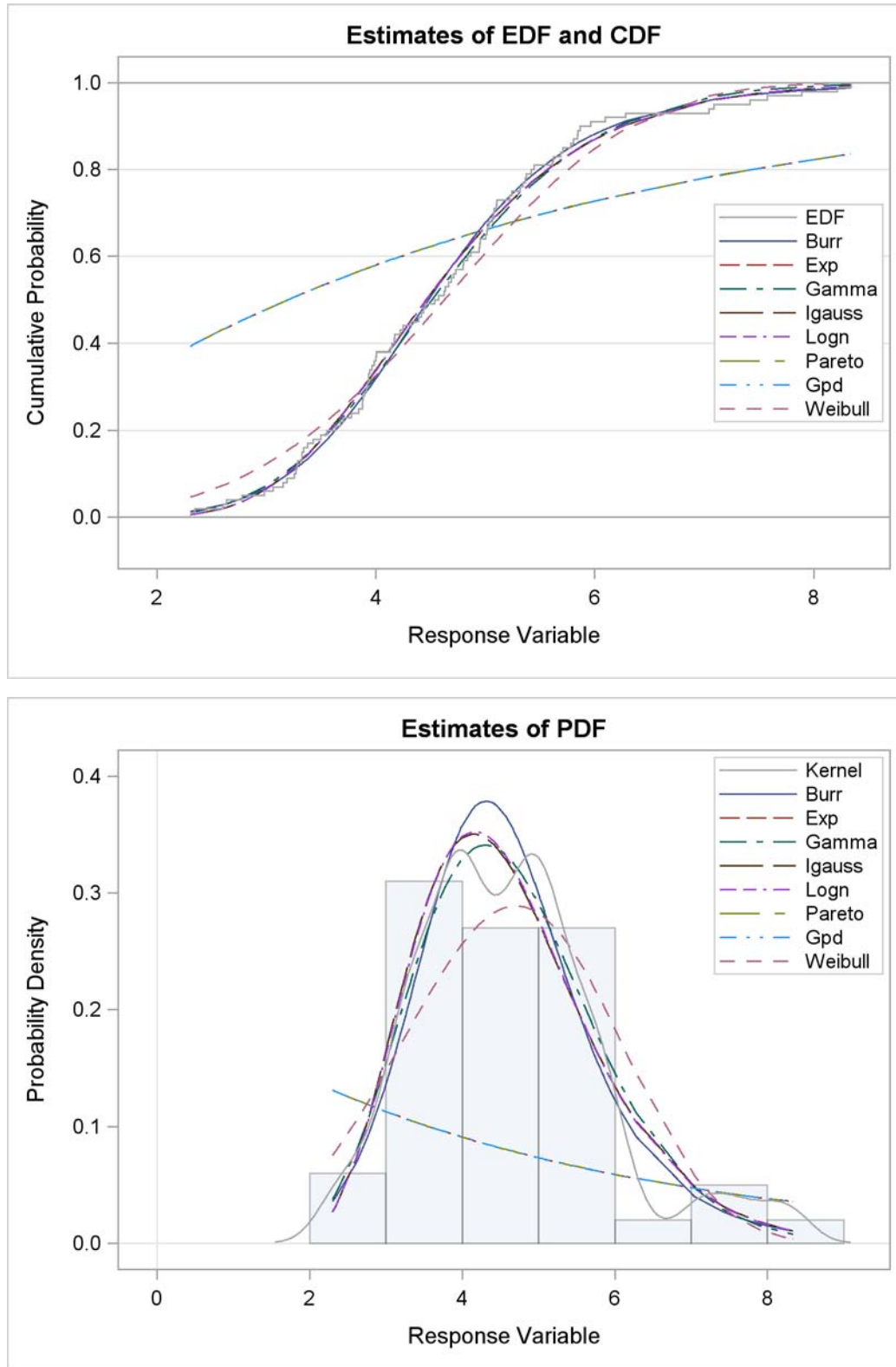
The comparative plots are followed by the estimation information for each of the candidate models. The information for the lognormal model, which is the best-fitting model, is shown in Figure 3. The first table displays a summary of the distribution. The second table displays the convergence status. This is followed by a summary of the optimization process which indicates the technique used, the number of iterations, the number of times the objective function was evaluated, and the log likelihood attained at the end of the optimization. Since the model with lognormal distribution has

converged, PROC SEVERITY displays its statistics of fit and parameter estimates. The estimates of *Mu*=1.49605 and *Sigma*=0.26243 are quite close to the population parameters of *Mu*=1.5 and *Sigma*=0.25 from which the sample was generated. The p-value for each estimate indicates the rejection of the null hypothesis that the estimate is 0, implying that both the estimates are significantly different from 0.

Figure 3 Estimation Details for the Lognormal Model

```
                        The SEVERITY Procedure

                        Distribution Information

        Name                                            Logn
        Description                      Lognormal Distribution
        Number of Distribution Parameters                   2

                Convergence Status for Logn Distribution

             Convergence criterion (GCONV=1E-8) satisfied.

                Optimization Summary for Logn Distribution

             Optimization Technique          Trust Region
             Number of Iterations                       2
             Number of Function Evaluations             8
             Log Likelihood                    -157.72104

                  Fit Statistics for Logn Distribution

        -2 Log Likelihood                           315.44208
        Akaike's Information Criterion              319.44208
        Corrected Akaike's Information Criterion    319.56579
        Schwarz's Bayesian Information Criterion    324.65242
        Kolmogorov-Smirnov Statistic                  0.50641
        Anderson-Darling Statistic                    0.31240
        Cramer-von Mises Statistic                    0.04353

                Parameter Estimates for Logn Distribution

                                    Standard            Approx
           Parameter     Estimate      Error   t Value  Pr > |t|

           Mu             1.49605     0.02651     56.43   <.0001
           Sigma          0.26243     0.01874     14.00   <.0001
```

ACCOUNTING FOR THE UNOBSERVED: TRUNCATION AND CENSORING EFFECTS

One of the key features of PROC SEVERITY is that it enables you to specify whether the severity event's magnitude is observable and if it is observable, then whether the exact value of the magnitude is known. An event that is unobservable when the magnitude is in certain intervals is said to have a truncation effect. An event whose magnitude is not known exactly, but is known to have a value in a certain interval, is said to have a censoring effect.

PROC SEVERITY allows a severity event to be left-truncated and right-censored. An event is said to be left-truncated if it is observed only when $Y > T$, where Y denotes the random variable for the magnitude and T denotes a random variable for the truncation threshold. An event is said to be right-censored if the exact magnitude Y is not known, but it is known to be greater than C, where C is a random variable for the censoring limit.

PROC SEVERITY assumes that the input data is given as a triplet (y_i, t_i, δ_i) for $i = 1, \ldots, N$, where N is the number of observations (in a BY group), y_i is the *observed* value (magnitude) of the response (event) variable, t_i is the left-truncation threshold, and δ_i is a right-censoring indicator. You can specify a set of values that indicate right-censoring; the default set is {0}. If δ_i is equal to one of those values, then the censoring limit c_i is assumed to be equal to the recorded value y_i. If δ_i is not equal to one of the indicator values or has a missing value, then y_i is assumed to be the exact event value; that is, the observation is uncensored. You can also specify a common (global) left-truncation threshold T_g such that $t_i = T_g$ for all i. Similarly, you can also specify a global right-censoring limit C_g. For each observation, y_i is then compared with C_g. If $y_i < C_g$, then $\delta_i = 1$ to indicate exact (uncensored) observation; otherwise, $\delta_i = 0$ to indicate right-censored observation.

For left-truncated data, PROC SEVERITY also enables you to provide additional information in the form of *probability of observability*. It is defined as the probability that the underlying severity event is observed (and recorded) for the specified left-truncation threshold value. For example, if you specify a value of 0.75, then for every 75 observations recorded above a specified threshold, 25 more events have happened with a severity value less than or equal to the specified threshold. Although the exact severity value of those 25 events is not known, PROC SEVERITY can use the information about the number of those events while estimating the distribution of Y.

See the *SAS/ETS 9.22 User's Guide* for details about the syntax for specifying left-truncation and right-censoring.

Likelihood with Censoring and Truncation

Specification of right-censoring and left-truncation affects the likelihood of the data. Let $f_\Theta(y)$ and $F_\Theta(y)$ denote the PDF and CDF, respectively, evaluated at y for a set of parameter values Θ. If probability of observability is not specified, then the likelihood of the data L is

$$L = \prod_{i \in E} f_\Theta(y_i) \prod_{j \in E_l} \frac{f_\Theta(y_j)}{1 - F_\Theta(t_j)} \prod_{k \in C} 1 - F_\Theta(c_k) \prod_{m \in C_l} \frac{1 - F_\Theta(c_m)}{1 - F_\Theta(t_m)}$$

where E denotes the set of uncensored observations that are not left-truncated, E_l denotes the set of uncensored observations that are left-truncated, C denotes the set of right-censored observations that are not left-truncated, and C_l denotes the set of right-censored observations that are left-truncated. Note that $(E \cup E_l) \cap (C \cup C_l) = \emptyset$.

If the probability of observability $p (0 < p < 1)$ is specified, then for each left-truncated observation with truncation threshold t, there exist $(1 - p)/p$ observations with a response variable value that is less than or equal to t. The likelihood of the data L is as follows:

$$L = \prod_{i \in E} f_\Theta(y_i) \prod_{j \in E_l} f_\Theta(y_j) F_\Theta(t_j)^{\frac{1-p}{p}} \prod_{k \in C} 1 - F_\Theta(c_k) \prod_{m \in C_l} (1 - F_\Theta(c_m)) F_\Theta(t_m)^{\frac{1-p}{p}}$$

EDF Estimation with Censoring and Truncation

Specification of right-censoring and left-truncation also affects estimation of the empirical distribution function (EDF), used for computing EDF-based statistics of fit. The Kaplan-Meier (KM) estimator, also known as the product-limit estimator, was first introduced by Kaplan and Meier (1958) for censored data. Lynden-Bell (1971) derived a similar estimator for left-truncated data. PROC SEVERITY uses the definition that combines both censoring and truncation information (Klein and Moeschberger 1997, Lai and Ying 1991). The EDF estimate $\hat{F}_n(y_i)$ at observation i is computed as

$$\hat{F}_n(y_i) = 1 - \prod_{\tau \le y_i} \left(1 - \frac{n_\tau}{R_n(\tau)} \right)$$

where n_τ and $R_n(\tau)$ are defined as follows:

- $n_\tau = \sum_{k=1}^{N} I[y_k = \tau \text{ and } \delta_k \ne 0]$, which is the number of uncensored observations with response variable value equal to τ.

- $R_n(\tau) = \sum_{k=1}^{N} I[y_k \ge \tau > t_k]$, which is the size (cardinality) of the risk set at τ. The term *risk set* has its origins in survival analysis; it contains the events that are at the risk of failure at a given time, τ. In other words, it contains the events that have survived up to time τ and might fail at or after τ. For PROC SEVERITY, *time* is equivalent to the magnitude of the event and *failure* is equivalent to an uncensored and observable event, where observable means it satisfies the left-truncation threshold.

The preceding estimator might not work well if the risk set size becomes very small. PROC SEVERITY enables you to use a modification of the estimator proposed by Lai and Ying (1991), which computes the EDF estimate as

$$\hat{F}_n(y_i) = 1 - \prod_{\tau \le y_i} \left(1 - \frac{n_\tau}{R_n(\tau)} \cdot I[R_n(\tau) \ge cN^\alpha] \right)$$

Use of appropriate values c and α ensures that the effect of sufficiently small risk sets is ignored.

EXAMPLE: Truncation and Censoring in Insurance Industry

The following DATA step prepares a data set to simulate a scenario that is typically encountered by an automobile insurance company. This data was used by Klugman and Rioux (2006). Although inspired by the liability claims seen by an insurance company, the data are not from any specific insurance company. The values of the variable LOSS represent the loss values on claims that are reported to the insurance company. The variable DED records the deductible on the insurance policy, which serves as the left-truncation of LOSS. The CAPPED variable indicates whether the loss exceeds the policy limit. If CAPPED is equal to 1, then the actual value of LOSS is greater than the recorded value; that is, LOSS is right-censored. If CAPPED is not equal to 1, then the recorded value of LOSS is the actual value of the loss (the payment made by the insurance company plus the deductible).

```
/*----- Insurance Loss Data -----*/
data test_sev2(label='A Sample With Censoring and Truncation');
    label loss='Observed Loss' ded='Deductible' capped='Exceeds Limit? (1:Yes,0:No)';
    input loss ded capped @@;
datalines;
 184 100 0  1100 100 1  1100 100 1   793 100 0  1100 100 1  1100 100 1   960 100 0
1100 100 1  1100 100 1  1100 100 1  3100 100 1  3100 100 1   872 100 0  3100 100 1
 547 100 0  1495 100 0   401 100 0  1409 100 0  3100 100 1  1556 100 0   182 100 0
1744 100 0  1383 100 0  1301 100 0  1141 100 0  1148 100 0  1751 100 0   771 100 0
1044 100 0  4510 100 0  1215 250 0  1250 250 1  1178 250 0  1060 250 0   491 250 0
 885 250 0   296 250 0  1105 250 0   381 250 0   974 250 0   685 250 0  1250 250 1
 505 250 0  1811 250 0   331 250 0  1647 250 0  1564 250 0   495 250 0  3250 250 1
 674 250 0  1614 250 0   931 250 0   693 250 0  1259 250 0  1441 250 0  3250 250 1
2880 250 0  1294 250 0   616 250 0  4254 250 0   929 250 0  1737 250 0   825 250 0
2275 250 0   927 250 0   913 250 0  1016 250 0  1434 250 0  2031 250 0   553 250 0
1213 500 0  1500 500 1   840 500 0  1500 500 1  1500 500 1  1500 500 1  1500 500 1
1500 500 1  1156 500 0  1500 500 1  2671 500 0  3469 500 0  1064 500 0  2080 500 0
1807 500 0  3500 500 1  2263 500 0  1768 500 0   514 500 0  3500 500 1  1240 500 0
 653 500 0   708 500 0  1131 500 0  5500 500 1  1372 500 0   601 500 0  1122 500 0
2752 500 0  1200 500 0
;
run;
```

The following statements use the AICC criterion to analyze which of the eight predefined distributions has the best fit for the data:

```
proc severity data=test_sev2 print=all;
    model loss(lt=ded rc=capped(1)) / crit=aicc;
run;
```

The MODEL statement uses the LT= option to specify the left-truncation variable and the RC= option to specify the right-censoring indicator variable. The numbers in the parentheses after the variable CAPPED specify the set of values that indicate right-censoring, which is {1} in this case. The PRINT= option in the PROC SEVERITY statement requests that all the displayed output be prepared.

Some of the key results prepared by PROC SEVERITY are shown in Figure 4 through Figure 9. The descriptive statistics of LOSS are shown in the second table of Figure 4. In addition to the range, mean, and standard deviation of LOSS, the table also indicates the number of observations that are right-censored, left-truncated, and both right-censored and left-truncated. The "Model Selection Table" in Figure 4 shows that models with all the candidate distributions have converged and that the Logn (lognormal) model has the best fit for the data according to the AICC criterion.

Figure 4 Summary Results for the Truncated and Censored Data

```
                      The SEVERITY Procedure

                         Input Data Set

              Name                           WORK.TEST_SEV2
              Label     A Sample With Censoring and Truncation


                 Descriptive Statistics for Variable loss

         Number of Observations                                100
         Number of Observations Used for Estimation            100
         Minimum                                         182.00000
         Maximum                                              5500
         Mean                                                 1478
         Standard Deviation                             982.23759
         Number of Left Truncated Observations                100
         Number of Right Censored Observations                 25
         Number of Left Truncated and Right Censored Observations   25
```

Figure 4 *continued*

```
                         Model Selection Table

                                       Corrected
                                        Akaike's
                                      Information
             Distribution    Converged    Criterion    Selected

             Burr            Yes               1257     No
             Exp             Yes               1258     No
             Gamma           Yes               1259     No
             Igauss          Yes               1259     No
             Logn            Yes               1257     Yes
             Pareto          Yes               1261     No
             Gpd             Yes               1261     No
             Weibull         Yes               1260     No
```

PROC SEVERITY also prepares a table that shows all the fit statistics for all the candidate models. It is useful to see which model would fit best according to each of the criteria. The table prepared for this example is shown in Figure 5. It indicates that none of the models is the best fit according to all the criteria. The Burr model is chosen by the log likelihood and the EDF-based statistics, the lognormal model is chosen by AIC and AICC, and the exponential model is chosen by the BIC criterion.

Figure 5 Comparing All Statistics of Fit for the Truncated and Censored Data

```
                       All Fit Statistics Table

                    -2 Log
   Distribution   Likelihood      AIC         AICC        BIC         KS

   Burr            1251*         1257        1257        1265      0.82990*
   Exp             1256         1258        1258        1261*     0.89249
   Gamma           1255         1259        1259        1264      1.03554
   Igauss          1255         1259        1259        1264      0.92024
   Logn            1253         1257*       1257*       1262      0.93747
   Pareto          1256         1260        1261        1266      0.89503
   Gpd             1256         1260        1261        1266      0.89248
   Weibull         1256         1260        1260        1265      1.01407

                       All Fit Statistics Table

                 Distribution       AD          CvM

                 Burr            0.83717*     0.07795*
                 Exp            1.55729      0.26230
                 Gamma          0.97066      0.14298
                 Igauss         1.35559      0.10962
                 Logn           0.93736      0.09946
                 Pareto         1.58336      0.26756
                 Gpd            1.55732      0.26230
                 Weibull        1.07105      0.16237
```

The parameter estimates of the three models that are chosen by different criteria are shown in Figure 6. The parameter estimates are all significantly different from 0.

Figure 6 Parameter Estimates for the Truncated and Censored Data

```
               Parameter Estimates for Burr Distribution

                                   Standard                 Approx
          Parameter     Estimate      Error    t Value    Pr > |t|

          Theta             1208    461.47060     2.62      0.0103
          Alpha          0.91341      0.51146     1.79      0.0772
          Gamma          2.07127      0.50666     4.09      <.0001

               Parameter Estimates for Exp Distribution

                                   Standard                 Approx
          Parameter     Estimate      Error    t Value    Pr > |t|

          Theta             1598    185.42351     8.62      <.0001
```

Figure 6 *continued*

```
            Parameter Estimates for Logn Distribution

                                    Standard                 Approx
            Parameter     Estimate     Error     t Value     Pr > |t|

            Mu             7.16304     0.10044      71.32      <.0001
            Sigma          0.85888     0.09074       9.47      <.0001
```

You can take a closer look at the three competing models by using some diagnostic plots. The following statements fit only Burr, lognormal, and exponential models to the data:

```
ods graphics on;
proc severity data=test_sev2 print=all plots(markcensored marktruncated)=pp;
    model loss(lt=ded rc=capped(1)) / crit=aicc;

    dist burr init=(theta=1208 alpha=0.9134 gamma=2.072);
    dist logn init=(mu=7.163 sigma=0.86);
    dist exp  init=(theta=1598);
run;
```

Each candidate distribution needs to be specified by using a separate DIST statement. The convergence is made faster by using the INIT= option in the DIST statement to provide initial values that are the final parameter estimates obtained from the previous run of PROC SEVERITY. Use of the INIT= option overrides the default parameter initialization that is defined for each distribution. The PLOTS= option in the PROC SEVERITY statement requests that the P-P plots for each candidate distribution be prepared in addition to the default plots. It also instructs the procedure to mark the left-truncated and right-censored observations in the CDF plot.

The plot that compares EDF and CDF estimates is shown in Figure 7. When left-truncation is specified, both the EDF and CDF estimates are conditional on the response variable being greater than the smallest left-truncation threshold in the sample. Notice the markers close to the X-axis of the plot. These indicate the values of LOSS that are left-truncated or right-censored.

Figure 7 EDF and CDF Estimates for the Truncated and Censored Data
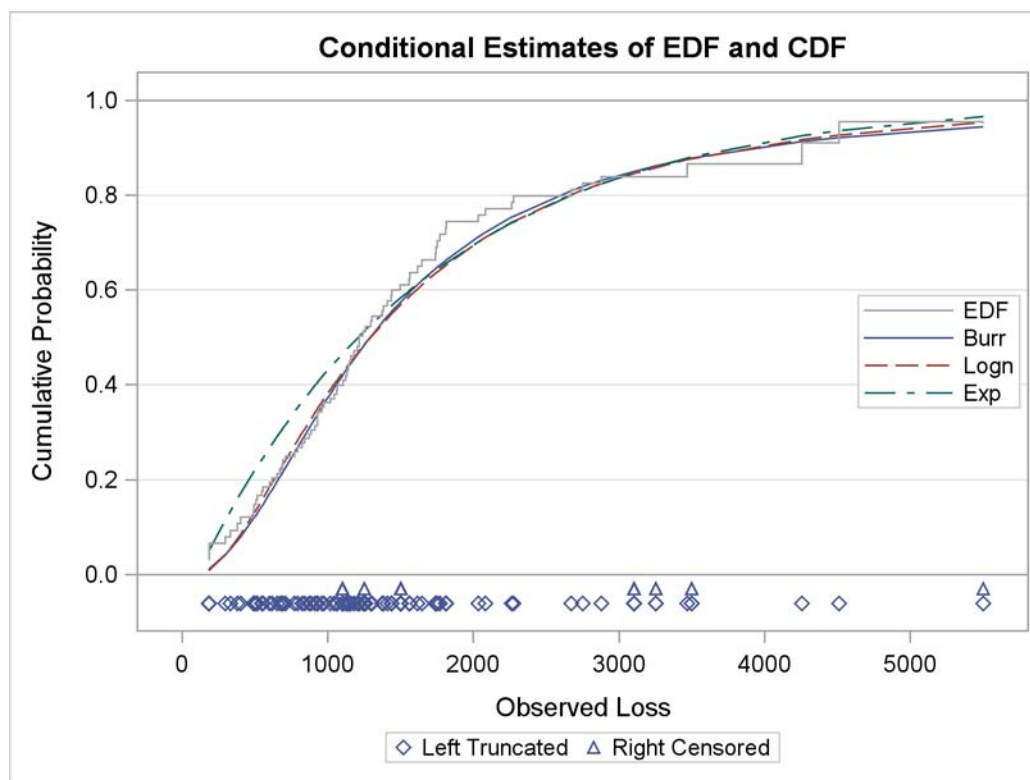
In addition to the comparative plots, PROC SEVERITY produces a P-P plot for each of the models that has not failed to converge. The P-P plot is a scatter plot of the EDF and the CDF estimates. The model for which the points are scattered closer to the unit-slope reference line is a better fit. The P-P plots for the Burr and lognormal models are shown in Figure 8 and Figure 9, respectively. Burr distribution's CDF estimates are a better fit for the EDF estimates in the middle region, but the lognormal distribution fits slightly better in the tail regions. So, based on the business needs and preferences, you can choose either of the two.

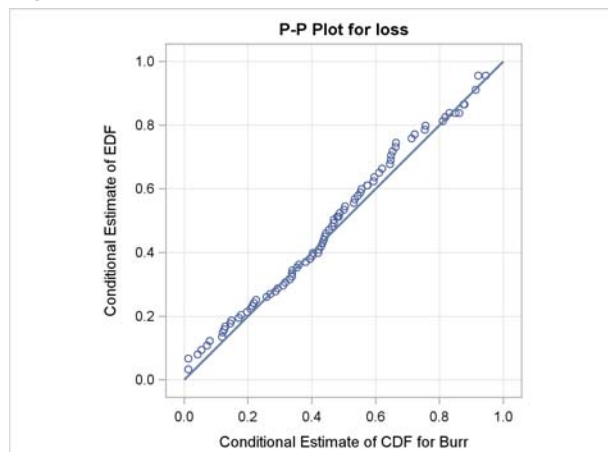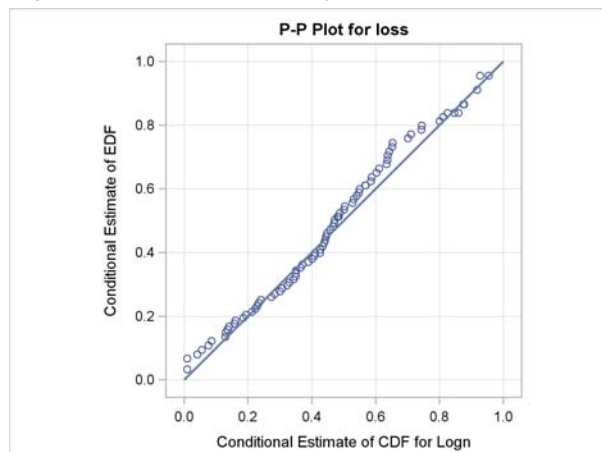**Figure 8**   P-P Plot for the Burr Model

Figure 9   P-P Plot for the Lognormal Model

ACCOUNTING FOR THE EXTERNAL FORCES: REGRESSION EFFECTS

The SEVERITY procedure enables you to estimate the effects of regressor (exogenous) variables while fitting a distribution model if the distribution has a scale parameter or a log-transformed scale parameter.

Let x_j $(j = 1, \ldots, k)$ denote the k regressor variables. Let β_j denote the regression parameter that corresponds to the regressor x_j. The regression effects are modeled by extending the error model to the following form:

$$Y \sim \exp(\sum_{j=1}^{k} \beta_j x_j) \cdot \mathcal{F}(\Theta)$$

Under this model, the distribution of Y is valid and belongs to the same parametric family as $\mathcal{F}$ if and only if $\mathcal{F}$ has a scale parameter. Let θ denote the scale parameter and Ω denote the set of nonscale distribution parameters of $\mathcal{F}$. Then the model can be rewritten as

$$Y \sim \mathcal{F}(\theta, \Omega)$$

such that θ is affected by the regressors as

$$\theta = \theta_0 \cdot \exp(\sum_{j=1}^{k} \beta_j x_j)$$

where θ_0 is the *base* value of the scale parameter. Thus, the regression model consists of the following parameters: θ_0, Ω, and $\beta_j (j = 1, \ldots, k)$.

Given this form of the model, distributions without a scale parameter cannot be considered when regression effects are to be modeled. If a distribution does not have a direct scale parameter, then PROC SEVERITY accepts it only if it has a log-transformed scale parameter—that is, if it has a parameter $p = \log(\theta)$.

EXAMPLE: Modeling Regression Effects

This example simulates a scenario of modeling claims against a worker's compensation insurance policy. An insurance company wants to analyze the effect of regressors on the distribution of the severity of the claims that it paid for one of its corporate clients. The regressors are usually referred to as *key risk indicators* (KRIs) in insurance and banking industries.

The following SAS statements prepare the sample data. The first DATA step creates the Work.WorkComp data set that records the loss amount for each claim and the year in which the loss was incurred. The second DATA step creates the Work.Kri data set that records the annual values of six KRIs: annual revenue in millions (revenue), ratio of temporary to permanent employees (tempratio), average number of customer complaints (complaints), number of sick days per employee (sickdays), attrition rate (attrition), and average number of employees (nemp). The final DATA step combines the loss data with KRI data to prepare the Work.Test_Sev3 data set.

```
data workcomp(label='Workers Compensation Claims');
    input year loss @@;
    datalines;
01  1830  02  1031  02  1075  02  1113  02  1115  02  1119  02  1130  02  1131  02  1214
02  1260  02  1261  02  1263  02  1318  02  1319  02  1370  02  1370  02  1386  02  1440
02  1512  02  1631  02  1717  02  1748  02  1828  02  1912  02  2012  02  2030  02  2077
02  2093  02  2248  02  2263  02  2288  02  2293  02  2328  02  2366  02  2376  02  2442
02  2508  02  2603  02  2608  02  2718  02  2759  02  2767  02  2777  02  2826  02  2828
02  2838  02  2868  02  2896  02  2906  02  2906  02  2972  02  2988  02  3030  02  3075
02  3208  02  3218  02  3292  02  3391  02  3459  02  3503  02  3548  02  3552  02  3685
02  3747  02  3823  02  3833  02  3870  02  3900  02  3929  02  4020  02  4021  02  4082
02  4098  02  4147  02  4155  02  4249  02  4275  02  4524  02  4766  02  4788  02  4857
02  4970  02  4974  02  5149  02  5484  02  5770  02  5890  02  6455  02  6461  02  7073
02  7423  02  7854  02  8040  02  8449  02  8607  02 10127  02 10667  02 10939  02 12228
02 19445  03  1037  03  2516  03  2540  03  2662  03  3070  03  3293  03  6419  03  7188
03  8914  03 10055  03 18792  03 18897  03 19730  03 21833  03 30161  03 31942  03 40741
03 57995  03 62152  03 65814  04  1134  04  1248  04  2255  04  2548  04  2577  04  2759
04  2835  04  2864  04  3070  04  3332  04  3644  04  5110  04  5185  04  5671  04  6412
04  9124  04 13295  04 19505  04 28209  04 56688  05  1007  05  1698  05  2185  05  2786
05  3329  05  4686  05  4834  05  8940  05  9878  05 11602  05 15411
;
run;

data kri;
    input year revenue tempratio complaints sickdays attrition nemp @@;
    datalines;
01 20.0 0.15 4.91 2.56 0.07 53.44   02 20.0 0.51 8.96 3.01 0.08 12.37
03 19.5 0.71 9.01 4.65 0.16 44.17   04 19.0 0.59 4.44 4.26 0.05 39.47
05 20.5 0.19 2.49 0.06 0.16 15.71
;
run;

data test_sev3(drop=year); merge workcomp kri; by year; run;
```

The following PROC SEVERITY step fits regression models with each of the predefined distributions to the data that are recorded in the Work.Test_Sev3 data set. The regressors are specified in the MODEL statement. The NLOPTIONS statement sets the maximum number of iterations that can be used by the nonlinear optimizer to obtain maximum likelihood estimates of the model parameters.

```
proc severity data=test_sev3 outest=est3 print=all plots=none;
    model loss = revenue tempratio complaints sickdays attrition nemp / crit=ad;
    nloptions maxiter=100;
run;
```

Some of the key results prepared by PROC SEVERITY are shown in Figure 10 and Figure 11. The "Model Selection Table," shown in Figure 10, indicates that the Burr model is the best model according to the Anderson-Darling statistic.

Figure 10 Model Selection Table for the Regression Example

```
                    The SEVERITY Procedure

                    Model Selection Table

                                  Anderson-
                                   Darling
          Distribution   Converged  Statistic   Selected

          Burr           Yes        1.26719     Yes
          Exp            Yes        7.45680     No
          Gamma          Yes        3.84381     No
          Igauss         Yes        2.48764     No
          Logn           Yes        2.29535     No
          Pareto         Yes        7.45820     No
          Gpd            Yes        7.45680     No
          Weibull        Yes        4.63597     No
```

The initial values and final estimates for the parameters of the Burr model are shown in Figure 11. The "Initial Parameter Values and Bounds" table indicates that regressors revenue and sickdays are redundant; that is, they are linearly dependent on other regressors. Redundant regressors are removed from all the models prior to estimating the parameters. The "Parameter Estimates" table indicates that p-values for regressors complaints and nemp are too high to reject the null hypothesis that their coefficients are 0.

Figure 11 Initial and Final Parameter Estimates for the Burr Model with Regressors

```
                   Initial Parameter Values and
                   Bounds for Burr Distribution

                          Initial          Lower           Upper
           Parameter        Value          Bound           Bound

           Theta             1542     1.05367E-8           Infty
           Alpha          1.82063     1.05367E-8           Infty
           Gamma          2.00000     1.05367E-8           Infty
           revenue       Redundant             .               .
           tempratio      2.64718    -709.78271       709.78271
           complaints    -0.08609    -709.78271       709.78271
           sickdays      Redundant             .               .
           attrition      8.06157    -709.78271       709.78271
           nemp           0.00422    -709.78271       709.78271


               Parameter Estimates for Burr Distribution

                                   Standard                 Approx
           Parameter    Estimate      Error    t Value     Pr > |t|

           Theta       684.08880   286.84830      2.38       0.0184
           Alpha         0.68884     0.27171      2.54       0.0123
           Gamma         2.61451     0.48605      5.38       <.0001
           tempratio     2.11688     1.20947      1.75       0.0822
           complaints   -0.05504     0.06001     -0.92       0.3606
           attrition     7.90205     2.81177      2.81       0.0056
           nemp          0.00451     0.01041      0.43       0.6655
```

The search for significant regressors is usually done by eliminating the insignificant regressors one at a time starting with the one that has the highest p-value, which in this case is nemp. The following PROC SEVERITY step includes only three regressors in the model: tempratio, complaints, and attrition. In addition to fitting the Burr model, it also fits the lognormal model, which was the closest contender to the Burr model when all regressors were included. The INEST= option in the PROC SEVERITY statement specifies the data set that contains initial values of the parameters. The data set Work.Est3 contains the final estimates that are prepared by the previous PROC SEVERITY step.

```
    proc severity data=test_sev3 inest=est3 print=all plots=none;
        model loss = tempratio complaints attrition;
        dist burr;
        dist logn;
    run;
```

The "All Fit Statistics Table," shown in Figure 12, indicates that the lognormal model is chosen by most of the likelihood-based statistics, whereas the Burr model is chosen by all the EDF-based statistics. The insurance company can choose either model based on the statistics of fit that align with its policies or preferences.

Figure 12 Comparison of Statistics of Fit for Burr and Lognormal Regression Models

```
                       The SEVERITY Procedure

                      All Fit Statistics Table

                 -2 Log
    Distribution Likelihood     AIC         AICC        BIC           KS

    Burr             2859*      2871        2871        2889      1.01507*
    Logn             2860       2870*       2870*       2885*     1.33957

                      All Fit Statistics Table

                   Distribution        AD           CvM

                   Burr             1.28932*      0.20534*
                   Logn             2.32926       0.38933
```

The final parameter estimates of the Burr and lognormal models are shown in Figure 13. Estimates of all the distribution and regression parameters are significant for both models. The estimate that is reported for the parameter *Theta* of the Burr distribution is the base value for its scale parameter θ. Let t, c, and a denote the values observed in a particular year for regressors tempratio, complaints, and attrition, respectively. If the Burr distribution is chosen to model the loss, then the effective value of the scale parameter varies with the observed values of regressors as $\theta \approx 689.55 \exp(2.52\,t - 0.076\,c + 8.24\,a)$. The estimate that is reported for the parameter *Mu* of the lognormal distribution is the base value for μ, which is the log-transformed scale parameter. If the lognormal distribution is chosen to model the loss, then the effective value of the parameter μ varies with regressor values as $\mu \approx 6.724 + 3.08\,t - 0.106\,c + 8.51\,a$. For each model, the attrition rate has the largest impact on the scale of the distribution. As the proportion of temporary employees or the attrition rate increases, the scale of the distribution increases, which means that the distribution becomes more spread out (fatter tails), leading to higher probability (risk) of incurring larger losses.

Figure 13 Parameter Estimates for the Burr and Lognormal Regression Models

```
             Parameter Estimates for Burr Distribution

                                Standard             Approx
         Parameter      Estimate      Error    t Value    Pr > |t|

         Theta       689.55059    292.00821       2.36      0.0195
         Alpha         0.68086      0.26747       2.55      0.0120
         Gamma         2.62537      0.48905       5.37      <.0001
         tempratio     2.52092      0.80887       3.12      0.0022
         complaints   -0.07600      0.03567      -2.13      0.0348
         attrition     8.24398      2.73147       3.02      0.0030


             Parameter Estimates for Logn Distribution

                                Standard             Approx
         Parameter      Estimate      Error    t Value    Pr > |t|

         Mu            6.72395      0.36645      18.35      <.0001
         Sigma         0.79242      0.04637      17.09      <.0001
         tempratio     3.08115      0.62758       4.91      <.0001
         complaints   -0.10598      0.03480      -3.05      0.0028
         attrition     8.50982      1.83739       4.63      <.0001
```

WIDENING THE HORIZON: DEFINING YOUR OWN DISTRIBUTIONS

One of the most powerful features of the SEVERITY procedure is that it enables you to define and fit any arbitrary continuous probability distribution of your choice if none of the eight predefined distributions satisfy your need.

For PROC SEVERITY, a severity *distribution model* consists of a set of functions and subroutines that are defined by using the FCMP procedure. The FCMP procedure is part of Base SAS software. Each function or subroutine must be named as <*distribution-name*>_<*keyword*>, where *distribution-name* is the identifying short name of the distribution and the *keyword* suffix identifies one of the functions or subroutines. The total length of the name should not exceed 32 characters. The summary of all the recognized keyword suffixes and their expected behavior is given in Table 2.

Table 2 List of Functions and Subroutines That Define a Distribution Model

Keyword Suffix	Type	Required	Expected to Return
CDF	Function	YES	Cumulative distribution function value
PDF	Function	YES	Probability distribution function value
CDFGRADIENT	Subroutine	NO	Gradient of the CDF
CDFHESSIAN	Subroutine	NO	Hessian of the CDF
CONSTANTPARM	Subroutine	NO	Constant parameters
DESCRIPTION	Function	NO	Description of the distribution
LOWERBOUNDS	Subroutine	NO	Lower bounds on parameters
PARMINIT	Subroutine	NO	Initial values for parameters
PDFGRADIENT	Subroutine	NO	Gradient of the PDF
PDFHESSIAN	Subroutine	NO	Hessian of the PDF
SCALETRANSFORM	Function	NO	Type of relationship between the scale parameter and the first distribution parameter
UPPERBOUNDS	Subroutine	NO	Upper bounds on parameters

Each function or subroutine must have a specific signature, which consists of the number of arguments, sequence and types of arguments, and return value type. See the *SAS/ETS 9.22 User's Guide* for more details. All the predefined distributions have been defined in this manner by using PROC FCMP. They are available in the Sashelp.Svrtdist library. An inventory of many other useful severity distributions can be found in Klugman, Panjer, and Willmot (1998). The following subsection illustrates how you can define and fit your own model with a distribution that is typically required by the insurance industry.

EXAMPLE: Defining a Model for Mixed Tail Distributions

In some applications, a few severity values tend to be extreme as compared to the typical values. The extreme values represent the worst case scenarios and cannot be discarded as outliers. Instead, their distribution must be modeled to prepare for their occurrences. In such cases, it is often useful to fit one distribution to the non-extreme values and another distribution to the extreme values. The *mixed tail* distribution mixes two distributions: one for the *body* region, which contains the non-extreme values, and another for the *tail* region, which contains the extreme values. The tail distribution is typically a generalized Pareto distribution (GPD), because it is usually good for modeling the conditional excess severity above a threshold. The body distribution can be any distribution. The following definitions are used in describing a generic formulation of a mixed tail distribution:

$g(x)$	PDF of the body distribution
$G(x)$	CDF of the body distribution
$h(x)$	PDF of the tail distribution
$H(x)$	CDF of the tail distribution
θ	scale parameter for the body distribution
Ω	set of nonscale parameters for the body distribution
ξ	shape parameter for the GPD tail distribution
x_r	normalized value of the response variable at which the tail starts
p_n	mixing probability

Given these notations, the PDF $f(x)$ and the CDF $F(x)$ of the mixed tail distribution are defined as

$$f(x) = \begin{cases} \frac{p_n}{G(x_b)} g(x) & \text{if } x \leq x_b \\ (1 - p_n) h(x - x_b) & \text{if } x > x_b \end{cases}$$

$$F(x) = \begin{cases} \frac{p_n}{G(x_b)} G(x) & \text{if } x \leq x_b \\ p_n + (1 - p_n) H(x - x_b) & \text{if } x > x_b \end{cases}$$

where $x_b = \theta x_r$ is the value of the response variable at which the tail starts.

These definitions indicate the following:

- The body distribution is conditional on $X \leq x_b$, where X denotes the random response variable.

- The tail distribution is the generalized Pareto distribution of the $(X - x_b)$ values.

- The probability that a response variable value belongs to the body is p_n. Consequently the probability that the value belongs to the tail is $(1 - p_n)$.

The parameters of this distribution are θ, Ω, ξ, x_r, and p_n. The scale of the GPD tail distribution θ_t is computed as

$$\theta_t = \frac{G(x_b; \theta, \Omega)}{g(x_b; \theta, \Omega)} \frac{(1 - p_n)}{p_n} = \theta \frac{G(x_r; \theta = 1, \Omega)}{g(x_r; \theta = 1, \Omega)} \frac{(1 - p_n)}{p_n}$$

The parameter x_r is usually estimated by using a tail index estimation algorithm such as the Hill's algorithm (Danielsson et al. 2001). The parameter p_n is usually determined by the domain expert and is based on the fraction of losses that are expected to belong to the tail.

The following SAS statements define the LOGNGPD distribution model for a mixed tail distribution with the lognormal distribution as the body distribution and GPD as the tail distribution:

```
proc fcmp library=sashelp.svrtdist outlib=work.sevexmpl.models;
    function LOGNGPD_DESCRIPTION() $64;
        return("Lognormal Body-GPD Tail Distribution");
    endsub;

    function LOGNGPD_PDF(x, Mu,Sigma,Xi,Xr,Pn);
        cutoff = exp(Mu) * Xr;
        p = CDF('LOGN',cutoff, Mu, Sigma);
        if (x < cutoff + constant('MACEPS')) then
            return ((Pn/p)*PDF('LOGN', x, Mu, Sigma));
        gpd_scale = p*((1-Pn)/Pn)/PDF('LOGN', cutoff, Mu, Sigma);
        h = (1+Xi*(x-cutoff)/gpd_scale)**(-1-(1/Xi))/gpd_scale;
        return ((1-Pn)*h);
    endsub;

    function LOGNGPD_CDF(x, Mu,Sigma,Xi,Xr,Pn);
        cutoff = exp(Mu) * Xr;
        p = CDF('LOGN',cutoff, Mu, Sigma);
        if (x < cutoff + constant('MACEPS')) then
            return ((Pn/p)*CDF('LOGN', x, Mu, Sigma));
        gpd_scale = p*((1-Pn)/Pn)/PDF('LOGN', cutoff, Mu, Sigma);
        H = 1 - (1 + Xi*((x-cutoff)/gpd_scale))**(-1/Xi);
        return (Pn + (1-Pn)*H);
    endsub;

    function LOGNGPD_SCALETRANSFORM() $3;
        return ("LOG");
    endsub;

    subroutine LOGNGPD_CONSTANTPARM(Xr,Pn);
    endsub;

    subroutine LOGNGPD_PARMINIT(dim,x[*],nx[*],F[*], Mu,Sigma,Xi,Xr,Pn);
        outargs Mu,Sigma,Xi,Xr,Pn;
        array m[2] / nosymbols;
        array xe[1] / nosymbols;
        array nxe[1] / nosymbols;
        /* Initialize lognormal parameters */
        call logn_parminit(dim, x, nx, F, Mu, Sigma);
        if (not(missing(Mu))) then do;
            /* Initialize Xr and Pn */
            _status_ = .;
            call streaminit(56789);
            Xb = svrtutil_hillcutoff(dim, x, 100, 25, _status_);
            if (missing(_status_) or _status_ = 1) then
                Xb = svrtutil_percentile(Pn, dim, x, F);
            Xr = Xb/exp(Mu);
            Pn = 0.8;
            /* Initialize GPD's shape parameter using excess values */
            i = 1;
            do while (i <= dim and x[i] < Xb+constant('MACEPS'));
                i = i + 1;
            end;
            dime = dim-i+1;
            call dynamic_array(xe, dime);
            call dynamic_array(nxe, dime);
            j = 1;
            do while(i <= dim);
                xe[j] = x[i] - Xb;
                nxe[j] = nx[i];
                i = i + 1;
                j = j + 1;
            end;
            call gpd_parminit(dime, xe, nxe, F, theta_gpd, Xi);
        end;
        else do;
            Xi=.; Xr=.; Pn=.;
        end;
    endsub;
```

```
       subroutine LOGNGPD_LOWERBOUNDS(Mu,Sigma,Xi,Xr,Pn);
           outargs Mu,Sigma,Xi,Xr,Pn;
           Mu    = .; /* Mu has no lower bound */
           Sigma = 0; /* Sigma > 0 */
           Xi    = 0; /* Xi > 0 */
       endsub;
   quit;
```

The statements define the two functions that are required of any distribution model (LOGNGPD_PDF and LOGNGPD_CDF). The statements also define other subroutines, which are usually optional, but are required for the proper definition of the mixed tail distribution. The name of each function or subroutine starts with the model's short or identifying name, which is 'LOGNGPD' in this case, followed by an underscore '_', followed by a keyword suffix such as 'PDF'.

Following is the description of each function and subroutine suffix defined in this example:

- The DESCRIPTION suffix enables you to provide a label to the distribution model. It is used by PROC SEVERITY in the displayed output.

- The PDF and CDF suffixes define functions that return the probability density function and cumulative distribution function values, respectively, given the values of the random variable and the distribution parameters.

- The SCALETRANSFORM suffix specifies that the first parameter of the distribution is a log-transformed scale parameter. You must define this subroutine in order to use this distribution in a model with regression effects.

- The CONSTANTPARM suffix specifies that parameters *Xr* and *Pn*, which represent parameters x_r and p_n of the mixed tail distribution, respectively, are *constant* parameters. A constant parameter is not subject to optimization and retains its initial value throughout the parameter estimation process. It must be assigned a valid, nonmissing initial value. The signature of the subroutine must list only the constant parameters.

- The PARMINIT suffix defines a subroutine that returns the initial values for the parameters by using the sample data or the empirical distribution function (EDF) estimate computed from it. The OUTARGS statement must be used to specify the arguments that represent distribution parameters as the output arguments. In this example, arguments *Mu*, *Sigma*, *Xi*, *Xr*, and *Pn* represent distribution parameters μ, σ, ξ, x_r, and p_n, respectively. The statements that follow the OUTARGS statement assign initial values to the parameter arguments.

  First, the parameters of the lognormal body (μ and σ) are initialized using the parameter initialization subroutine LOGN_PARMINIT of the predefined lognormal distribution.

  Then, the constant parameters are initialized. The parameter p_n is fixed to 0.8. The parameter x_r is estimated by using the Hill's tail index estimation algorithm (Danielsson et al. 2001), which is implemented by the predefined utility function SVRTUTIL_HILLCUTOFF available to you in the Sashelp.Svrtdist library. The algorithm and the utility function are described in detail in the *SAS/ETS 9.22 User's Guide*. The function computes an estimate of x_b, which is used to compute an estimate of x_r because $x_r = x_b/\theta$, where $\theta = \exp(\mu)$ is the estimate of the scale parameter of the body distribution. If SVRTUTIL_HILLCUTOFF fails to compute a valid estimate, then the SVRTUTIL_PERCENTILE utility function, also available in the Sashelp.Svrtdist library, is used to set x_b to the p_nth percentile of the data.

  Finally, the ξ parameter of the GPD tail is initialized using the parameter initialization subroutine GPD_PARMINIT of the predefined generalized Pareto distribution. Only the values that exceed the cutoff point (x_b) are used for estimating the tail parameters.

  The Sashelp.Svrtdist library is specified with the LIBRARY= option in the PROC FCMP statement to enable the LOGNGPD_PARMINIT subroutine to use the predefined utility functions (SVRTUTIL_HILLCUTOFF and SVRTUTIL_PERCENTILE) and parameter initialization subroutines (LOGN_PARMINIT and GPD_PARMINIT).

- The LOWERBOUNDS suffix defines the lower bounds for all parameters. This subroutine is required because the parameter *Mu* has a non-default lower bound. The bounds for *Sigma* and *Xi* must be specified. If they are not specified, they are returned as missing values, which are interpreted as having no lower bound by PROC SEVERITY. You do not need to specify bounds for the constant parameters *Xr* and *Pn*.

The following DATA step statements simulate a sample from a mixed tail distribution with a lognormal body and GPD tail. The parameter p_n is fixed to 0.8, the same value used in the LOGNGPD_PARMINIT subroutine defined previously:

```
data testmixdist(keep=y label='Lognormal Body-GPD Tail Sample');
    call streaminit(45678);
    label y='Response Variable';
    N = 100; Mu = 1.5; Sigma = 0.25; Xi = 1.5; Pn = 0.8;
    /* Generate data comprising the lognormal body */
    Nbody = N*Pn;
    do i=1 to Nbody;
        y = exp(Mu) * rand('LOGNORMAL')**Sigma;
        output;
    end;

    /* Generate data comprising the GPD tail */
    cutoff = quantile('LOGNORMAL', Pn, Mu, Sigma);
    gpd_scale = (1-Pn) / pdf('LOGNORMAL', cutoff, Mu, Sigma);
    do i=Nbody+1 to N;
        y = cutoff + ((1-rand('UNIFORM'))**(-Xi) - 1)*gpd_scale/Xi;
        output;
    end;
run;
```

You can now fit the LOGNGPD distribution model to this data with PROC SEVERITY. First, you must communicate the location of the library that contains the definition of the distribution and the locations of libraries that contain any functions and subroutines that are used by your distribution model. The following OPTIONS statement sets the CMPLIB= system option to include the FCMP library Work.Sevexmpl in the search path used by PROC SEVERITY to find FCMP functions and subroutines. It is optional to specify Sashelp.Svrtdist in the search path, because PROC SEVERITY internally appends that library to the path.

```
/*--- Set the search path for functions defined with PROC FCMP ---*/
options cmplib=(work.sevexmpl);
```

The following statements use PROC SEVERITY to fit the LOGNGPD distribution model to the simulated sample. They also fit three other predefined distributions (BURR, LOGN, and GPD). The final parameter estimates are written to the Work.Parmest data set.

```
ods graphics on;
proc severity data=testmixdist print=all plots=pp outest=parmest;
    model y;
    dist logngpd;
    dist burr;
    dist logn;
    dist gpd;
run;
```

Some of the results prepared by PROC SEVERITY are shown in Figure 14 through Figure 16. The "All Fit Statistics Table" in Figure 14 shows that the model with LOGNGPD distribution has the best fit according to almost all the statistics of fit. The Burr distribution model is the closest contender to the LOGNGPD model. The lognormal and GPD distributions fit the data poorly, GPD being the worst.

Figure 14 Statistics of Fit for the Lognormal Body-GPD Tail Sample

```
                        All Fit Statistics Table

                    -2 Log
     Distribution   Likelihood      AIC         AICC        BIC          KS

     logngpd        418.78232*   428.78232*   429.42062*   441.80817     0.62140*
     Burr           424.93728    430.93728    431.18728    438.75280*    0.71373
     Logn           459.43471    463.43471    463.55842    468.64505     1.55267
     Gpd            558.13444    562.13444    562.25815    567.34478     3.43470


                        All Fit Statistics Table

                    Distribution       AD           CvM

                    logngpd          0.31670*     0.04972*
                    Burr             0.57649      0.07860
                    Logn             3.27122      0.48448
                    Gpd             16.74156      3.31860
```

The P-P plots of Figure 15 and Figure 16 provide a visual confirmation that the LOGNGPD distribution fits the tail region better than the Burr distribution.

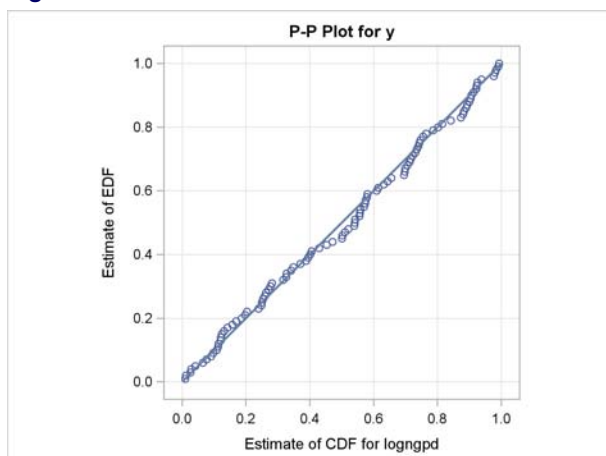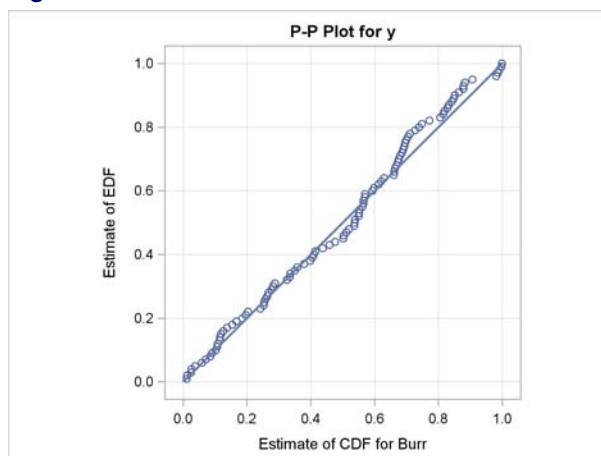**Figure 15**   P-P Plot for the LOGNGPD Model

Figure 16   P-P Plot for the Burr Model

The final parameter estimates of the LOGNGPD distribution are shown in Figure 17. The table indicates that parameters *Xr* and *Pn* are constant and displays the values used for them. The estimates of all free (nonconstant) parameters are significantly different from 0.

Figure 17   Detailed Results for the LOGNGPD Distribution

```
                Parameter Estimates for logngpd Distribution

                                    Standard                 Approx
           Parameter      Estimate     Error    t Value     Pr > |t|

           Mu              1.57921     0.06426     24.57     <.0001
           Sigma           0.31868     0.04459      7.15     <.0001
           Xi              1.03771     0.38205      2.72     0.0078
           Xr              1.27395    Constant        .          .
           Pn              0.80000    Constant        .          .
```

The following data step uses the Work.Parmest data set prepared by PROC SEVERITY to compute the value where the tail region is estimated to start $(x_b = e^\mu x_r)$ and the scale of the GPD tail distribution $(\theta_t = \frac{G(x_b)}{g(x_b)} \frac{(1-p_n)}{p_n})$:

```
data xb_thetat(keep=x_b theta_t);
    set parmest(where=(_MODEL_='logngpd' and _TYPE_='EST'));
    x_b = exp(Mu) * Xr;
    theta_t = (CDF('LOGN',x_b,Mu,Sigma)/PDF('LOGN',x_b,Mu,Sigma)) * ((1-Pn)/Pn);
run;
```

The computed values of *x_b* and *theta_t* are approximately 6.18 and 1.28, respectively. Equipped with this additional derived information, you can now interpret the results of fitting the mixed tail distribution as follows:

- The tail starts at $y \approx 6.18$. The primary benefit of using the scale-normalized cutoff (x_r) as the constant parameter instead of using the actual cutoff (x_b) is that the absolute cutoff is optimized by virtue of optimizing the scale of the body region $(\theta = e^\mu)$.

- The values $y \le 6.18$ follow the lognormal distribution with parameters $\mu \approx 1.58$ and $\sigma \approx 0.32$. These parameter estimates are reasonably close to the parameters used for simulating the sample.

- The values $y_t = y - 6.18$ $(y_t > 0)$ follow the GPD distribution with scale $\theta_t \approx 1.28$ and shape $\xi \approx 1.04$.

CONCLUSION

Modeling severity distributions is an important step towards the model-based risk management that is required for running a profitable business and complying with regulatory requirements in the insurance and banking industries. The new SEVERITY procedure in the latest release of SAS/ETS software provides a simple yet powerful tool for fitting severity distributions to continuous-valued data. It enables you to model the effects of censoring, truncation, and exogenous variables on the distribution. You can use any of eight predefined distributions commonly used for modeling the severity of losses, but the most powerful feature of PROC SEVERITY is that it enables you to define and fit any arbitrary continuous distribution to your data, including mixture distributions.

This paper has illustrated the key features of PROC SEVERITY via examples. More details about the syntax and other features of PROC SEVERITY are provided in the *SAS/ETS 9.22 User's Guide*.

REFERENCES

Danielsson, J., De Haan, L., Peng, L., and de Vries, C. G. (2001), "Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation," *Journal of Multivariate Analysis,* 76, 226–248.

Kaplan, E. L. and Meier, P. (1958), "Nonparametric Estimation from Incomplete Observations," *Journal of American Statistical Association,* 53, 457–481.

Klein, J. P. and Moeschberger, M. L. (1997), *Survival Analysis: Techniques for Censored and Truncated Data,* New York: Springer-Verlag.

Klugman, S. A., Panjer, H. H., Willmot, G. E. (1998), *Loss Models: From Data to Decisions,* New York: John Wiley & Sons.

Klugman, S. A. and Rioux, J. (2006), "Toward a Unified Approach to Fitting Loss Models," *North American Actuarial Journal,* 10(1).

Lai, T. L. and Ying, Z. (1991), "Estimating a Distribution Function with Truncated and Censored Data," *Annals of Statistics,* 19(1), 417–442.

Lynden-Bell, D. (1971), "A Method of Allowing for Known Observational Selection in Small Samples Applied to 3CR Quasars," *Monthly Notices of the Royal Astronomical Society,* 155, 95–118.

SAS Institute Inc. (2010), *SAS/ETS 9.22 User's Guide,* Cary, NC: SAS Institute Inc. http://support.sas.com/documentation/onlinedoc/ets/index.html

ACKNOWLEDGMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author:

Mahesh V. Joshi
SAS Institute Inc.
SAS Campus Drive
Cary, NC 27513
Mahesh.Joshi@sas.com