Using the QUANTLIFE Procedure for Survival Analysis

Guixian Lin and Robert N. Rodriguez, SAS Institute Inc.

ABSTRACT

The QUANTLIFE procedure implements quantile regression, which provides a direct and flexible approach to modeling survival times without the proportionality hazard constraint of the Cox model. In clinical studies, quantile regression is helpful for identifying and distinguishing important prognostic factors for patient subpopulations that are characterized by short or long survival times. This paper compares the quantile regression model with the Cox and accelerated failure time models, which are commonly used in survival analysis. An understanding of the differences between these approaches is essential for deciding which model to use in practice. An example illustrates how to estimate regression parameters and survival functions.

INTRODUCTION

In fields such as medicine, engineering, and demography, survival analysis is commonly used for modeling the time until the occurrence of some event of interest. Examples of outcome variables are the survival time for a heart transplant patient, the lifetime of a diesel engine, or the duration of employment for an individual. One of the goals of survival analysis is to characterize the survival function of the outcome response. Achieving this goal is complicated because incomplete observations are frequently encountered in lifetime data.

With the addition of the QUANTLIFE procedure, SAS/STAT$^\circledR$ software now provides four procedures (the LIFETEST, LIFEREG, PHREG, and QUANTLIFE procedures) for estimating the survival function. Each of these procedures follows a different approach. The LIFETEST procedure estimates the survival function by using a nonparametric method such as the Kaplan-Meier estimator. The LIFEREG, PHREG, and QUANTLIFE procedures incorporate covariates. The LIFEREG procedure directly models the lifetime by using a parametric regression method. In contrast, the PHREG procedure directly models the hazard function by using a semiparametric method. The PHREG procedure can flexibly account for a variety of lifetime distributions, because the underlying baseline function is left unspecified. However, in situations where you are willing to assume a parametric lifetime distribution, the parametric models available in the LIFEREG procedure lead to an efficient estimator for regression parameters.

The QUANTLIFE procedure, which was introduced in SAS/STAT 12.1, implements quantile regression for survival analysis. Because lifetime distributions are generally skewed, the quantiles of the lifetime (such as 25th, 50th, and 75th percentiles) are more informative than the mean for summarizing the lifetime distribution.

Quantile regression (Koenker and Bassett 1978) provides you with a direct approach for modeling the quantiles of a response variable as a function of covariates. By modeling the quantiles of a lifetime, you can gain insights about the dependence of the response distribution on its covariates. For example, in clinical studies, quantile regression is useful for identifying and distinguishing prognostic factors that can have different effects for patient subpopulations that are characterized by short (lower quantile) or long (upper quantile) survival times.

Compared with the parametric models in the LIFEREG procedure, the quantile regression model has two advantages: you do not need to specify a parametric form for the lifetime distribution, and you can fit a regression model for each quantile. Furthermore, you can estimate the survival function from a set of quantiles.

The next section provides more detail about the various regression approaches that are available in the LIFEREG, PHREG, and QUANTLIFE procedures for estimating the survival function. Subsequent sections provide an example and syntax details for the QUANTLIFE procedure.

ESTIMATION OF THE CONDITIONAL SURVIVAL FUNCTION

Let T be a nonnegative random variable that represents the lifetime of an individual from a homogeneous population. It is well known that you can characterize the distribution of T by the survival function $S(t)$, the cumulative distribution $F(t)$, or the hazard function $\lambda(t)$. Assuming that T is continuous, the relationship among these functions is

$$F(t) = 1 - S(t)$$
$$\lambda(t) = f(t)/S(t)$$

where $f(t)$ is the probability density function of T. Given one of these functions, you can derive the others.

A fourth (and less familiar) function that characterizes the lifetime distribution is the quantile function $Q(\tau) = F^{-1}(\tau), 0 < \tau < 1$. The set of quantile function values, $\{Q(\tau) : 0 < \tau < 1\}$, provides a complete picture of the distribution. The value τ is referred to as the quantile level. For example, $Q(0.5)$ is the median lifetime, and 0.5 is the median level.

Regardless of how you characterize the lifetime distribution, a regression model is useful for answering the question of how prognostic factors affect the lifetime. After you obtain the parameter estimates for the model, you can estimate the survival function conditional on given covariates. The LIFEREG, PHREG, and QUANTLIFE procedures estimate the survival function in different ways.

The LIFEREG procedure fits a parametric accelerated failure time (AFT) model for the log transformation of the lifetime,

$$\log(T) = x'\boldsymbol{\beta} + \sigma\,\boldsymbol{\epsilon}$$

where x is a vector of covariates, $\boldsymbol{\beta}$ is a vector of unknown regression parameters, σ is an unknown scale parameter, and $\boldsymbol{\epsilon}$ is an error term that has an assumed distribution, such as the standard normal distribution. If $S_0(t)$ denote the survival function of the error term $\boldsymbol{\epsilon}$, then the conditional survival function of T is

$$S(t|x) = S_0(\frac{\log(t) - x'\boldsymbol{\beta}}{\sigma})$$

The PHREG procedure fits the Cox proportional hazards model,

$$\lambda(t|x) = \lambda_0(t)e^{x'\boldsymbol{\beta}}$$

where $\lambda_0(t)$ is the baseline hazard function. The corresponding baseline survival function is $S_0(t) = e^{-\int_0^t \lambda_0(u)\mathrm{d}u}$. You can use the PHREG procedure to obtain the estimates $\hat{\boldsymbol{\beta}}$, $\hat{S}_0(t)$, and $\hat{S}(t|x) = \hat{S}_0(t)e^{x'\hat{\boldsymbol{\beta}}}$.

The QUANTLIFE procedure fits a linear quantile regression model for a log transformation of the lifetime as

$$Q_{\log(T)}(\tau|x) = x'\boldsymbol{\beta}(\tau)$$

where $Q_{\log(T)}(\tau|x)$ is the τth quantile of $\log(T)$ at x. The quantile regression model for T is $Q_T(\tau|x) = e^{x'\boldsymbol{\beta}(\tau)}$, because the quantile function is invariant under a monotone transformation. Thus, the QUANTLIFE procedure applies to survival analysis when your goal is to assess varying covariate effects on the lifetime.

On the other hand, if your goal is to estimate the survival function at given x, you can also use the QUANTLIFE procedure by taking advantage of the relationship between the quantile function and the survival function. By the definition of the quantile function,

$$F(Q_T(\tau|x)) = P(T \leq Q_T(\tau|x)) = \tau$$

In other words, the cumulative distribution function $F_T(t|x)$ maps $Q_T(\tau|x)$ to τ, so the corresponding survival function $S_T(t|x)$ maps $Q_T(\tau|x)$ to $1 - \tau$. The following example illustrates the relationship between the quantile function and the survival function in a regression model setting.

Relationship of Quantile Function and Survival Function

Given a covariate x, suppose that the lifetime T follows the model

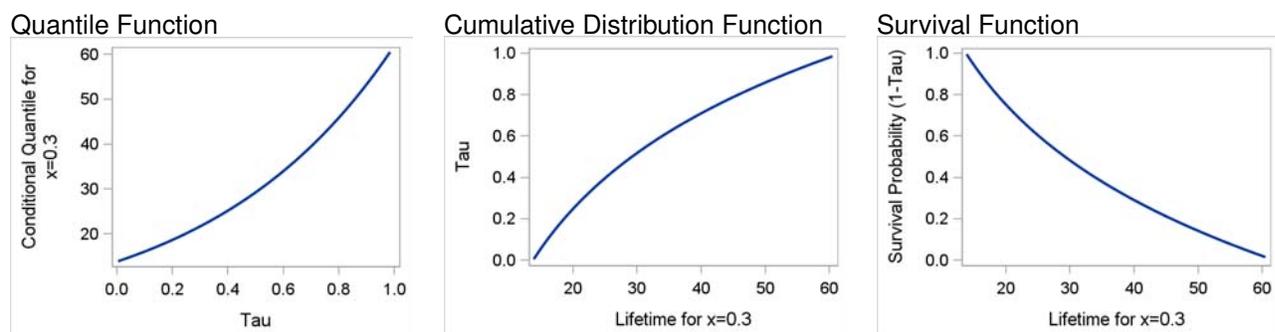$$\log(T) = 3 + 1.25x + \sigma(x)\,\epsilon,\ x > 0$$

where ϵ has the uniform distribution $U(-0.5, 0.5)$, and $\sigma(x) = 5x$ is a scale parameter that depends on the covariate x.

Because ϵ has the uniform distribution $U(-0.5, 0.5)$, the quantile function of ϵ is $Q_\epsilon(\tau) = \tau - 0.5$. Thus, the quantile function of the untransformed lifetime T conditional on x is

$$Q_T(\tau|x) = e^{3+1.25x+5xQ_\epsilon(\tau)} = e^{3+(1.25+5(\tau-0.5))x}$$

Figure 1 shows the equivalence of the quantile function, the cumulative distribution function, and the survival function at $x = 0.3$.

Figure 1   Conditional Quantile, Cumulative Distribution, and Survival Functions

The appendix provides a SAS® program that constructs Figure 1.

Comparison of Fitted Survival Functions

This section compares the survival functions that are estimated for a simulated data set by using the LIFEREG, PHREG and QUANTLIFE procedures. The lifetimes are generated from a regression model that has heterogeneous errors,

$$\log(T) = 3 + 1.25x + \sigma(x)\,\epsilon$$

where x has the uniform distribution $U(0, 0.5)$, ϵ has the uniform distribution $U(-0.5, 0.5)$, and $\sigma(x) = 5x$ is a scale parameter that depends on the covariate x. Note that the choice of $\sigma(x)$ introduces heterogeneity in the data.

Among the 200 observations, 47% are randomly censored. The SAS program for this example is provided in the appendix.

Figure 2 compares the three estimated survival functions with the true underlying survival function $S(t|x = 0.3)$. The QUANTLIFE procedure provides a better estimate than the other two procedures do. The reason is that the true model for $\log(T)$ is a linear quantile model, whereas the assumptions for the Cox proportional hazard model and the AFT model are violated. In the AFT model, the scale parameter is assumed to be constant.

See the appendix for the SAS program that was used to construct Figure 2.

Figure 2 Estimated Survival Functions from the LIFEREG, PHREG, and QUANTLIFE Procedures
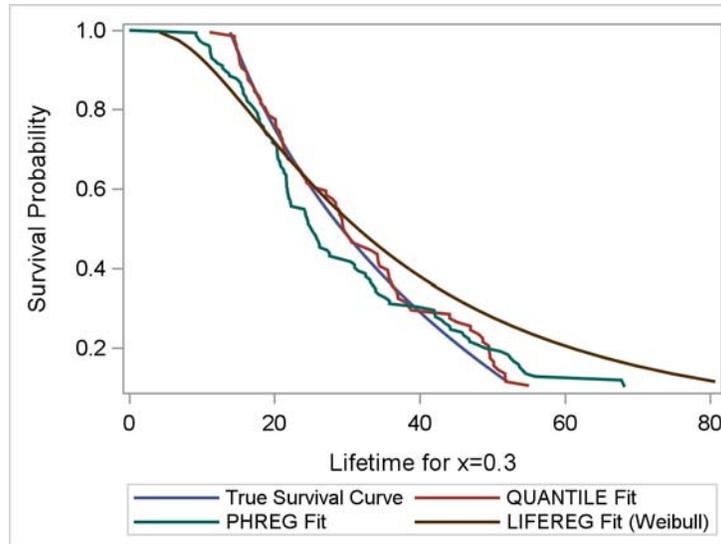
In the regression model, the conditional quantile function of log(T) is $Q_{\log(T)}(\tau|x) = 3 + (5\tau - 1.25)x$. That is, the true parameter $\beta_0(\tau)$ for the intercept is 3, and the true parameter $\beta(\tau)$ for x is $5\tau - 1.25$, which varies with τ. Figure 3 displays estimates for $\beta_0(\tau)$ and $\beta(\tau)$ from the QUANTLIFE procedure at the quantile levels 0.1, 0.15, 0.2, ..., 0.9. The plot on the right displays the increasing effect of x. Specifically, x has a negative effect on short survival times (quantile levels less than 0.25) and a positive effect on long survival times.

Figure 3 Parameter Estimates with 95% Confidence Bands
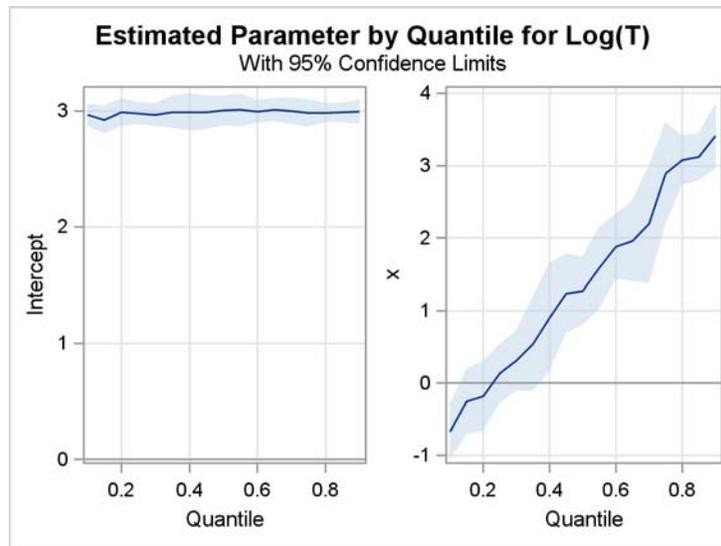
METHODS FOR ANALYZING CENSORED DATA IN PROC QUANTLIFE

Lifetime data often contain incomplete observations, which result from censoring (Klein and Moeschberger 2003; Hosmer, Lemeshow, and May 2008). The regression methods in the LIFEREG and PHREG procedures allow for censoring in a variety of ways; for more information, see the *SAS/STAT 12.1 User's Guide*.

Likewise, the QUANTLIFE procedure provides two quantile regression methods that are appropriate for incomplete data with right censoring: the Kaplan-Meier-type method and the Nelson-Aalen-type method. Portnoy (2003) proposed the Kaplan-Meier-type method to estimate conditional quantile functions by generalizing the idea of the Kaplan-Meier estimator of the survival function, and Peng and Huang (2008)

4

proposed the Nelson-Aalen-type method to estimate conditional quantile functions by generalizing the Nelson-Aalen estimator of the cumulative hazard function.

In the special case where lifetimes are fully observed, you can use the QUANTREG procedure to model the lifetimes with a standard quantile regression model. The regression methods in the QUANTREG and QUANTLIFE procedures have much in common: they are distribution-free, they are appropriate with heterogeneous data, and they use linear programming techniques to estimate the parameters. However, when the data are censored, the standard quantile regression approach leads to biased estimates, so the QUANTLIFE procedure is strongly recommended.

EXAMPLE: PRIMARY BILIARY CIRRHOSIS STUDY

This example shows how you can use the QUANTLIFE procedure to detect varying covariate effects on the survival time, and it illustrates key features of the procedure.

The data for this example are from a study of primary biliary cirrhosis, a rare but fatal chronic liver disease; see Fleming and Harrington (1991). Researchers followed 418 patients who had this disease, 161 of whom died during the study.

The data set contains the following variables:

- **Time**, follow-up time in years

- **Status**, event indicator with value 1 if death time is observed and with value 0 if the time is censored

- **Age**, age in years from birth to study registration

- **Albumin**, serum albumin level in g/dL

- **Bilirubin**, serum bilirubin level in mg/dL

- **Edema**, presence of edema

- **Protime**, prothrombin time in seconds

The following statements create the data set:

```
data pbc;
   input Time Status Age Albumin Bilirubin Edema Protime @@;
   label Time="Follow-up Time in Days";
   logAlbumin   = log(Albumin);
   logBilirubin = log(Bilirubin);
   logProtime   = log(Protime);
   datalines;
  400 1 58.7652 2.60 14.5 1.0 12.2 4500 0 56.4463 4.14  1.1 0.0 10.6
 1012 1 70.0726 3.48  1.4 0.5 12.0 1925 1 54.7406 2.54  1.8 0.5 10.3
 1504 0 38.1054 3.53  3.4 0.0 10.9 2503 1 66.2587 3.98  0.8 0.0 11.0
 1832 0 55.5346 4.09  1.0 0.0  9.7 2466 1 53.0568 4.00  0.3 0.0 11.0
 2400 1 42.5079 3.08  3.2 0.0 11.0   51 1 70.5599 2.74 12.6 1.0 11.5
 3762 1 53.7139 4.16  1.4 0.0 12.0  304 1 59.1376 3.52  3.6 0.0 13.6

    ... more lines ...

  989 0 35.0000 3.23  0.7 0.0 10.8  681 1 67.0000 2.96  1.2 0.0 10.9
 1103 0 39.0000 3.83  0.9 0.0 11.2 1055 0 57.0000 3.42  1.6 0.0  9.9
  691 0 58.0000 3.75  0.8 0.0 10.4  976 0 53.0000 3.29  0.7 0.0 10.6
;
```

The next statements fit a linear quantile regression model for the log of the survival time by using the covariates **logBilirubin**, **logProtime**, **logAlbumin**, **Age**, and **Edema**:

```
ods graphics on;
proc quantlife data=pbc log method=na plot=(quantplot survival) seed=1268;
   model Time*Status(0)=logBilirubin logProtime logAlbumin Age Edema
                    / quantile=(0.1 0.2 0.3 0.4 0.5 0.6 0.75);
run;
```

Note that the basic syntax for the MODEL statement is the same as that for the MODEL statements in the LIFEREG and PHREG procedures. The LOG option in the PROC QUANTLIFE statement specifies that the log of **Time** be analyzed as the response variable. When you use the QUANTLIFE procedure (as when you use the QUANTREG procedure), you can include the QUANTILE= option to specify the quantiles of interest for comparing quantile-specific covariate effects. The METHOD= option specifies the Nelson-Aalen method for estimating the regression parameters.

The QUANTLIFE procedure provides resampling methods for computing confidence limits for the parameters. By default, the repetition number is 200. You can request a different number of repetitions with the NREP= option in the PROC QUANTLIFE statement. By specifying the SEED= option for generating random numbers, you can later reproduce the confidence interval results.

Figure 4 displays the model information and information about censoring in the data. Out of 418 observations, 257 are censored; the censoring rate is 61.48%.

Figure 4   Model Information

```
                    The QUANTLIFE Procedure


                       Model Information


    Data Set                                 WORK.PBC
    Dependent Variable                      Log(Time)
    Censoring Variable                         Status
    Censoring Value(s)                              0
    Number of Observations                        418
    Method                              Nelson-Aalen
    Replications                                  200
    Seed for Random Number Generator             1268



    Summary of the Number of Event and Censored Values


                                          Percent
          Total       Event    Censored   Censored

            418         161         257      61.48
```

Figure 5 shows the parameter estimates. Each quantile level has a set of parameter estimates and confidence limits.

Figure 5 Parameter Estimates at Different Quantiles

```
                              Parameter Estimates

                                     Standard    95% Confidence
    Quantile Parameter      DF Estimate    Error        Limits       t Value Pr > |t|

      0.1000 Intercept       1  14.8030   4.0967    6.7736   22.8325    3.61   0.0003
             logBilirubin    1  -0.4488   0.1485   -0.7398   -0.1578   -3.02   0.0027
             logProtime      1  -3.6378   1.4560   -6.4915   -0.7841   -2.50   0.0129
             logAlbumin      1   1.9286   0.9756    0.0165    3.8408    1.98   0.0487
             Age             1  -0.0244   0.0107   -0.0455  -0.00334   -2.27   0.0237
             Edema           1  -1.0712   0.6688   -2.3820    0.2396   -1.60   0.1100
      0.2000 Intercept       1  15.1800   2.6664    9.9540   20.4060    5.69   <.0001
             logBilirubin    1  -0.6532   0.0886   -0.8268   -0.4796   -7.37   <.0001
             logProtime      1  -3.3273   0.9401   -5.1699   -1.4847   -3.54   0.0004
             logAlbumin      1   1.6842   0.6888    0.3343    3.0342    2.45   0.0149
             Age             1  -0.0291  0.00687   -0.0425   -0.0156   -4.23   <.0001
             Edema           1  -0.7265   0.3179   -1.3497   -0.1034   -2.29   0.0228
      0.3000 Intercept       1  13.2382   2.5296    8.2804   18.1961    5.23   <.0001
             logBilirubin    1  -0.6013   0.0762   -0.7506   -0.4521   -7.90   <.0001
             logProtime      1  -2.5816   0.8907   -4.3273   -0.8359   -2.90   0.0039
             logAlbumin      1   1.7246   0.7142    0.3248    3.1245    2.41   0.0162
             Age             1  -0.0244  0.00716   -0.0385   -0.0104   -3.41   0.0007
             Edema           1  -0.8577   0.2763   -1.3992   -0.3163   -3.10   0.0020
      0.4000 Intercept       1  13.4716   3.0874    7.4204   19.5228    4.36   <.0001
             logBilirubin    1  -0.6047   0.0846   -0.7705   -0.4389   -7.15   <.0001
             logProtime      1  -2.1632   1.1726   -4.4615    0.1351   -1.84   0.0658
             logAlbumin      1   0.9819   0.7191   -0.4274    2.3912    1.37   0.1728
             Age             1  -0.0255  0.00681   -0.0389   -0.0122   -3.74   0.0002
             Edema           1  -1.0589   0.3104   -1.6672   -0.4506   -3.41   0.0007
      0.5000 Intercept       1  10.9205   2.8047    5.4235   16.4175    3.89   0.0001
             logBilirubin    1  -0.5315   0.0904   -0.7087   -0.3543   -5.88   <.0001
             logProtime      1  -1.2222   1.2142   -3.6020    1.1577   -1.01   0.3148
             logAlbumin      1   1.5700   0.6284    0.3383    2.8016    2.50   0.0129
             Age             1  -0.0318  0.00883   -0.0491   -0.0145   -3.60   0.0004
             Edema           1  -0.7316   0.3743   -1.4653   0.00202   -1.95   0.0513
      0.6000 Intercept       1  11.2381   2.6294    6.0846   16.3917    4.27   <.0001
             logBilirubin    1  -0.5701   0.0852   -0.7370   -0.4031   -6.69   <.0001
             logProtime      1  -1.3508   1.1402   -3.5856    0.8840   -1.18   0.2368
             logAlbumin      1   1.3704   0.5091    0.3726    2.3682    2.69   0.0074
             Age             1  -0.0226   0.0109   -0.0440  -0.00111   -2.06   0.0399
             Edema           1  -0.5141   0.3088   -1.1193    0.0912   -1.66   0.0968
      0.7500 Intercept       1  10.0954   3.1893    3.8445   16.3463    3.17   0.0017
             logBilirubin    1  -0.6366   0.1071   -0.8466   -0.4267   -5.94   <.0001
             logProtime      1  -0.9670   1.2343   -3.3862    1.4521   -0.78   0.4338
             logAlbumin      1   1.8148   0.5883    0.6618    2.9678    3.08   0.0022
             Age             1  -0.0203   0.0156   -0.0509    0.0102   -1.30   0.1931
             Edema           1  -0.3529   0.3120   -0.9644    0.2586   -1.13   0.2587
```

For comparison, the following statements use the LIFEREG procedure to fit a Weibull distribution to the data. The LIFEREG procedure fits an accelerated failure time model, which assumes that the effect of independent variables is multiplicative on the event time. The effect of the independent variables is additive for log(**Time**).

```
proc lifereg data=pbc;
   model Time*Status(0)=logBilirubin logProtime logAlbumin Age Edema;
run;
```

Figure 6 shows the parameter estimates that are computed by the LIFEREG procedure.

Figure 6 Parameter Estimates from PROC LIFEREG

```
                    The LIFEREG Procedure

           Analysis of Maximum Likelihood Parameter Estimates

                           Standard    95% Confidence       Chi-
    Parameter      DF Estimate    Error        Limits       Square Pr > ChiSq

    Intercept       1  12.2155   1.4539   9.3658  15.0651    70.59     <.0001
    logBilirubin    1  -0.5770   0.0556  -0.6861  -0.4680   107.55     <.0001
    logProtime      1  -1.7565   0.5248  -2.7850  -0.7280    11.20     0.0008
    logAlbumin      1   1.6694   0.4276   0.8313   2.5074    15.24     <.0001
    Age             1  -0.0265   0.0053  -0.0368  -0.0162    25.35     <.0001
    Edema           1  -0.6303   0.1805  -0.9842  -0.2764    12.19     0.0005
    Scale           1   0.6807   0.0430   0.6014   0.7704
    Weibull Shape   1   1.4691   0.0928   1.2980   1.6628
```

The parameter estimates are similar to those obtained for the median by the QUANTLIFE procedure. Note that the p-value from the LIFEREG procedure for **logProtime** is very small. For this same variable, the p-values from the QUANTLIFE procedure are 0.3148 for the median and 0.4338 for the 0.75th quantile, and the p-values are much smaller for the lower quantiles. The QUANTLIFE procedure reveals that the effect of **logProtime** depends on which side of the response distribution is being modeled.

The PLOT=QUANTPLOT option in the PROC QUANTLIFE statement requests quantile process plots. The first four of these plots are shown in Figure 7; they plot the estimated regression parameter against the quantile level. You can use these plots to compare quantile-specific covariate effects. A curve that is not constant can indicate heterogeneity in the data. The interpretation of the regression coefficients at a given quantile is similar to the interpretation of the coefficient in classical regression analysis. That is, the coefficient for a given covariate indicates the effect on log(**Time**) of a unit change in the covariate, assuming that the other covariates are fixed.

Figure 7 Quantile Processes with 95% Confidence Bands
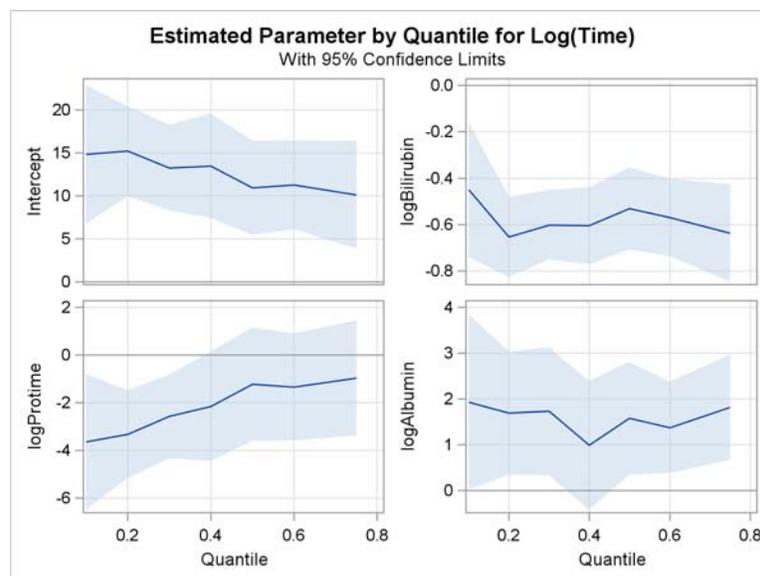
In Figure 7, you can see that the effect of **logProtime** is negative over the lower quantiles and that the effect diminishes in magnitude at the median and upper quantiles. You would miss this insight if you used the accelerated failure time model.

The PLOT=SURVIVAL option in the PROC QUANTLIFE statement requests the survival function plot. Figure 8 shows the estimated survival function at the reference set of covariate values, which are reference levels for the CLASS variables and are average values for the continuous variables.

Figure 8 Survival Function
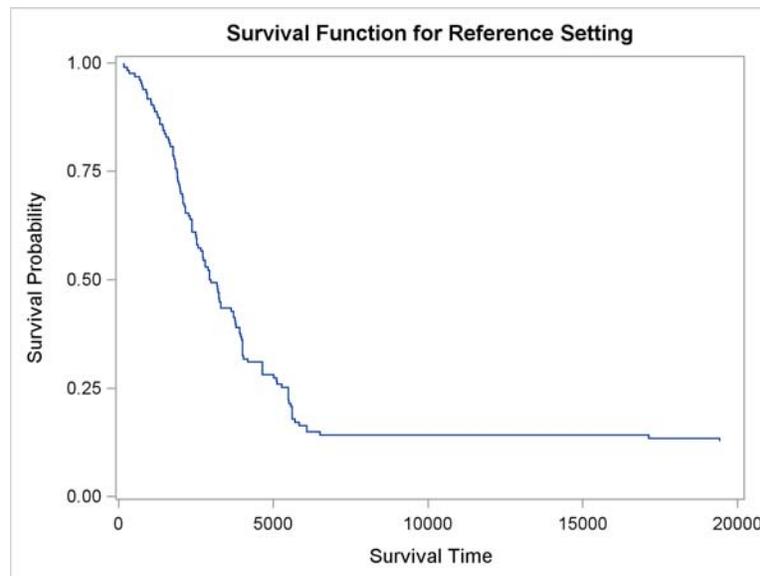
MAIN FEATURES AND SYNTAX

The syntax for PROC QUANTLIFE closely follows that of PROC PHREG and PROC QUANTREG. The main statements available in the QUANTLIFE procedure are as follows:

> **PROC QUANTLIFE** < *options* > **;**
>   **BASELINE** < *options* > **;**
>   **CLASS** *variables* **;**
>   **EFFECT** *name = effect-type (variables < / options >)* **;**
>   **MODEL** *response* < ∗ *censor*(*list*) > **=** < *effects* > < / **QUANTILE=** > **;**
>   **OUTPUT** < **OUT=**SAS-data-set > < keyword**=**name . . . keyword**=**name > **;**
>   **TEST** *effects* < / *options* > **;**

The PROC QUANTLIFE statement invokes the procedure. The PROC QUANTLIFE and MODEL statements are required. The MODEL statement specifies the variables to be used in the regression. The QUANTILE= option specifies the quantile levels of interest. The CLASS statement specifies which explanatory variables to treat as categorical variables. The EFFECT statement enables you to construct special collections of columns for design matrices. These collections are referred to as constructed effects to distinguish them from the usual model effects that are formed from continuous or classification variables.

The BASELINE statement creates an output data set that contains the survival function estimates or the conditional quantile function estimates for every set of covariates in the COVARIATES= data set that is optionally specified in this statement. If the COVARIATES= data set is not specified, PROC QUANTLIFE uses a reference set of covariates that consists of the reference levels for the CLASS variables and the average values for the continuous variables. The OUTPUT statement creates an output data set that contains predicted values, residuals, and estimated standard errors. The TEST statement requests linear tests for the model parameters. See Chapter 77, "The QUANTLIFE Procedure (Experimental)" (*SAS/STAT User's Guide*), for the full syntax.

SUMMARY

SAS/STAT 12.1 software introduces the QUANTLIFE procedure, which uses quantile regression to analyze survival data. To decide when to use this approach, you should understand how it compares with standard methods for survival analysis that are available in the LIFETEST, LIFEREG, and PHREG procedures. Each of these methods has its advantages and limitations.

Table 1 is an overview of the major distinctions among these procedures. The similarities and differences among the modeling methods are subtle; see Koenker and Geling (2001) for a helpful explanation.

Table 1   Comparison of Procedures for Survival Analysis

Procedure	Focus	Method	Covariates	Censoring Type
LIFETEST	Survival function	Nonparametric	No	Right
LIFEREG	Lifetime	Parametric	Yes	Right, left, interval
PHREG	Hazard function	Semiparametric	Yes	Right
QUANTLIFE	Lifetime	Semiparametric	Yes	Right

Quantile regression is especially useful when you are interested in modeling the survival time and when the effects of covariates on the lifetime distribution differ with the covariate level. You can use the QUANTLIFE procedure to explore these effects in lifetime data that involve certain patterns of heterogeneity—for example, when the variation in the lifetime increases with a continuous covariate.

Quantile regression is a distribution-free approach in the sense that inference about the regression parameters for a particular quantile of the lifetime depends only on the conditional distribution near that quantile. By comparison, the accelerated failure time (AFT) model in the LIFEREG procedure is more restrictive in its parametric assumption.

After a suitable tranformation of the survival time, the Cox proportional model in the PHREG procedure and the AFT model involve an iid error assumption (Koenker and Geling 2001). This implies that the covariate effect can shift the location but not the shape of the conditional density for the transformed lifetime. The additional flexibility of quantile regression for modeling the shape can be important in applications where you are concerned about treatment effects on longer lifetimes.

The QUANTLIFE procedure and the LIFEREG procedure are similar in that both use a regression method to model the lifetime. The LIFEREG procedure provides an efficient estimator for the regression parameters if you are willing to assume a parametric distribution for the lifetime. The regression coefficients that are obtained using the LIFEREG procedure are interpreted as the effect on the mean or median of the lifetime, whereas the regression coefficients that are obtained using the QUANTLIFE procedure apply to specified quantiles of the lifetime.

Unlike the QUANTLIFE procedure, the PHREG procedure models the hazard function. Both of these procedures are semiparametric, but in different ways. The Cox model requires no parametric assumption about the baseline hazard function. A second major advantage of the Cox model is that it can incorporate time-dependent covariates.

APPENDIX: SAS PROGRAM

```
/* Plots for the true quantile function and survival function */

%let x=0.3;
data QuantileSurvival(keep=t surv tau);
   do tau=0.005 to 0.99 by 0.02;
      t=exp(3 + 1.25* &x + 5*&x*(tau-0.5));
      surv=1-tau;
      output;
   end;
run;


proc sgplot data=QuantileSurvival;
```

```
      yaxis label="Conditional Quantile for x=0.3";
      xaxis label="Tau";
      series x=tau y=t/lineattrs=(thickness=2);
run;
proc sgplot data=QuantileSurvival;
   yaxis label="Tau";
   xaxis label="Lifetime for x=0.3";
   series x=t y=tau/lineattrs=(thickness=2);
run;
proc sgplot data=QuantileSurvival;
   yaxis label="Survival Probability (1-Tau)";
   xaxis label="Lifetime for x=0.3";
   series x=t y=surv/lineattrs=(thickness=2);
run;


/* Example comparing survival functions from LIFEREG, PHREG, and QUANTLIFE */

%let n=200;   /* Sample size*/
%let x=0.3;   /* Estimated survival function at x=0.3 */

data noniid;
   control=0;
   do i=1 to &n;
     x=ranuni(1);
     e=5*x*(ranuni(1)-0.5);
     censor=0;

     T =exp(3 + 1.25*x + e);   /* Lifetime T from the true model */
     ci=exp(2.8 + 1.4*x + 2.5*(ranuni(1)-0.2)); /* Censoring time variable ci */

     if(T>ci) then do;
        censor=1;   /* Right censoring */
        T=ci;
      end;
     output;
     if(i=&n) then do;
      x=&x;   /* Predict survival at x=&x; in the paper, x=0.3 */
      T=.;
      control=1;
      output;
      end;
   end;
run;


data PredAtX;
   x=&x;
run;


/* Compare survival functions from LIFEREG, PHREG, and QUANTLIFE */

ods graphics on;
proc lifereg data=noniid;
   model T*censor(1)=x / distribution=lnormal;
   output out=lfout control=control quantile=0.005 to 0.99 by 0.02 predicted=pred;
run;


proc phreg data=noniid;
   model T*censor(1)=x;
   baseline  out=phout covariates=PredAtX survival=phsurv;
run;


proc quantlife data=noniid method=na log plot=quantplot;
   model T*censor(1)=x / quantile=0.1 to 0.9 by 0.05;
```

```
        baseline out=qlout covariates=PredAtX survival=qlsurv quantile=qly;
    run;

    %let survfcutoff=0.1; /* Survival curves are plotted with survival probability>=survfcutoff */

    /* Survival function from PROC LIFEREG */
    data lfsurvdata(keep=lfy lfsurv);
        set lfout;
        lfy=pred;
        lfsurv=1-_PROB_;
        if lfsurv>=&survfcutoff;
    run;

    /* Survival function from PROC PHREG */
    data phsurvdata(keep=phy phsurv);
        set phout;
        if phsurv>=&survfcutoff;
        phy=T;
    run;

    /* Survival function from PROC QUANTLIFE */
    data qlsurvdata(keep=qly qlsurv);
        set qlout;
        if qlsurv>=&survfcutoff;
    run;

    /* True survival function for simulated model */
    data truesurvdata(keep=y0 surv0);
        do tau=0.005 to 0.99 by 0.02;
            y0=exp(3 + 1.25* &x + 5*&x*(tau-.5));
            surv0=1-tau;
            if(surv0>=&survfcutoff) then output;
        end;
    run;

    data survdata;
        set  lfsurvdata phsurvdata qlsurvdata truesurvdata;
    run;
    proc sgplot data=survdata;
        yaxis label="Survival Probability";
        xaxis label="Lifetime for x=0.3";
        series x=y0  y=surv0 / legendlabel='True Survival Curve' lineattrs=(thickness=2);
        series x=qly y=qlsurv/ legendlabel='QUANTILE Fit' lineattrs=(thickness=2);
        series x=phy y=phsurv/ legendlabel='PHREG Fit' lineattrs=(thickness=2);
        series x=lfy y=lfsurv/ legendlabel='LIFEREG Fit (Weibull)' lineattrs=(thickness=2);
    run;
```

REFERENCES

Fleming, T. R. and Harrington, D. P. (1991), *Counting Processes and Survival Analysis*, New York: John Wiley & Sons.

Hosmer, D. W., Jr., Lemeshow, S., and May, S. (2008), *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*, 2nd Edition, John Wiley & Sons.

Klein, J. P. and Moeschberger, M. L. (2003), *Survival Analysis: Techniques for Censored and Truncated Data*, 2nd Edition, New York: Springer-Verlag.

Koenker, R. and Bassett, G. W. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50.

Koenker, R. and Geling, O. (2001), "Reappraising Medfly Longevity: A Quantile Regression Survival Analysis," *Journal of the American Statistical Association*, 96, 458–468.

Peng, L. and Huang, Y. (2008), "Survival Analysis with Quantile Regression Models," *Journal of the American Statistical Association*, 103, 637–649.

Portnoy, S. (2003), "Censored Regression Quantiles," *Journal of the American Statistical Association*, 98, 1001–1012.

ACKNOWLEDGMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. You can contact the authors at the following addresses:

Guixian Lin   Robert N. Rodriguez
SAS Institute Inc.   SAS Institute Inc.
SAS Campus Drive   SAS Campus Drive
Cary, NC 27513   Cary, NC 27513
Guixian.Lin@sas.com   Bob.Rodriguez@sas.com